AIR®
AMERICAN INSTITUTES FOR RESEARCH®

Results of the Three Year Impact Evaluation of Zambia's Cash Transfer Program in Monze District
Final Report
June 2011

David Seidenfeld
dseidenfeld@air.org
American Institutes for Research
Washington, D.C. USA

Sudhanshu Handa
shanda@email.unc.edu
University of North Carolina
Department of Public Policy
Chapel Hill, NC USA

Additional contributors:
Leah Prencipe
leahprencipe@gmail.com

Carolyn Huang
Department of Public Policy, UNC
chuang@email.unc.edu

Gelson Tembo
Palm Associates
tembogel@gmail.com

Dan Sherman
American Institutes for Research
dsherman@air.org

Table of Contents

Acknowledgements

This is a technical report conducted by the American Institutes for Research (AIR), contracted by UNICEF, and funded by the Cooperating Partners – UNICEF, DFID, and Irish Aid.

We would like to recognize the contributions of many individuals and organizations without which it would not have been possible to complete this study. Our thanks go to the Zambian Ministry of Community Development and Social Services (MCDSS), the Department for International Development (DFID), the United Nations Children Fund (UNICEF), Irish Aid, and Palm Associates for the opportunity to carry out this study, and for the financial and/or technical support that they rendered. Our special thanks go to Dr. Gelson Tembo of Palm Associates for carrying out the data collection, Keri Culver of AIR for her help with data collection, and Dr. Charlotte Harland (UNICEF) and Ms. Kelley Toole (DfID) for their technical support during the design and field work. The value of the logistical support obtained from Mr. Stanfield Michelo and Mr. Manzunzo Zulu (MCDSS, Lusaka) also cannot be over-emphasized. Mr. Manzunzo Zulu (MCDSS in Monze) provided valuable logistical support during data collection in Monze, including provision of programme background information and access to the scheme database, which was a key input.

Our acknowledgements would be incomplete without mentioning our team of very able research assistants. Specifically, we acknowledge the input of the team of enumerators and supervisors from Palm Associates whose dedication during data collection ensured that the data collected were of high quality. The highly competent team of data entry personnel at Palm Associates is also greatly acknowledged.

The patience exercised by the households and community leaders and members during long hours of interviews are also greatly acknowledged. It is our hope that the insights from the information that they provided will translate into valuable interventions in their communities.

David Seidenfeld
Sudhanshu Handa

Acronym List

ADEQ	Adult Equivalent
ADLs	Activities of Daily Living
AIR	American Institutes for Research
CCT	Conditional Cash Transfer
CT/ OVC	Cash Transfer / Orphans and Vulnerable Children
CWAC	Community Welfare Assistance Committee
DD	Difference-in-Differences
DEF	Reports Design Effects
HAZ	Height-for-Age Z Score
HIV/ AIDS	Human Immunodeficiency Virus/Acquired Immune Deficiency Syndrome
MCDSS	Ministry of Community Development and Social Services
MCDSS/SSN	Ministry of Community Development and Social Services Operations Manual
MDES	Minimum Detectable Effect Size
LCMS	Living Conditions and Monitoring Survey
PSM	Propensity Score Matching
RCT	Randomized Control Trial
SCT	Social Cash Transfer
SSA	Sub-Saharan Africa
UNAIDS	Joint United Nations Programme on HIV/AIDS
UNICEF	United Nations Children's Fund
WAZ	Weight-for-Age Z Score
WHZ	Weight-for-Height Z Score
ZDHS	Zambian Demographic and Health Survey

Executive Summary

Background

In 2007 Zambia's Ministry of Community Development and Social Services (MCDSS) began implementing a cash transfer program in Monze district and an experimental evaluation design with baseline data collection accompanied the program[1]. The Monze cash transfer program is based on the Kalomo model that targets labor constrained and destitute households as defined by the operations manual (MCDSS/SSN 2007). Beneficiary households receive 40,000 or 50,000 kwacha a month (equivalent to $8 or $10 respectively) depending on if the household has children, in which case they receive the higher amount. Payments are made every other month and there are no conditions to receive the money. AIR was contracted by UNICEF Zambia in 2010 to conduct the follow up data collection, analysis and reporting for the three year impact evaluation of the program. This report presents findings from AIR's work on three aspects of the program:

- primary effects that include education, health, spending, and consumption;
- secondary effects including expectations of the future, discount rate, and quality of life; and
- program operations, including validating payments, accessibility to payments, and understanding program policies.

These results cover a three-year period and include 510 beneficiary households

Study Design

The Monze impact evaluation was initially designed to be a randomized control trial with assignment of communities to treatment and control. However, the evaluation presented here used a quasi-experimental design with random assignment at the community level and selection at the household level, requiring a matched comparison group. It was necessary to employ a quasi-experimental approach for defining a comparison group rather than randomized assignment to measure treatment effects due to the selection process that occurred in the treatment Community Welfare Assistance Committees (CWACs) but not in the control CWACs. This situation resulted because the baseline data collection occurred before the final selection of beneficiaries in both the treatment and control CWACs. The need to model selection in control CWACs potentially weakens our ability to make causal inferences because we cannot account for unobserved differences between treatment and control samples. We implement a propensity score matching approach to create comparison groups within the context of a differences in differences estimation framework, which has been shown to perform extremely well at replicating the experimental benchmark in social experiments (Heckman, Todd, and Ichimura, 1997). Therefore, we believe that we can identify the effects of the cash transfer program on

[1] The baseline study was conducted by Mazdar and Palm Associates

beneficiaries and have a strong argument for attributing observed differences to the impact of the cash transfer program.

Results

Our analyses investigate effects over a three-year period on a range of outcomes including nutrition, health, education, labor, and agricultural activity. In addition to these primary outcomes, we examine the program's impacts on individuals' expectations of the future, discount rates, and self assessed quality of life. Last, we also consider the quality of program implementation by the MCDSS, the ministry administering the cash transfer program in Monze.

Main Results

We find mixed results for impacts on primary outcomes.

- On the production side we find strong impacts on livestock ownership, particularly goats and chicken, and among smaller households, pig ownership for beneficiaries. Program households are more likely to purchase fertilizer and to produce a greater quantity of cash crops; there also appears to be a shift away from maize for direct consumption and towards more cash cropping (groundnut, sweet potato) for sale. The erratic schedule of payments by the ministry to beneficiaries could be one contributing factor to this finding as beneficiaries would receive several payments at one time, enabling them to make investments that might not otherwise be possible if the payments were smaller and more regular as was intended.

- We find strong impacts on school enrollment, in a similar range to other programs (seven percentage points), and very strong impacts on enrollment of younger children (20 percentage points) indicating that the program has an effect on on-time school entry.

- We find no impacts on food expenditures or food composition. We believe this is because the expenditure module, which only covers food, is missing important items and is not sensitive enough to capture changes in food expenditure, especially at such low levels of spending. Additionally, the delays in payments to beneficiaries, especially in the months prior to the follow-up data collection would affect their spending in the month prior to data collection, the expenditure period assessed in the follow-up instrument.

- There are no statistically significant impacts on health outcomes such as having an under five card, attending checkups, and curative care for either young children (age five and under) or school-age children, which is consistent with findings from the Kenya CT-OVC evaluation. For young children, the sample size is extremely small (720) and the study therefore lacks sufficient power to detect effects among this group given its size, even if they were to exist.

These results suggest that the program impacts economic production and investment in education, but that these impacts do not necessarily carry through to nutrition and health outcomes. The frequent and long delays in payment of funds to beneficiaries might explain these findings.

Secondary Results

Although cash transfer programs primarily focus on affecting expenditures at the time transfers are made they can potentially affect attitudes and expectations in a way that influence future behavior. We find interesting results on secondary outcomes related to expectations of future quality of life and preferences for delayed gratification that, as far as we know, have never been tested before in a cash transfer evaluation. These outcomes are linked to important behavior change for investing, saving, and avoiding unnecessary risk.

- We find a strong impact on beneficiaries' expectations about their future quality of life, with recipients being up to nine percentage points more likely to believe the future will be better than non-beneficiaries (21 vs 30).

- Similar to their expectations about the future, the beneficiaries of the cash transfer program consistently reported a willingness to delay gratification at a higher rate than the comparison group. We find that on average treatment households are as much as 10 percentage points more likely to wait for future money (e.g., money that may become available in one or more months) than households not receiving the cash transfers.

These results suggest that the cash transfer program makes people feel more secure, less desperate, and affects their discount rate and willingness to save.

Implementation Results

We investigate the implementation of the program around four areas: verification of last payment, timeliness and regularity of payments, access to payments, and understanding of program policies among beneficiaries. We find mixed results that the program is being successfully implemented along these measures.

- Verification of Payment: Recipients overwhelmingly report receiving the correct amount of money and at the right time for their most recent payment, with 99 percent of recipients responding accordingly. Thus, there is some evidence that the ministry is able to deliver the proper amount of cash in a timely manner to beneficiaries.

- Timeliness and regularity of payments: according to payment data, the ministry was slow to roll out the program to all beneficiary CWACs with over 70 percent not receiving payments in the first year of implementation. Additionally, the ministry delayed payments over 40 percent of the time, sometimes delaying several consecutive payments, leaveing beneficiaries without any payment for up to six months. These delayed

payments often occurred during the lean season when recipients are most vulnerable due to food shortages.

- Access to Payment: A majority of recipients (70 percent) reported that there travel to pay point locations is very easy or easy. More impressively, over 99 percent of beneficiaries reported that they incur no financial cost to receive their cash payments. These results suggest that the ministry has successfully designed and implemented the cash transfer program in Monze so that beneficiaries can easily access their funds.

- Beneficiaries understanding of the policies of the program regarding the conditions they have to meet. We find that over two-thirds of beneficiaries have a strong understanding of program conditions, demonstrating that the Ministry has educated the people about the program.

Chapter 1: Introduction and Background

This report provides the results of the Monze cash transfer impact evaluation in Zambia's Southern Province. In 2007 Zambia's Ministry of Community Development and Social Services (MCDSS) began implementing the cash transfer program in Monze district and an experimental evaluation design with baseline data collection accompanied the program[2]. AIR was contracted by UNICEF Zambia in 2010 to conduct the follow up data collection, analysis and reporting for the three year impact evaluation of the program. The report contains the findings from AIR's work and is presented in eight sections: background, study design, conceptual framework, analysis, main results, secondary results, implementation results, and conclusion.

Background

In 2007 Zambia's MCDSS started the rollout of a cash transfer program in the Monze district. Zambia had been implementing cash transfer programs since 2004 in three districts, trying different targeting models in each district. The government decided to scale up the Kalomo model to new districts including Monze. This model targets labor constrained and destitute households as defined by the operations manual (MCDSS/SSN 2007). Beneficiary households receive 40,000 or 50,000 kwacha a month (equivalent to $8 or $10 respectively) depending on if the household has children, in which case they receive the higher amount. Payments are made every other month and there are no conditions to receive the money. The purpose of this program is to supplement the income for poor households to ensure that they can eat at least one meal a day, improve diet diversity, and help them access government services such as schools and health clinics.

Targeting

Monze implements a community-based targeting method to identify beneficiary households. Community Welfare Assistance Committees (CWAC) first meet to nominate households in their community that they believe meet the labor constrained or destitute criteria defined in the operations manual. Next, CWAC members collect data on the nominated households and the data are confirmed by the village headman as valid. At a second meeting, all of the nominated households are ranked by their level of destitution and cutoff line is drawn to identify the most destitute 10 percent of the community. At a third meeting the entire list with the cutoff score and identified most destitute is presented to the community for transparency and open debate about the household scores. After the community agrees on the list of identified beneficiary households, the CWAC members submit the list to the District Social Welfare Office (DSWO) where the list further scrutinized by district officers in the presence of CWAC members. The DSWO makes the final decision about household eligibility and determines who will become beneficiaries. An explanation is provided for each beneficiary and rejected household. Finally, households are then notified of their final eligibility status.

[2] The baseline study was conducted by Mazdar and Palm Associates

Baseline Evaluation

UNICEF Zambia contracted Palm Associates and Masdar in 2007 to conduct a baseline analysis of the Monze cash transfer program. These two firms designed a randomized controlled experiment, where randomization occurred at the community (CWAC) level. CWACs were randomly assigned to either the treatment or delayed control group with 65 in the treatment group and 40 in the delayed control group, where households were expected to receive cash transfers three years later. Masdar and Palm associates also designed the baseline instruments that collected household data on demographic information, food expenditures, education, health, wealth, and nutrition including height and weight. The baseline data collection was conducted in July and August of 2007 with a corresponding report submitted to UNICEF in 2008.[3] The first payments to beneficiaries in treatment CWACs began in September and October of 2007.

Follow-up Evaluation

In 2010, UNICEF Zambia contracted AIR to conduct a three- year impact evaluation of the Monze cash transfer program. The delayed control CWACs were still being delayed and had not yet received the program. AIR with Palm Associates conducted the first follow-up round of data collection in July and August of 2010, three years after the baseline data collection and start of the program for treatment CWACs. AIR used the same instrument from the baseline data collection to maintain the longitudinal quality of the study and measure changes over time. Several new sections were added to investigate program implementation, time-value preferences, current quality of life indicators, and expectations of future quality of life. These additional sections only occur in the follow-up instrument and are cross-sectional data instead of longitudinal. These new sections are discussed in the secondary results and implementation chapters of this report.

Data Collection

Similar to the baseline data collection, follow up data were collected at the home of each beneficiary. Enumerators, who are fluent in the local language Tonga, conducted interviews with the female head of household and the named beneficiary if it differed from the female head. In order to maintain consistency between rounds of data collections, height and weight measurements were taken for every household member using the same scales and measuring tapes as used at baseline.

[3] Tembo G. and Freeland N. (2008) Baseline Survey Report for the Monze Social Cash Transfer Programme. UNICEF Zambia. July 2008.

Chapter 2: Study Design

The study design for the Monze impact evaluation changed from a randomized control trial to a design with random assignment at the community level and selection at the household level, requiring a matched comparison group. This design change occurred because the baseline data collection happened in the middle of the community selection process, before the final beneficiaries were identified. This section reviews the original study design and changes that occurred to motivate the final design.

Randomization

The Monze impact study was originally designed as a randomized controlled trial with random assignment at the CWAC level. A randomized controlled trial is the most powerful research design for drawing unambiguous conclusions about the impacts of an intervention on specific outcomes. In an RCT, some subjects are assigned to a treatment group that receives the intervention and others are assigned to a control group, against which comparisons of outcomes can be made. An RCT permits us to directly attribute any observed differences between the intervention and control groups to the intervention program as the result of the random assignment of participants to these groups.[4] Randomization is used to balance the observed and unobserved characteristics that affect the outcomes between the treatment and control conditions of the sample.

On average, households in the randomly assigned treatment and control CWACs looked similar at baseline, indicating that randomization worked to create equivalent groups. Table 2.1 compares the means between households in treatment and control CWACs at baseline for outcomes of interest and characteristics related to them. Only the proportion buying fertilizer is statistically significantly different between treatment and control groups when conducting a t-test to compare proportions, but the difference is less than 0.1 standard deviations and is substantively meaningless. Therefore, the households in treatment CWACs look similar to households in control CWACs. The total sample size of the study is roughly 2,300 households at baseline with close to an even split between treatment and control conditions.

Table 2.1: Mean Differences between Original Treatment and Control Status

	Original Controls		Original Treatment		
	Mean	***Std***	***Mean***	***std***	***P-value***
Total food consumption per month per capita (Kw)	21910.02	21381.89	20974.29	19988.73	0.289
bought fertilizer (1=Yes 2=No)	1.90	0.30	1.92	0.27	**0.035**
DHS style wealth index	0.09	1.05	0.01	0.96	0.080
Head's age (years)	59.40	16.65	59.31	17.00	0.897
Head's education (years)	3.87	3.56	3.85	3.53	0.898
Female Headed Household	0.59	0.49	0.59	0.49	0.792
Household size	5.44	3.12	5.58	3.22	0.307

[4] Campbell, D.T. & Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Houghton Mifflin: Hopewell, N.J.

Dependency ratio	2.43	1.72	2.42	1.67	0.820
# of orphans	2.03	1.86	2.07	1.98	0.634
# of children (0-18)	3.20	2.43	3.23	2.43	0.783
# of adults (19-64)	1.66	1.42	1.74	1.46	0.246
# of seniors (65+)	0.56	0.64	0.60	0.68	0.202
distance to nearest secondary school	27.67	124.67	22.41	19.01	0.172
distance to nearest clinic	5.84	6.29	6.23	5.76	0.126
No toilet	0.57	0.49	0.60	0.49	0.193
Unprotected water source	0.57	0.49	0.60	0.49	0.193
Observations	1072		1145		

P-value is for t-test for statistical difference in means between treatment and control groups (bold indicates significance at 5 percent or less).

Selection in the Treatment Group

Although treatment and control groups were selected to provide equivalence, the baseline data collection occurred before final selection of beneficiaries, making the new identified treatment group look dissimilar to the control group and the original treatment group. This additional round of selection introduced differences between the two groups and threw off the balance that randomization had achieved.

Table 2.2 compares the original control group with the treatment group that was selected after baseline, called the true treatment group. The households in the true treatment group (those in randomly assigned treatment CWACs and selected to receive the program) are poorer and have more orphans. The heads of these households are older, less educated and more likely to be women. This selection process never occurred in the control CWACs. The baseline data had already been collected and AIR was stuck with the sample provided from the baseline survey. Eleven of the variables are statistically significantly different in terms of a t-test of mean differences. The selection process that occurred in the treatment CWACs after baseline created made the two samples unbalanced, introducing selection bias to the original study design. Additionally, the sample size in the true treatment condition has been reduced to around 516 households, less than half the number in the original treatment group.

Table 2.2: Mean Differences between Actual Treatment and Original Control Group

	Original Controls		Actual Treatment Group		
	Mean	***Std***	***Mean***	***std***	***P-value***
Total food consumption per month per capita (Kw)	21924.43	21517.85	21947.36	21031.39	0.984
bought fertilizer (1=Yes, 2=No)	1.89	0.31	1.95	0.22	**0.000**
DHS style wealth index	0.11	1.05	-0.27	0.77	**0.000**
Head's age in years	59.21	16.67	64.43	15.01	**0.000**
Head's years of schooling	3.91	3.57	2.81	3.14	**0.000**
Female headed household	0.59	0.49	0.68	0.47	**0.001**
Household size	5.48	3.12	4.89	2.91	**0.000**
Dependency ratio	2.42	1.72	2.43	1.61	0.969
# of orphans	2.04	1.86	2.09	1.88	0.600

# of children (0-18)	3.22	2.43	2.71	2.17	**0.000**
# of adults (19-64)	1.69	1.43	1.42	1.32	**0.000**
# of seniors (65+)	0.56	0.64	0.75	0.65	**0.000**
distance to nearest secondary school	28.06	126.41	23.74	18.65	0.280
distance to nearest clinic	5.84	6.36	6.34	5.51	0.110
no toilet	0.58	0.49	0.68	0.47	**0.000**
unprotected water source	0.58	0.49	0.68	0.47	**0.000**
Observations	1072		516		

P-value is for t-test for statistical difference in means between treatment and control groups (bold indicates significance at 5 percent or less).

Creating a Matched Comparison Group to Restore Balance at Baseline

Although the balance that resulted from randomization was lost with the selection of beneficiaries after baseline, AIR was able to statistically restore balance by creating a comparison group from the control households that resembles the true treatment group. We selected households from the control group that appear to be most similar to the selected treatment group by using a statistical technique called propensity score matching (PSM) – (Heckman, Ichimura, and Todd 1998). The PSM method estimates the likelihood (propensity) a household is selected for the program based on the characteristics of households that actually were selected to receive the program. Households from the control group are matched to households from the true treatment group by their likelihood to receive the program, creating a comparison group from the control group that best matches the beneficiaries in the treatment CWACs.[5] Table 2.3 contains the mean differences between the true treatment and PSM constructed comparison group. The number of differences and the magnitude of difference in household characteristics between the true treatment and comparison groups greatly reduced after implementing PSM. There are now only two statistically significantly different variables, no toilet and no access to a protected water source, instead of the 11 that resulted from comparing the true treatment group with the control group. Additionally, the magnitude of the differences between the true treatment and comparison groups are not substantively meaningful because they are 0.12 standard deviations difference on these variables. Thus, the PSM method successfully created a comparison group that looks very similar to the true treatment group and removed observed differences that resulted from the selection process in the treatment CWACs that never occurred in the control CWACs. Additionally, we are back to having a balanced sample size between the treatment and comparison groups, with 516 and 508 households respectively. This sample size is less than half of the original sample due to the greatly reduced number of households that were selected to receive the program, but meets requirements that the two groups have similar characteristics.

[5] See Appendix **A** for a technical explanation of PSM and specifics to our analysis.

Table 2.3: Mean Differences between Actual Treatment and Matched Comparison Group

	Matched Comparisons		Actual Treatment Group		
	Mean	***Std***	***Mean***	***std***	***P-value***
Total food consumption per month per capita (Kw)	23126.74	23267.50	21947.36	21031.39	0.396
bought fertilizer (1=Yes, 2=No)	1.93	0.26	1.95	0.22	0.173
DHS style wealth index	-0.31	0.77	-0.27	0.77	0.442
Head's age in years	64.65	15.06	64.43	15.01	0.817
Head's years of schooling	2.73	3.08	2.81	3.14	0.665
Female headed household	0.69	0.46	0.68	0.47	0.817
Household size	4.69	2.64	4.89	2.91	0.267
Dependency ratio	2.43	1.83	2.43	1.61	0.979
# of orphans	2.00	1.82	2.09	1.88	0.461
# of children (0-18)	2.63	2.05	2.71	2.17	0.518
# of adults (19-64)	1.40	1.32	1.42	1.32	0.782
# of seniors (65+)	0.69	0.62	0.75	0.65	0.127
distance to nearest secondary school	33.19	180.30	23.74	18.65	0.240
distance to nearest clinic	6.27	7.09	6.34	5.51	0.860
no toilet	0.61	0.49	0.68	0.47	**0.040**
unprotected water source	0.61	0.49	0.68	0.47	**0.040**
Observations	508		516		

P-value is for t-test for statistical difference in means between treatment and comparison group (bold indicates significance at 5 percent or less).

Final Sample

The sample used in this study that resulted from community selection and matched comparison group contains 1,024 households with 516 in the treatment group and 508 in the matched comparison group. The treatment group represents households that the community believes are the most vulnerable and labor constrained in the area and the matched comparison group looks very similar to them. Table 3 contains the means and standard deviations for demographic and poverty information about the sample. The average head of household is 64.5 years old, has less than three years of education, and is female 70 percent of the time. The average household has 4.7 people, with 2.6 children and 2 orphans. Over 60 percent of the households do not have access to their own toilet (including pit latrine) or a protected water source and live over six kilometers from the nearest clinic.

Chapter 3. Framework, Literature Review, Hypotheses and Outcomes

Framework for Understanding Impacts

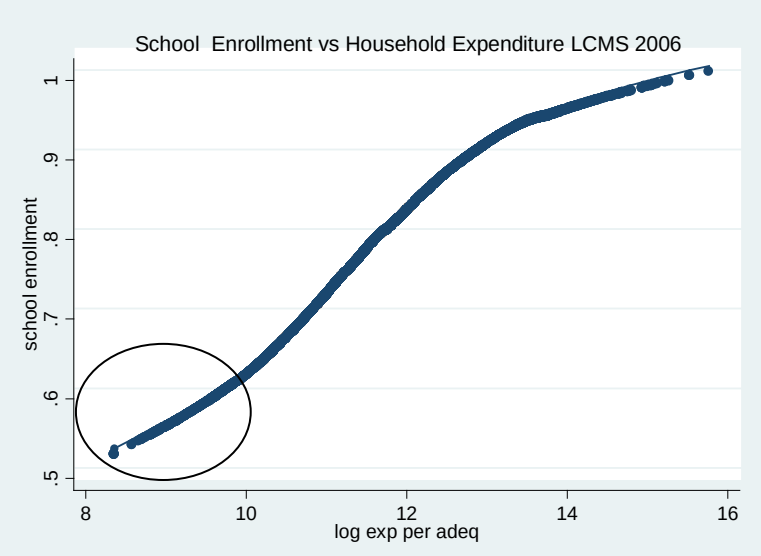
The SCT is an unconditional cash transfer program targeted to ultra-poor, labor-constrained households. Those with higher dependency ratios are given preference in case there are more eligible households than funds can accommodate. Unlike the conditional cash transfer programs common in Latin America (CCTs) which exert both a price and income effect (Handa and Davis 2006), the Zambian SCT will have only an income effect on household demands for consumption goods. Note that we define goods broadly to include services, human resources such as child schooling, health, and nutrition as well as regular commodities that are purchased in the market such as cooking oil and food. We assume that child schooling, health and nutrition include an important consumption component as well as an investment component.

The size of the program's effect on consumption of goods will depend on two factors: the sensitivity of demand to income and the size of the transfer relative to total household income. The greater these two factors, the larger will be the program impact for consumption of a particular good, holding other factors constant. In the SCT Kw100,000 is transferred every two months (Kw80,000 for households without children) and average family size is 5 in our sample. Assuming half of all recipients have children, the mean monthly transfer per capita is Kw 9,000 (45,000/5) or approximately USD1.80 per month or 6 cents per person per day. The official Zambian poverty line is about 85 US cents per day per person but SCT recipients are ultra poor and can be expected to have a daily income of half this figure, or 43 US cents per day. Therefore we estimate that the transfer size is approximately 14 percent (6/43) of household per capita income which is on the lower end of transfer values according to a review in UNICEF (2008). Though there is no hard and fast rule about the optimal transfer size, the successful programs in Latin America transfer at least 20 percent of mean household income to recipients and this number is slowly taking hold in operational contexts as an appropriate figure to aim for to ensure relevant impacts (UNICEF 2008).

A full ex-ante prediction of possible program effects would entail estimating income (or total expenditure) elasticities from baseline data, weighting these by the size of transfer and 'simulating' program responses. Unfortunately the Monze survey instrument does not have a complete non-food expenditure module and even the food expenditure module is small, so it is impossible to estimate expenditure elasticities from the baseline data. However we provide some evidence based on Living Conditions and Monitoring Survey (LCMS) 2006 data on the relationship between total household expenditure (per adult equivalent) and two key indicators that have been found to respond to cash transfers in Africa—school enrollment and food expenditure. The figure immediately below shows the local linear regression relationship between school enrollment for children age 6-16 and total household spending (in expenditure per adult equivalent, ADEQ). The region of interest is at very low levels of spending (the left portion of the graph) which corresponds to the households in the evaluation sample, and we see a

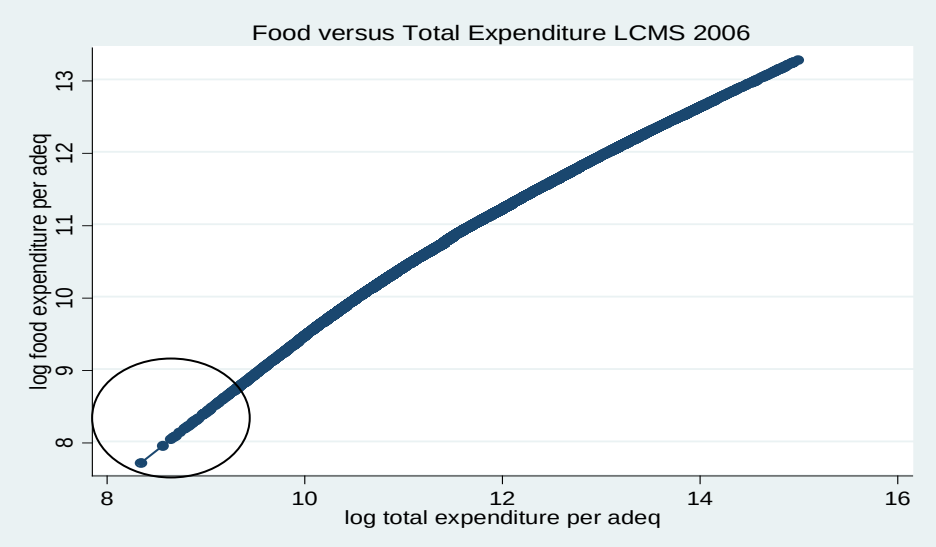
fairly steep slope, indicating that school enrollment is responsive to changes in total expenditure among poor households in Zambia.

Figure 3.1



Below we present a similar type of graph depicting the relationship between food spending and total spending, again from LCMS 2006. Both variables are measured in logarithmic units, therefore the estimated slope of the curve is approximately the food expenditure elasticity (i.e. the percent change in food spending associated with a percent change in total spending). At the lower end of the expenditure distribution (lower left side of graph) the slope of the line is quite steep, indicating that for very poor households, a large share of additional money is spent on food, as we would expect. Thus for both school enrollment and food elasticities with respect to expenditures are large. However this is offset by the total size of the transfer which is relatively low—we might therefore expect a positive but small impact of the program on these two indicators. An additional factor to consider is that 2010, the year of the follow-up survey, was a record agricultural year in Zambia with the largest volume of maize harvested ever in Zambia at 2.7 million tones, a 48 percent increase from the previous year (Sianjalika 2010). Assuming this record productivity also existed in Monze, it could lead to lower impacts of the program on food consumption.

Figure 3.2



Our discussion so far has focused on consumption goods, but there is increasing interest in the potential for cash transfers to contribute directly to economic growth by raising the

productivity of recipient households, for example by allowing them to invest in improved agricultural inputs or open small businesses. The economic theory of the agricultural household predicts that an unconditional cash transfer will have *no impact* on productive activity if labor and credit markets are well-functioning (Handa et al 2010, de Janvry & Sadoulet 1995). This well known result in economics, also known as the separability condition, implies that if indeed a cash transfer has an impact on productive activity, there must be some market failure that the cash transfer alleviates. The obvious candidate for such a market failure is the credit market. Poor rural households are likely to be liquidity constrained and either unable or unwilling to borrow due to lack of collateral, risk aversion, high discount rates, or monopolistic creditors in local geographical areas. All these phenomena are likely to exist in our study setting so that productive impacts of the social cash transfer are theoretically plausible. On the other hand the target group itself is labor constrained and possibly less able to generate productive activity out of the cash transfer.

Brief review of Cash Transfer Impacts in Africa

We summarize impacts from three recent studies of unconditional, government executed cash transfer programs in sub-Saharan Africa (SSA) to provide context for results we find in the this evaluation. These three studies have used relatively rigorous methods such as randomized control trials to adjust for confounders and self-selection: 1) The Kenya CT-OVC evaluation which used a cluster-randomized longitudinal design similar to this study (Ward et al. 2010); 2) the Mchinji (Malawi) SCT evaluation which also used a cluster-randomized longitudinal design (Miller, et al. 2010), and; 3) the South African Child Support Grant which used a longitudinal propensity score matching difference-in-differences design (Samson, et al 2010). All three programs are unconditional, poverty targeted programs with slightly different demographic eligibility criteria.

All three studies demonstrated significant impacts on school enrollment in the range of 4-9 percentage points, and the Kenya study also demonstrated large (13 percentage points) impacts among younger children (less than age 9) indicating an improvement in on-time school entry. Impacts on health are mixed in these studies. There were no significant impacts on health in the Kenya program and a positive impact on curative health visits in Malawi as well as a reduction in morbidity among children in the previous four weeks (health outcomes were not analyzed in the South Africa study). Both the Kenya and Malawi programs demonstrated strong impacts on food expenditure and diet diversity (driven by an increase in the consumption of meat). In South Africa there was an improvement (decline) in self-reported prevalence of hunger of 7 percentage points. The Kenya study did not measure productive outcomes but in Malawi program participants were more likely to hire labor, acquire small farming implements such as hoes and axes, and own livestock (goats and chickens), while in South Africa program households were more likely to continue to engage in agricultural activities. Finally, in terms of child protection, the Kenya program demonstrated strong increases in children with birth certificates and a slight decline in child labor among younger children, while the Malawi program also demonstrated

strong declines (10 percentage points) in the proportion of children engaged in income-earning activities.

This brief review, based on three longitudinal studies from SSA with rigorous approaches to dealing with self-selection and confounders, indicates that unconditional cash transfers can have positive impacts on a range of household and individual outcomes, both on the consumption side (food, diet diversity, schooling, health) and production side (livestock and productive input purchases), as well as on child protection concerns (child labor, birth registration). However equally clear is that effects are context specific, and difficult to find for nutrition and health indicators.

Indicators in this study

In light of the brief review of impacts cited above, we provide estimates of program impacts on a range of individual and household level outcomes. Unfortunately the Monze survey instrument is particularly weak in the areas of expenditure and schooling, two areas where consistent impacts of cash transfer programs have been found in SSA. There are only a handful of highly aggregated questions on food expenditure and *no non-food expenditure module*. There are only three questions on schooling (current enrollment, grade completed and days absent last week) and so impossible to dig deeper into more nuanced schooling responses such as repetition, recent drop-out or returner to school. On the other hand the instrument is quite strong on productive activity and young child health. For example, there is detailed information on crop production, land use, spending on fertilizer and other productive inputs, as well as small tools and livestock.

In terms of health there is a detailed module on children under 5 but the number of children under 5 years old is typically very small in labor-constrained households. In our final estimation sample for example, there are an average of only 0.6 children age 5 or less per household, in contrast to an average of 1.7 children age 6-16 per household. The same demographic pattern is found among program recipients in Kenya, Malawi and Ghana as all these programs ultimately target vulnerability *and* poverty and not poverty per se. It will be difficult to exploit this area of the survey instrument and find statistically significant impacts (even if they existed, which itself is questionable in light of existing evidence) due to low power associated with small sample sizes.[6] Given the target population and experiences from other evaluation studies from the region, a more appropriate instrument would devote more space to expenditures, schooling, and health, and behavioral outcomes appropriate for older children, maintain the agriculture and production modules and reduce the space for young child health.

We provide impact estimates in seven distinct behavioral areas as described below.

[6] AIR was not involved in the original research design and questionnaire development for this evaluation.

1. Food Spending: We begin with an analysis of spending behavior, looking at total food, diet diversity and the composition of food (food shares). There are a few questions on non-food spending (firewood, charcoal, grinding) that we also report.
2. Demographic Composition: We assess whether the composition of the household has changed over time, in terms of the elderly, orphans, and the total dependency ratio since dependency is a key program eligibility criteria.
3. Productive Activity: We take advantage of the strength of the survey instrument and investigate crop production, input purchases, livestock ownership, and small tools accumulation.
4. Outcomes for children 6-16: These include school enrollment, days missed in the reference week, the number of meals eaten in the last day, morbidity in the last 4 weeks, and curative health care visits.
5. Outcomes for children 0-5: We report impact estimates for possession of a health card, well-baby (preventive0 check-ups, morbidity and curative visits. In appendix C we present a detailed analysis of the anthropometric data, which we believe is not suitable for inclusion in the main body of the report because of data quality.
6. Adult physical health: The survey gathers information on activities of daily living (ADLs) which we convert into a score and report as a measure of adult physical health status.
7. Other indicators: A few additional indicators were included in the follow-up survey to measure perceptions about the quality of life and discount rates (the propensity to save or to delay gratification).

Chapter 4: Approach

The statistical approach we take to derive average treatment effects of the SCT is the difference-in-differences (DD) estimator. This entails calculating the change in an indicator (Y) such as food consumption between baseline (prior to program initiation--2007) and post intervention (2010) for treatment and comparison group units, and comparing the magnitude of these changes. Figure 4.1 illustrates how the estimate of differences in differences between treatment (T) and control (C) groups is computed. The top row shows the baseline and post-intervention values of the indicator (Y) and the last cell in that row depicts the change or difference in the value of the outcome for T units. The second row shows the value of the indicator at baseline and post-intervention for comparison group units and the last cell illustrates the change or difference in the value of this indicator over time. The difference between these two differences, shown in the shaded cell in Figure 4.1, is the difference-in-differences or double-difference estimator.

Figure 4.1: The Difference-in-Differences (DD) Estimator

	Baseline (2007)	Post (2010)	1st difference
Treatment (T)	Y^{T}_{2007}	Y^{T}_{2010}	$\Delta Y^{T}=(Y^{T}_{2010}-Y^{T}_{2007})$
Comparison (C)	Y^{C}_{2007}	Y^{C}_{2010}	$\Delta Y^{C}=(Y^{C}_{2010}-Y^{C}_{2007})$
			Difference in differences $DD = (\Delta Y^{T} - \Delta Y^{C})$

The DD is one of the strongest estimators available in the evaluation literature (Shadish, et al. 2002). There are two critical features of this design that are particularly attractive for deriving unbiased program impacts. First, using pre- and post-treatment measures allows us to 'difference' out unmeasured fixed (i.e. time-invariant) characteristics of the family or individual which may affect outcomes, such as motivation, health endowment, mental capacity or unobserved productivity. It also allows us to 'benchmark' the change in the indicator against its value in the absence of treatment. Second, using the change in a control group as a comparison allows us to account for general trends in the value of the outcome. For example if there is a general increase in school enrollment due to expansion of school access, deriving treatment effects based only on the treatment group will confound program impacts on schooling with the general trend increase in schooling.

The key assumption underpinning the DD is that there is no systematic unobserved *time-varying* difference between the T and C groups. For example, if the T group changes its preference for schooling over time while the C group does not, then we would attribute a greater increase in schooling in T to the program rather than to this unobserved time-varying change in characteristic. In practice, the random assignment to T and C, the geographical proximity of the samples and the rather short duration between pre- and post-intervention measurements will make this assumption quite reasonable. In the present study the comparison units are a sub-set of the overall randomized control group. This potentially weakens our ability to make causal

inferences using DD, though the PSM approach within the context of the DD has been shown to perform extremely well at replicating the experimental benchmark in social experiments (Heckman, Todd, and Ichimura – 1998).

When treatment and comparison units are selected randomly and their characteristics are perfectly balanced then simple mean differences as shown in Figure 4.1 are usually sufficient to derive unbiased estimates of program impact. However in large scale social experiments it is typical to estimate the DD in a multivariate framework, controlling for other potential intervening factors that might not be perfectly balanced across T and C units and/or are strong predictors of the outcome (Y). Not only does this allow us to control for possible confounders, it also increases the efficiency of our estimates by reducing the residual variance in the model. Of course there is an important weakness to the multivariate approach, which is that 'over-fitting' the statistical model can 'wash-away' program effects that work through the control variables. For example, if we control for the number of young children in the household when estimating treatment effects on nutrition, and if the program improves nutrition through decreases in fertility (via the well-known child quantity-quality trade-off) then we may not estimate a positive treatment effect when controlling for the number of young children, even though the program actually has an impact on nutrition.

Our approach is twofold. First we present uncontrolled treatment effects, essentially comparing mean difference-in-differences as depicted in Figure 4.1. Second, we estimate treatment effects controlling for a small set of variables that are measured at baseline only, thus minimizing the risk that we are including potential mediators in the model that might soak up true treatment effects. The control variables we include are total household size and the age (in years), education (years completed) and sex of the household head. We emphasize that all these measures are from the baseline data set only. The inclusion of household size is particularly important because the SCT provides essentially a flat transfer so the per-person transfer varies across households of different sizes. We also provide treatment effects separately for small (4 individuals or less) and large households (about 50 percent each of total households) to investigate whether the average treatment effect varies by household size, which would be driven by the difference in the average per capita transfer level.

In the multivariate analysis, the basic setup of the estimation model is shown in equation (1):

$$(1) \quad Y_{it} = \alpha + \beta_1(post)_{it} + \beta_2(T)_{it} + \beta_3(T * post)_{it} + \beta_4 X_{it} + \varepsilon_{it}$$

In this framework 'post' is a dummy (indicator) variable equal to 1 if the observation pertains to the post-intervention period (2010), T is a dummy variable if the observation receives the treatment, and the DD estimate of impact is given by β_3—the interaction between the two variables. The X vector captures control variables described above, and t and i indicate year of

survey and individual observation respectively. The units of observation may be individual children, adolescents, adults or households depending on the outcome. In this framework the coefficient β_2 is a measure of the pre-treatment mean difference in Y between T and C while β_1 measures general changes over time which will be important to control when outcomes are influenced by time trends (such as school enrollment). In the tables we present in the text we only report the coefficient of the DD (T*post) variable—this is the primary parameter of interest, in that it directly estimates the difference in the change in outcome variables over time between the treatment and control group.

Finally, in our analysis we adjust the reported standard errors (and resulting t-statistics) for clustering at the primary sampling unit—the CWAC. There are approximately 90 CWACS in the sample and the assumption is that sampled households from the same CWAC are not necessarily independent, that is, their outcomes may be correlated. This would be due to common contextual effects such as environmental (i.e. agro-ecological) constraints, policies, norms or preferences within CWACS, for example the same agricultural conditions affecting food. The correlation among units within CWACS may result in a dramatic reduction in power to detect statistically significant effects. For example, the Zambian Demographic and Health Survey (ZDHS) reports design effects (DEF) due to intra-cluster correlations ranging from 1.0 to 4.0 across a range of outcomes similar to some of those we estimate here—this effectively reduces the sample size by the amount of the design effect. In other words, the correlation of household measures within a CWAC can have the effect of reducing the effective sample size by as much as 75 percent.

To illustrate how the cluster-based sampling influences the power of the study we provide two examples in Table 4.1. In large-scale cluster-based samples it is usual to assume an average design effect of 2 across all indicators. With 90 clusters and a sample size of approximately 1700 children 6-16 years of age we have 19 children per cluster and our estimated minimum detectable effect size (MDES) is around 0.18 standard deviations. The MDES is the smallest effect in terms of standard deviations that, if found, will be statistically significant given the sample size. Mean school enrollment is 0.72 at baseline with a standard deviation of 0.43 so we will need a change in school enrollment of about 7.7 percentage points for statistical significance (0.18*0.43). This effect size is well within the range estimated for unconditional cash transfer programs in Kenya (Hurrell et al 2010) and South Africa (Samson et al 2010) so it seems plausible that, if the program is being implemented correctly, we can detect a significant impact on school enrollment.

However for household per capita food expenditure where we have a total of 1020 observations and only 11 observations per cluster, our MDES is around 0.23. The baseline mean per capita food expenditure is Kw 22,535 with a (rather large) standard deviation of Kw 22,170 so we require a treatment effect of about Kw 5,100 (Kw 22,170*0.23) for statistical significance. With a monthly transfer of 45,000 (taking the mid-point of the transfer values for households with

children and or those without) and a mean family size of 5, the per capita mean transfer is 9,000 per month. Poor households will typically have an expenditure food share of (at least) 0.60 thus we might expect an increase of (at least) 5,400 in average monthly food spending, which is an effect size of 0.24, just around the MDES. Note that if the intra-cluster correlation was 0 (if each observation was independent) the MDES would be 0.10 or only Kw 2,217.

Table 4.1: Minimum detectable effect sizes (MDES) given existing sample size and cluster-based sampling

Indicator	N	Mean	STD	MDES	Actual change
School enrollment 6-16 (DEF=2)	1700	0.72	0.43	0.18	0.077
School enrollment (6-16 DEF=0)	1700	0.72	0.43	0.14	0.060
PC Food Spending Kw (DEF=2)	1020	22,535	22,170	0.23	5,099
PC Food Spending Kw (DEF=0)	1020	22,535	22,170	0.10	2,217

Chapter 5. Main Impact Results

Spending

Total food expenditure per capita at baseline averages Kw 22,535 per person per month (in all the analysis 2007 values are inflated to 2010 values) or roughly 14 US cents per person per day. Table 5.1 shows DD impact estimates for total food spending, a diet diversity score and food shares for each of 10 different food groups to assess whether the pattern of spending has changed.

Table 5.1: Household Spending

	With Controls		Without Controls	
	Coefficient	R-squared	Coefficient	R-squared
Variable	(1)	(2)	(3)	(4)
Total Food (Kw)	2,133.18	0.005	2,614.42	0.001
	(0.59)		(0.73)	
Diversity Score	0.203	0.028	0.171	0.010
	(1.15)		(0.51)	
Food Shares				
Cereal	-0.011	0.024	-0.013	0.016
	(0.41)		(0.47)	
Tubers	-0.012	0.031	-0.013	0.028
	(1.08)		(1.11)	
Legumes	0.009	0.007	0.011	0.007
	(0.69)		(0.84)	
Meats	0.025	0.018	0.025	0.015
	(1.20)		(1.18)	
Dairy	-0.006	0.012	-0.006	0.004
	(1.15)		(1.14)	
Veg	-0.018	0.003	-0.018	0.001
	(0.73)		(0.76)	
Fruits	-0.001	0.003	-0.001	0.003
	(0.40)		(0.42)	
Sugars	0.007	0.013	0.008	0.009
	(0.97)		(1.04)	
Oils/Fats	0.008	0.018	0.008	0.01
	(1.00)		(0.97)	
Beverages	-0.001	0.004	0.000	0.001
	(0.18)		(0.06)	
Non-Food Spending (Kw)				
Wood, Charcoal, Grinding	152.46	0.016	194.416	0.013
	(0.40)		(0.50)	
Fertilizer*	**0.079**		**0.077**	
	(2.40)		(2.38)	
Land Purchases	-14.178	0.003	-13.506	0.001
	(0.06)		(0.05)	
Land Rent	290.009	0.019	287.073	0.004
	(1.57)		(1.58)	
Observations	2024		2036	

Each row is a separate regression with the dependent variable shown in the first column. Coefficient estimates are difference-in-differences (DD) impact estimate with t-statistics, clustered at CWAC level, in parentheses below. Outcomes with * indicate marginal effects from probit regressions which are used for dichotomous outcome variables; all others are OLS. Statistically significant effects ($p<0.05$) in bold.

The point estimate for total food spending is Kw2,133 in Column 1 which is well below our earlier prediction that a large share of the new money would be spent on food—not surprisingly it is not statistically significant. The impact on the food diversity score, which simply measures the number of food groups purchased, is also not statistically significant. Moreover none of the food share impacts are significant either though the point estimate and t-statistic is highest for meat, which is where we would expect to see an impact (note that spending includes the value of home consumption). We believe that the expenditure module in the questionnaire is simply too aggregated to identify changes in spending.

We carried out additional analysis, not reported in Table 5.1, to better understand the relationship between treatment status and food consumption. First, we separated out total food consumption into cash spending and own production/ gifts to see if the SCT affected the source of food consumption. Results from this analysis show that program households have higher cash purchases of food (by Kw 3194, 22 percent more than the baseline mean) though the difference is still not statistically significant (t=1.41). Second, we investigated whether crop failure or poor rains unevenly affected intervention and comparison households and discovered that while both groups experienced poor rains in equal proportions, intervention households were 8 percentage points more likely to experience crop failure in the year prior to baseline. We then tested whether the SCT had a protective effect on those households that experienced crop failure and found evidence of this. Crop failure prior to baseline reduced mean food consumption by Kw 5203 in 2010 among comparison households but reduced it by only Kw 1222 among intervention households. Thus, the protective effect of the SCT among these households resulted in improved food consumption of Kw 3980, significant at the ten percent level (t=1.75), an important finding.

Finally, to see if the bumper harvest in 2010 was affecting our results we tracked the mean diet diversity score and overall food consumption over time. Overall food consumption did not increase significantly across the sample in 2010 as we would have expected, but the diet diversity score did increase significantly in both intervention and comparison households in 2010 (relative to 2007). While the increase in diet diversity suggests that both groups of households were better off at follow-up and able to diversity their diets, we would have also expected overall food consumption to increase among these very poor households as a result of the very good harvest season.

The lower panel of Table 5.1 shows impact estimates on the non-food spending items included in the survey. Here we see some hints of productive impacts of the SCT. Program households are 8 percentage points more likely to purchase fertilizer, and spend about Kw290 more per person per month on land rental though the latter effect is just outside statistical significance (p=0.12).

Demographic Composition

Table 5.2 presents impact estimates for changes in demographic composition. There appears to have been an increase in the number of orphans in intervention households, particularly male orphans (p=0.06) as well as a decrease in the probability of experiencing a death (3.3 percentage

points) which is significant at 10 percent . Note that these effects are all statistically significant when we do not cluster on CWAC and assume independence of observations, reflecting the fact that clustering effectively reduces sample size

Table 5.2: Impact on Household Demographics

	With Controls		Without Controls	
Variables	Coefficient	R-squared	Coefficient	R-squared
Number of Elders	0.001	0.564	**0.005**	0.002
	(0.72)		(2.13)	
Number of Orphans	0.259	0.081	0.250	0.033
	(1.69)		(1.63)	
Number of Female Orphans	0.256	0.088	0.246	0.032
	(1.65)		(1.59)	
Number of Male Orphans	0.217	0.039	0.214	0.016
	(1.93)		(1.92)	
Dependency Ratio	0.011	0.301	0.012	0.265
	(0.09)		(0.09)	
Any deaths in previous year*	-0.033		-0.028	
	(1.64)		(1.32)	
	2036		2042	

See notes to Table 1 for explanation of table contents.

Productive Activity

We take advantage of the strength of the survey instrument and investigate impacts over a range of productive activities including crop production, input purchases, livestock ownership, and small tools accumulation. We begin by exploring ownership of commonly owned livestock and small tools in Table 5.3. All outcomes are dichotomous so estimation is via probit regression and coefficients reported are actually marginal effects to be interpreted as percentage point changes. We see that over this three year period intervention households are significantly more likely (27 percentage points) to own goats, and about 9 percentage points more likely to own chickens (though only at 10 percent significance level). There is also some indication of greater ownership of pigs (around 4 percentage points) but not statistically significant. On the other hand, the bottom panel of Table 3 shows no significant impact of the program on ownership of small tools, though almost all the point estimates are positive.

Table 5.3: Livestock and Small Tools Ownership

	With Controls	Without Controls	
Variable	Marginal Effects	Marginal Effects	Baseline Mean
Livestock			
Chicken	0.088	0.083	0.580
	(1.88)	(1.76)	
Pig	0.040	0.038	0.062
	(1.2)	(1.16)	
Goat	**0.271**	**0.272**	0.178
	(4.94)	(4.95)	
Cattle	-0.022	-0.024	0.065
	(1.06)	(1.06)	
Small Tools			

Jerrycan	0.001	0.002	0.386
	(0.01)	(0.03)	
Radio	-0.012	-0.009	0.144
	(0.34)	(0.24)	
Machete	0.014	0.014	0.020
	(0.91)	(0.92)	
Pick	-0.015	-0.02	0.054
	(0.89)	(1.09)	
Axe	0.042	0.044	0.731
	(0.93)	(0.99)	
Hoe	0.017	0.018	0.911
	(1.04)	(1.10)	
Oxplough	0.016	0.012	0.073
	(0.60)	(0.45)	
Bicycle	-0.02	-0.021	0.091
	(0.71)	(0.71)	
Observations	2036	2046	

All impacts derived via probit estimation. See notes to Table 1 for explanation.

In Table 5.4 we report estimates for the total number of crops planted, planting any crop, and then planting one of the three most popular crops in the sample—maize, groundnut and sweet potato. The pattern of results indicates no increase in new households planting crops. However those who did plant crops seem to have slightly increased the number of crops grown, reduced their propensity to grow maize and increase cash cropping (groundnut and sweet potato). This pattern is confirmed in column 3, which reports estimates for the quantity of crops produced (kilograms) among those who reported growing anything. Note that the dependent variable here is measured in logs so the coefficients (when multiplied by 100) are interpreted as percentage changes. We see a 50 percent increase in the kilograms of sweet potato produced (significant at 10 percent level), a 30 percent increase in the quantity of groundnut and a 16 percent reduction in the quantity of maize, though the latter two coefficients are not significant. The increase in cash cropping is consistent with the finding above that intervention households derive more of their food consumption from cash purchases rather than own production.

Table 5.4: Crop Production Impacts with Controls

			Households with some positive production only		
	Coefficient	Baseline Mean	Variable (in logs)	Coefficient	Baseline Mean
Variables	(1)	(2)	(3)	(4)	(5)
Number of Crops	0.114	1.301			
	(0.98)				
Planted any Crop*	-0.04	0.762	Total Kgs	0.055	555.41
	(1.01)			(0.45)	
Planted Maize*	**-0.092**	0.726	Maize Kgs	-0.161	460.58
	(2.09)			(0.92)	
Planted Groundnut*	0.041	0.291	Groundnut Kgs	0.300	49.94
	(0.93)			(1.11)	
Planted Sw. Potato*	0.071	0.140	Sweet potato Kgs	0.505	30.95
	(1.42)			(1.75)	
Observations	2048			1524	

Column (1) is estimated on the full sample while Column 2 is estimated only on households who reported any crop production. The dependent variables in column 3 are measured in logs; baseline mean in column 5 is measured in kgs (not logs).

Outcomes for children 6-16

Table 5.5 presents schooling and food security indicators for children of school going age (6-16). There is a significant positive impact on enrollment of 7 percentage points among this age group, and an even stronger (20 percentage points) effect among the youngest children age 6-8 years old, indicating a strong program effect of getting children into school on-time. A similar strong impact on on-time enrollment was found in the Kenya CT-OVC program. Enrollment effects at older ages are also somewhat larger than for the overall sample (9 percentage points) but significant at the 10 percent level only. We find no impact on absenteeism, nor on food security. The latter is measured by whether the child had less than 2 meals in the previous 24 hours (excluding snacks).

Table 5.5: Outcomes Children 6-16

Variable	With Controls Marginal Effects	Without Controls Marginal Effects	Observations	Baseline Mean
Whether Enrolled*	**0.070**	**0.070**	3446	0.787
	(2.47)	(2.54)		
Whether Enrolled Age 6-8*	**0.195**	**0.182**	920	0.501
	(2.28)	(2.17)		
Whether Enrolled Age 15-16*	0.091	0.090	606	0.864
	(1.93)	(1.90)		
Missed 2 or more days in last week*	-0.025	-0.033	1988	0.178
	(-0.46)	(-0.59)		
Had <2 meals in previous 24 hours*	0.042	0.044	3478	0.073
	(1.20)	(1.21)		

See Table 1 for explanation.

Outcomes for children 0-5

Table 5.6 reports selected indicators for children under 60 months at time of survey. Recall that the mean number of children of this age group per household is around 0.60 so there are only about 720 children in the working sample which, along with the intra-cluster correlation mentioned earlier, suggests that these indicators are under-powered for this study, i.e., difficult to detect an effect given sample size . Indeed Table 5.6 identifies no statistically significant program impacts and if anything, several of the point estimates suggest a slightly negative relationship between the program and these outcomes (e.g. possession of a health card and morbidity). See the appendix C for a detailed assessment of the anthropometric data collected in the survey.

Table 5.6: Outcomes for Children 0-5 Years

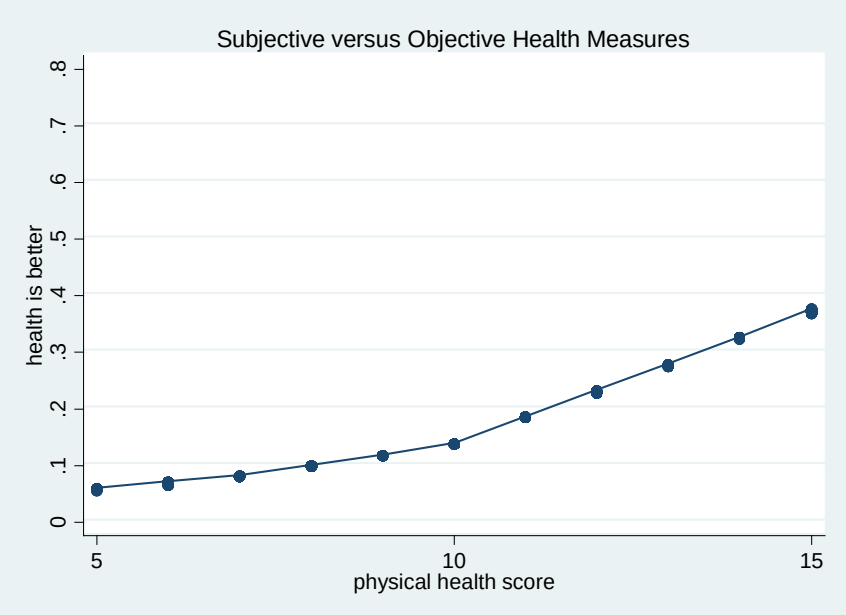
Variables	With Controls Marginal Effects	Without Controls Marginal Effects	Obs	Baseline Mean
Under-5 Health Card*	-0.106	-0.11	508	0.418
	(1.09)	(1.13)		
Well baby checkup*	0.003	0.006	508	0.778
	(0.04)	(0.07)		
Morbidity (last 4 weeks)*	0.106	0.106	586	0.258
	(1.29)	(1.29)		
Curative Care*	0.012	0.022	586	
	(0.16)	(0.29)		0.189
Had <2 meals in previous day*	-0.018	-0.017	574	0.064
	(0.41)	(0.36)		

Se explanatory notes to Table 1.

Adult Health

Adult health is measured through two sets of questions in the survey instrument. First, adults are asked if their health has improved/stayed the same/become worse over the last year. Second, a series of five questions are asked about the ease of accomplishing different activities of daily living (ADLs) (stand-up from a sitting position, draw a pail of water from a well, carry a heavy load for 20 meters, walk 5 kilometers, bend over or squat). The first measure is a subjective measure while the second is more objective. In addition the first measure incorporates psychological health as well as physical health while the second measure focuses on physical health only. We create a physical health score by adding up the responses to the 5 ADL questions (score range is 5 to 15) and code the subjective measure as equal to 1 if health has improved, and 0 otherwise. The figure below plots the relationship between the ADL score and the proportion reporting an improvement in health—there is clearly a correlation between the two measures, especially at larger ADL scores, indicating that physical health is an important component of the subjective health measure, although they do seem to capture slightly different aspects of health at lower levels of ADL scores.

Figure 5.1: Subject versus Object Health Measures



The figure below depicts the relationship between age and ADL score and demonstrates a strong negative relationship between age and physical health as we would expect, and thus indicates some degree of reliability of the score we have created.

Figure 5.2: Physical health Score by Age at Baseline

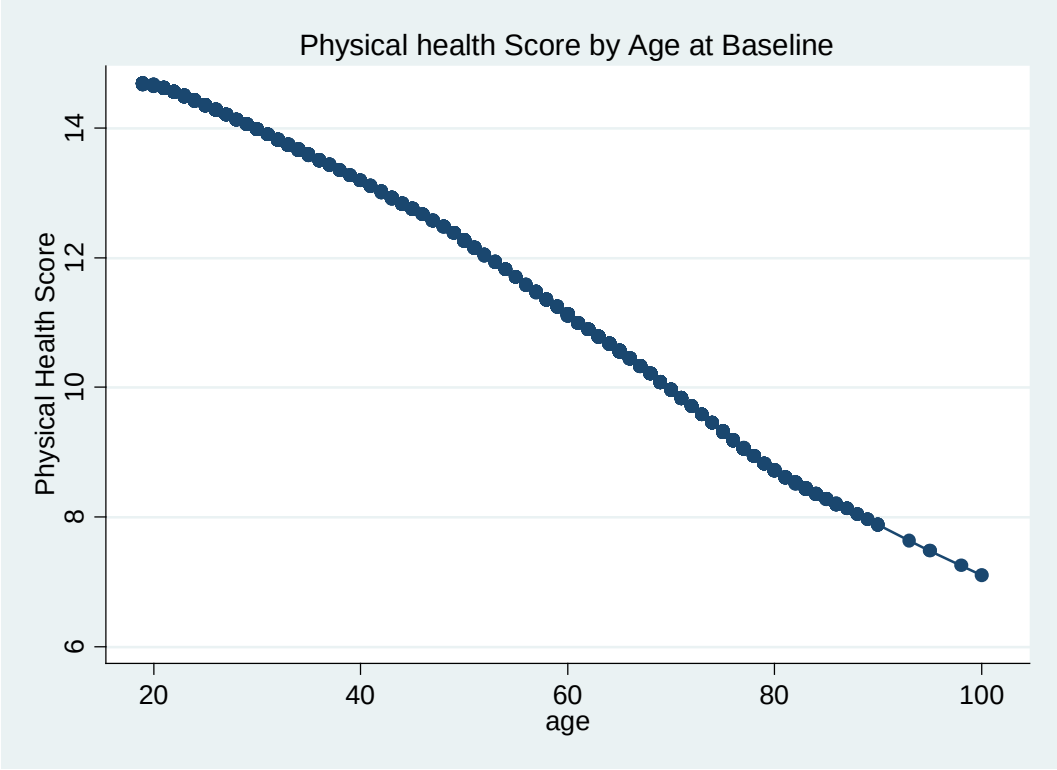


Table 5.7 reports impact estimates of the program on these adult health measures. While there is no impact on subjective health in the full sample of adults (18+), among older adults (65 and

older) there does seem to be a positive impact of the program on subjective health status (significant at 10 percent). For example those 65 and older are 7 percentage points more likely to believe their health has improved over the last year and those 70+ years are 9 percentage points more likely to believe their health has improved. On the other hand there is no impact of the program on physical health status for either the full sample or the older age cohorts, confirming that these two measures capture slightly different dimensions of health.

Table 5.7: Adult Health with Controls

	Coefficient	Obs	Baseline Mean
Health Improved*	0.032	4610	0.167
	(0.94)		
Health Improved Age 65+*	0.070	1418	0.088
	(1.68)		
Health Improved Age 70+*	0.089	1061	0.077
	(1.75)		
Physical health Score	0.091	4557	12.004
	(0.50)		
Physical health Score Age 65+	0.115	1412	9.297
	(0.32)		
Physical health Score Age 70+	0.117	1057	8.986
	(0.32)		

See notes to Table 1.

Impacts by Household Size

We indicated earlier that due to the structure of the program the per capita transfer level varies by household size, with smaller households, especially those with children, effectively receiving a larger transfer on a per person basis. This suggests that program impacts might vary by 'intensity of treatment', that is, there may be larger impacts among smaller households.

Median household size in our working sample is 5 and about half of all households have children (defined as a resident age 18 or younger). To test whether there is a heterogeneous program impact we estimate all impacts separately for small (4 or less) and large (5+) households. Mean household size is 2.62 and 6.86 people for small and large households respectively. Consequently the mean monthly transfer size per person is Kw 17,176 and Kw 6,560 respectively—the 'intensity' of the intervention in small households is almost three times the intensity in large households. Given this pattern in the intensity of treatment we would expect to see larger impacts among smaller households for household level indicators (indicators that are 'shared' across all household members. However for individual level outcomes there may not be enough individuals in smaller households to generate a significant program effect.

Results of estimating treatment effects by household size are reported in Appendix B and briefly summarized here. None of the expenditure results change though the point estimate for total food expenditure is much larger among smaller households at Kw7,575 as is the share devoted to meat. Moreover the significant increase in fertilizer purchases observed in the full sample seems to be driven by larger households rather than smaller households, suggesting that the availability

of labor in larger households may complement the use of other farm inputs in crop production. The marginally significant impact on increased number of male orphans in the full sample is wholly driven by larger households; now these increases are statistically significant for both male, female and total orphans. The point estimate suggests that there was an increase of 0.44 orphans in intervention households relative to comparison households over this period. Note that this is consistent with a marginally significant increase in the dependency ratio among larger households as well. The observed full sample impact on goat ownership is driven by smaller households, who also now have statistically significant greater likelihood of owning pigs. On the other hand larger households appear more likely to own chickens (p=0.07).

The treatment effects on school enrollment also differ by household size. In the full age-range 6-16 the treatment effect is only statistically significant among large households (8 percentage points). However for the younger age group (6-8) the treatment effect is much larger and significant among smaller households (29 percentage points) while among older kids (15-16) the treatment effect is only significant among larger households (13 percentage points).

Chapter 6: Secondary Results

This section presents and discusses the findings for three modules that AIR added to the follow-up survey that were not part of the baseline; these additions are: expectations about the future, discount rates, and quality of life. To our knowledge this is the first time that these outcomes have been investigated in a cash transfer evaluation. All three outcomes have an extensive literature in economics and psychology that link them to important behaviors such as investing, saving, avoiding risky behaviors, and household planning.

Overall, there is strong evidence that the Monze cash transfer program has positive effects across these indicators, suggesting that the program leads to important behavior changes that improve beneficiaries' economic and health outlook. We provide a brief motivation for investigating these outcomes in the context of the SCT.

A household's or individual's willingness to invest in the future is affected by expectations about the future and time preference or discounting (how heavily the future is regarded realtive to the present). For example, widespread prime-age mortality due to HIV/AIDS in Eastern and Southern Africa has lead to declining life expectancies and increased uncertainty about the quality of life in the future (UNAIDS 2008). Researchers have cited these factors as affecting individuals' propensity to invest in the future, either through financial or asset savings, or human capital accumulation in themselves or their family members. The economic concept of time preference or 'patience'—the importance an individual places on immediate versus delayed consumption—plays a key role in economic models of savings, growth and economic development (Rae, 1834; Samuelson 1937; Smith 1776). Indeed many development economists view this parameter as a major determinant of household poverty, citing 'myopia' as the reason why many households do not invest in the means of production (human, physical or financial) which could significantly alter their future standard of living (Bardhan and Udry, 1999). While economists have long recognized the importance of expectations about the future for current decision-making, this concept has also recently been recognized by social epidemiologists (who refer to it as 'hope') as important in understanding and predicting the behavioral response to health-related risk (Barnett & Weston 2008).

Cash transfer payments can interact with future expectations and discount rates in two ways. First, steady and predictable transfers can alter a household's expectations about the future quality of life and thus make it more interested and oriented toward investing in the future or engaging in less risky behavior. Second, a predictable source of income can alter individual discount rates, making recipients less myopic and more forward looking, and more willing to delay current consumption for future investment such as schooling of children, or health and economic activities which provide benefits in the future rather than the present. In this framework expectations and discount rates are conceived as mediators between the program and investment and consumption activity. However they can also be important moderators of program effects—holding other things constant the cash transfer program ought to have a larger impact on school enrollment or productive investment the lower a household's discount rate and the higher its (positive) expectations about the future.

In the follow-up survey, we measured expectations about the quality of life at one, two, five and 10 years in the future. We measure time preference through a series of hypothetical questions about whether the respondent would choose to receive Kw 200,000 today or a larger sum of money in one month. The amount of 'future money' is varied and ranges from Kw 200,000 (i.e. the same amount as the amount available today) up to Kw600,000. Hypothetical questions such as these have been used with success by development economists in field studies in similar study settings (Cardenas & Carpenter 2008; Anglewicz et al. 2009). Respondents who select the second option, waiting one month, for smaller amounts of money demonstrate a greater willingness to delay than those who require a higher delayed amount. To ensure that respondents understand the question we ask a further question about whether they would accept less money in the future (Kw 220,000 today versus Kw 200,000 in one month). Fewer than 10 percent of respondents chose the option of waiting for less money. Further analysis shows that these individuals always chose the option of waiting, so we exclude these observations from the results reported here on the grounds that these individuals likely did not understand the question.

Recall that these outcomes were not part of the baseline instrument, thus we only have data for treatment and comparison groups at the three year follow up round of data collection. All of the outcomes in this section were tested using regression analysis while controlling for the gender of the household head, the household size, the age of household head, and the household head's education level. In the case of a binary outcome, we ran probit regressions instead of OLS. Point estimates, t-statistics, and sample sizes (n) are presented for each test.

Expectations about the future

The beneficiaries of the cash transfer program consistently reported at a higher rate than the comparison group that their life will be better in the future. Heads of household were asked a series of questions about their belief of the quality of their life at four time points in the future: one year, two years, five years, and ten years. The response options were better, same, worse, or

I don't know. Figure 6.1 provides the proportion of households by treatment status that believed their life will be better at different time points in the future. The proportion who believes life will be better in the future decreases as they consider more distant time periods, regardless of their treatment condition. This decrease represents the uncertainty that comes with considering a longer period of time. However, a greater proportion of households receiving cash transfers consistently believe that their life will be better than households that do not receive the cash transfers at all time points. The difference in probability of believing the future will be better derived from probit regressions, are shown in Table 6.1. For 1-2 years in the future program participants are 9 percentage points more likely to believe the future will be better relative to households in the comparison group; for longer periods in the future (5-10 years) recipients are about 5 percentage points more likely to believe their life will be better.

Figure 6.1 Expectation of Quality of Life

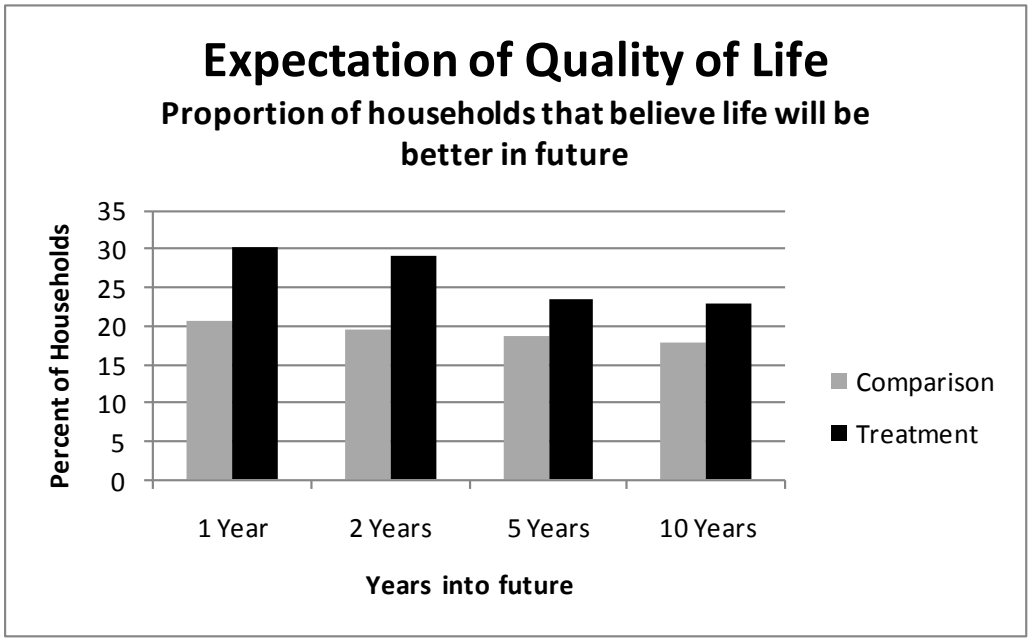


Table 6.1: Program Impact on Whether Household Believes Life will be Better in the Future

Years in future:	1 Year	2 Years	5 Years	10 Years
Marginal Effect	0.096	0.093	0.050	0.051
clustered t-statistic	(2.45)	(2.35)	(1.29)	(1.27)
unclustered t-statistic	(3.54)	(3.49)	(1.98)	(2.07)
Mean	0.256	0.244	0.212	0.206
Observations		1014		

Marginal effects (percentage point differences) derived from probit regression. Clustered and unclustered t-statistics reported in parentheses. A marginal effect with a t-statistic greater than 1.96 is statistically significant at the 5 percent alpha level. Data from follow-up survey only.

Willingness to Delay

Similar to their expectations about the future, the beneficiaries of the cash transfer program consistently reported a willingness to delay receiving money at a higher rate than the comparison group. Heads of household were asked a series of hypothetical questions about their preference to receive 200,000 kwacha today (equivalent to $40 in 2010) or a higher amount one month from now, with the delayed amount increasing with each question. Figure 6.2 depicts the proportion of households that prefer to wait one month to receive the amount listed. At 200,000 kwacha, the treatment and comparison groups are the same with only two percent willing to delay a month. This result seems logical because the delayed amount is the same as the initial offer, so there is little or no benefit to delaying. As the delayed offer increases, so does the proportion of households in each group (treatment and comparison) willing to wait a month for the larger amount of money, an indication that respondents seem to understand the question. However, the cash transfer beneficiaries have a greater proportion of households willing to wait for the delayed offer, at every amount. This suggests that the cash transfer program changes households' time preferences by increasing their willingness to delay gratification.

Figure 6.2 Time preferences for money for treatment and comparison groups.

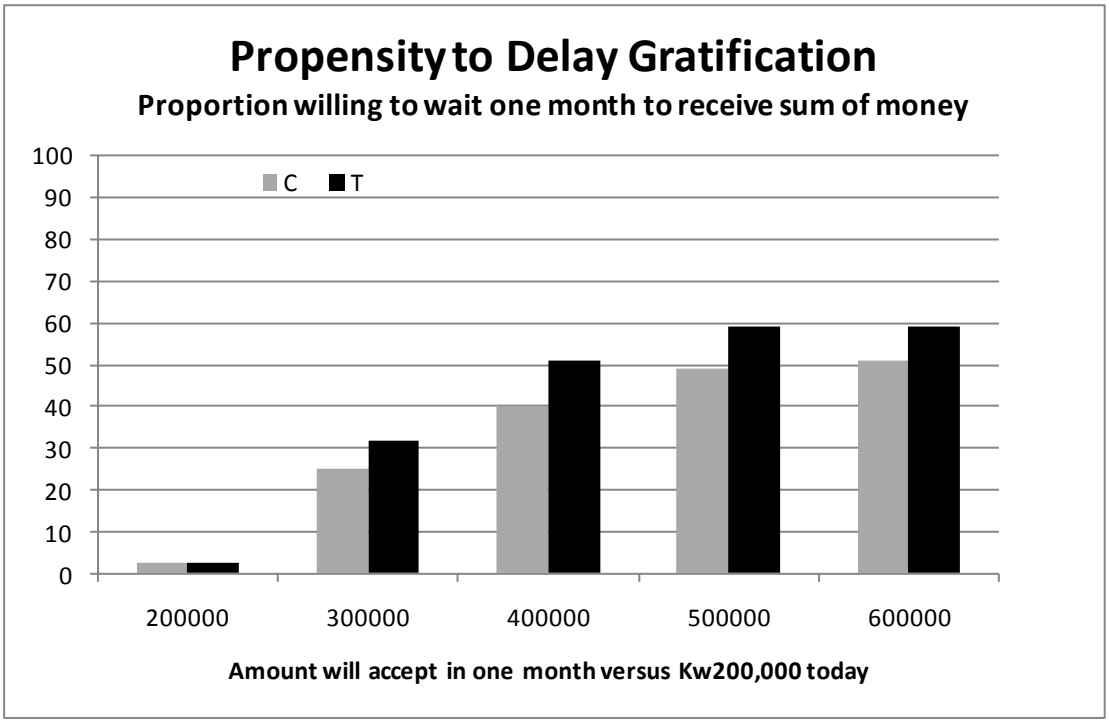


Table 6.2 provides estimates of the marginal effects from probit regressions of the likelihood to wait for a future sum of money. At kw 300,000 recipients are 6 percentage points more likely to wait; the largest difference is at Kw 400,000 where respondents in treatment households are 10 percentage points more likely to wait for future money. These effects are not statistically significant when standard errors are clustered, but become strongly significant when standard errors are not clustered, indicating the strong design effect at work in these findings.

Table 6.2: Program Impact on Propensity to Delay Receiving Kw 200,000 Money for One Month

Amount in 1 Month:	200,000	300,000	400,000	500,000	600,000
Marginal Effect	-0.005	0.060	0.101	0.084	0.072
clustered t-statistic	(-0.37)	(1.26)	(1.57)	(1.22)	(1.01)
unclustered t-statistic	(0.50)	(2.01)	(3.01)	(2.49)	(2.12)
Mean	0.026	0.288	0.458	0.539	0.552
Observations			910		

Marginal effects (percentage point differences) derived from probit regression. Clustered and unclustered t-statistics reported in parentheses. A marginal effect with a t-statistic greater than 1.96 is statistically significant at the 5 percent alpha level. Data from follow-up survey only.

Quality of Life

We asked respondents to rate their current satisfaction with their life on a 1 (very unsatisfied) to 5 (very satisfied) scale. The modal response was 3 (neither satisfied nor unsatisfied--35 percent) and 17 percent of the sample responded with a 4 or 5 indicating they were either satisfied or very satisfied with their life. We coded this variable 1 for responses 4 and 5 and 0 otherwise and estimated a probit regression for the probability of being satisfied or very satisfied with life. The marginal treatment effect indicated that respondents in treatment households were 4.2 percentage points more likely to be satisfied or very satisfied with their life ($p = 0.16$).

Chapter 7: Success of Implementation from Beneficiary Perspective

Zambia's MCDSS had been implementing the cash transfer program in Monze for three years by the time AIR conducted the follow up round of data collection. We used this opportunity to add a module to the survey that investigates the fidelity of program implementation from the beneficiaries' perspective. This section discusses the results of the implementation questions. We focus on four areas: verification of last payment, timeliness and regularity of payments, access to payments, and understanding of program policies. Verification refers to analyses that investigate if beneficiaries receive the proper payment amount based on their last payment . Timeliness and regularity of payments are investigated by looking at ministry records regarding when CWACs received each of their payments. We examine access by looking into the difficulty and cost incurred to beneficiaries for them to receive payments. Last, we ascertain beneficiaries understanding of program policies through questions about aspects of the program that they should be aware of such as conditions for receiving payment.

There are mixed results regarding the ministry's ability to successfully implement the cash transfer program. Beneficiaries report receiving the right amount of money for the last payment, can access the money without any cost, with relative ease, and have a good understanding of program policies. However, payments to beneficiaries were frequently delayed, sometimes for extended periods of time, and occurred more often during lean periods of the year when beneficiaries need them most. Data and analyses are presented through descriptive statistics due to the cross sectional nature of the data and lack of a comparison group. There are 559 households in the sample spread across 62 CWACs.

Verification of payments

Beneficiaries accurately corroborated the amount and timing of their last payment. AIR included several questions on the follow up survey specifically targeting beneficiaries to verify that they receive their money and the proper amount. Seventy-four percent of households report receiving their last payment in either June or July, with another 20 percent of households reporting their last payment in January or February, and the remaining six percent spread between March, April, and May. The ministry makes payments on a bimonthly basis, paying for two months at once. The household reports of last payment match the ministry's records for the payment schedule. Thus, it appears that all of the beneficiaries received their payments on time.

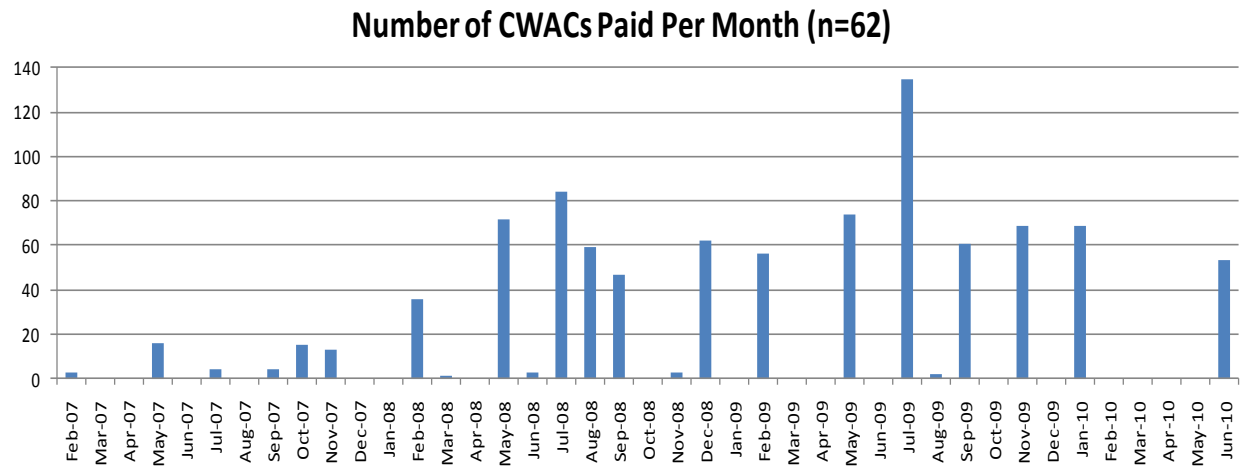
Perhaps more importantly, beneficiaries report receiving the proper amount of funds, suggesting that program implementers are not skimming or stealing beneficiaries' money. The Monze cash transfer program pays 80,000 kwacha every two months to households without children and 100,000 kwacha to households with children. When asked about the amount of money they receive, 83 percent of households report receiving 100,000 kwacha every two months, 16 percent report receiving 80,000 kwacha every two months, and less than one percent report that they do not know how much they receive. These responses support the claim that the ministry is

providing the proper amount of funds to beneficiaries without payment people skimming off the top.

Timeliness and regularity of payments

We reviewed administrative data from the ministry that documents each payment made over the three year period to each CWAC to assess the timeliness and regularity of payments to beneficiaries. Four concerns arise from the data: 1. the ministry was slow to roll out the program to all beneficiary CWACs with over 70 percent not receiving payments in the first year of implementation (February 2007 to February 2008). 2. the ministry delayed payments to beneficiaries over 40 percent of the time, sometimes delaying several consecutive payments, leaveing beneficiaries without any payment for up to six months. 3. The delays in payments often occurred during the lean season (January to May) when recipients are most vulnerable due to food shortages. 4. The greatest number of missed payments in a row occurred in the months immediately prior to the follow-up date collection in July 2010. This delay means that the follow-up data collection did not collect expenditure data reflective of how a family would spend if they had been receiving payments consistently. Figure 7.1 depicts the number of CWACs paid in a particular month during the study period. The ministry is supposed to pay CWACs every other month; thus, there should be a bar reaching to 62 on the y axis for every other month. Sometimes CWACs were paid double or triple the amount in one payment to compensate for missed previous payments, this cannot be seen by the figure. Other times, for example July 2009, CWACs were paid twice during the month to compensate for missed payments. These data suggest that the ministry did not consistently make payments on time, potentially reducing the impact of the cash transfer program on outcomes of interest.

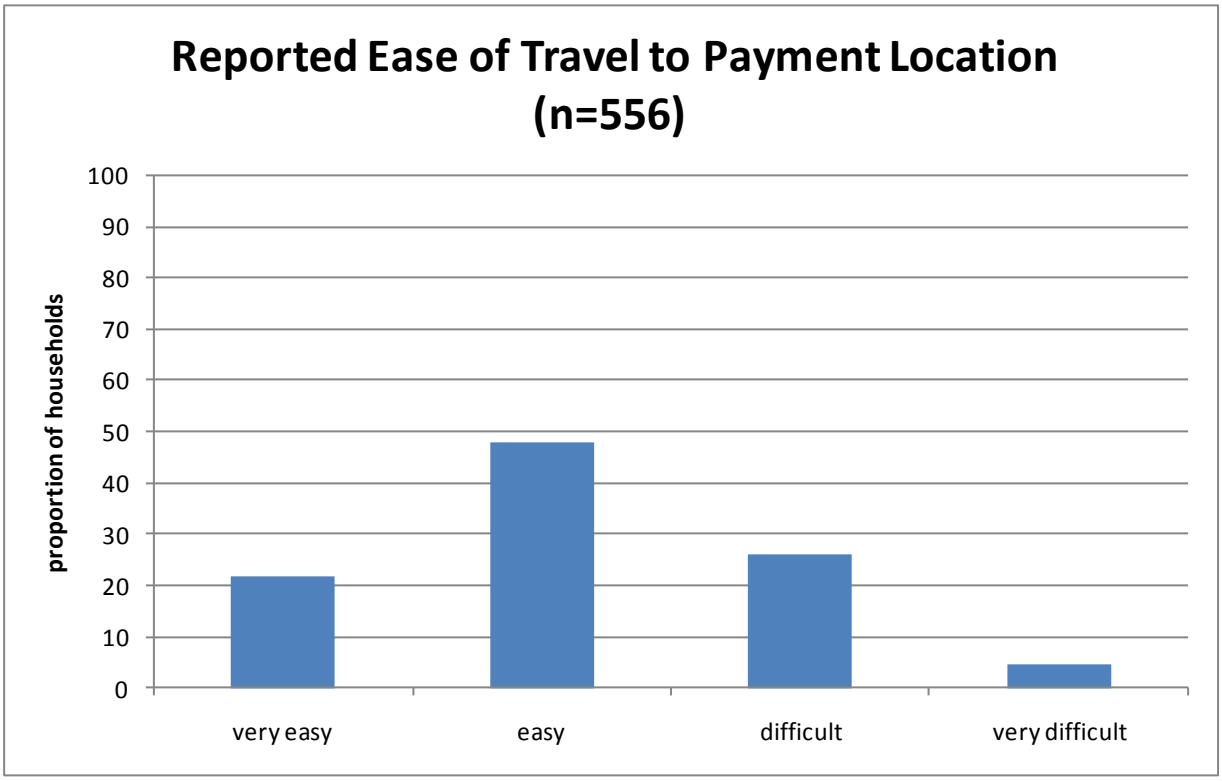
Figure 7.1 Number of CWACs paid per month



Access to payments

We investigated beneficiaries' access to funds through two questions about ease of travel to the pay point location and the expense incurred to reach the payment location. Figure 7.2 depicts the proportion of responses by beneficiaries when asked about the ease of their travel to the pay point location where they receive their funds. Over two-thirds of the households (70 percent) reported that the travel to receive their payment was either very easy or easy. Twenty-six percent stated that their travel was difficult, while only four percent said their travel was very difficult. In addition to inquiring about ease of travel, we also asked about the expense they incur from travel to collect their payment. Of the 551 beneficiaries who responded to this question, 549 stated that they incur zero cost to collect their payments and two people said it cost them 5,000 kwacha. These data suggest that the ministry has implemented the cash transfer program in Monze so that recipients incur no cost and have an easy travel experience to acquire their funds.

Figure 7.2: Reported Ease of Travel to Payment Location



Understanding of Policies

Beneficiaries demonstrated an understanding of the policies for the cash transfer program. This knowledge is important because it affects their expectations and behavior. AIR asked households if there are any conditions to receiving the money? The Monze cash transfer program implemented soft conditions on portion of their households. Soft conditions mean that the beneficiaries were told at the beginning that they had to meet certain conditions to remain in the program, but these conditions were never reinforced. Although the knowledge of their condition does not ultimately matter since they will receive the program anyway, it is an interesting measure to see how well the ministry was able to convey information to the beneficiaries. Table 7.1 contains the self reported condition from beneficiaries by their actual condition status. Over two-thirds of beneficiaries properly know if they have conditions or not.

Table 7.1 Proportion of Beneficiaries who Self Report Having a Condition by their actual condition status (n=549)

	Self Report Condition	Self Report No Condition
Condition	66.67	33.33
No Condtion	77.68	22.32

Chapter 8: Conclusion

In this chapter we review the findings from the Monze cash transfer impact evaluation, discuss the limitations to this study, and suggest future areas of research. Our analyses investigates three-year effects on a range of outcomes that cut across several sectors including nutrition, health, education, labor, and agriculture. In addition to these primary outcomes, we examine the program's psychological impacts on expectations of the future, discount rates, and self assessed quality of life. Last, we consider the quality of program implementation by the MCDSS, the ministry administering the cash transfer program in Monze.

Review of Main Results

We find mixed results for impacts on primary outcomes.

- On the production side we find strong impacts on livestock ownership, particularly goats and chicken, and among smaller households, pig ownership for beneficiaries. Program households are more likely to purchase fertilizer and to produce a greater quantity of cash crops; there also appears to be a shift away from maize for direct consumption and towards more cash cropping (groundnut, sweet potato) for sale. The erratic schedule of payments by the ministry to beneficiaries could be one contributing factor to this finding as beneficiaries would receive several payments at one time, enabling them to make investments that might not otherwise be possible if the payments were smaller and more regular as was intended.

- We find strong impacts on school enrollment, in a similar range to other programs (seven percentage points), and very strong impacts on enrollment of younger children (20 percentage points) indicating that the program has an effect on on-time school entry.

- We find no impacts on food expenditures or food composition. We believe this is because the expenditure module, which only covers food, is missing important items and is not sensitive enough to capture changes in food expenditure, especially at such low levels of spending. Additionally, the delays in payments to beneficiaries, especially in the months prior to the follow-up data collection would affect their spending in the month prior to data collection, the expenditure period assessed in the follow-up instrument.

- There are no statistically significant impacts on health outcomes such as having an under five card, attending checkups, and curative care for either young children (age five and under) or school-age children, which is consistent with findings from the Kenya study. For young children, the sample size is extremely small (720) and the study therefore lacks sufficient power to detect effects among this group given its size, even if they were to exist.

These results suggest that the program impacts economic production and investment in education, but that these impacts do not necessarily carry through to nutrition and health outcomes. The frequent and long delays in payment of funds to beneficiaries might explain these findings.

Secondary Results

Although cash transfer programs primarily focus on affecting expenditures at the time transfers are made they can potentially affect attitudes and expectations in a way that influence future behavior. We find interesting results on secondary outcomes related to expectations of future quality of life and preferences for delayed gratification that, as far as we know, have never been tested before in a cash transfer evaluation. These outcomes are linked to important behavior change for investing, saving, and avoiding unnecessary risk.

- We find a strong impact on beneficiaries' expectations about their future quality of life, with recipients being up to nine percentage points more likely to believe the future will be better than non-beneficiaries (21 vs 30).

- Similar to their expectations about the future, the beneficiaries of the cash transfer program consistently reported a willingness to delay gratification at a higher rate than the comparison group. We find that on average treatment households are as much as 10 percentage points more likely to wait for future money (e.g., money that may become available in one or more months) than households not receiving the cash transfers.

These results suggest that the cash transfer program makes people feel more secure, less desperate, and affects their discount rate and willingness to save. The observed impacts could be due to actual changes in discount rates among intervention households, or they may simply reflect optimism on the part of the 'lucky participants in the SCT' that future money will actually become available (though the questions were hypothetical). From a policy perspective however, the root cause of this behavioral change may not be important.

Implementation Results

We investigate the implementation of the program around three areas: verification of last payment, timeliness and regularity of payments, access to payments, and understanding of program policies among beneficiaries. We find mixed results that the program is being successfully implemented along these measures.

- Verification of Payment: Recipients overwhelmingly report receiving the correct amount of money and at the right time for their most recent payment, with 99 percent of recipients responding accordingly. Thus, there is some evidence that the ministry is able to deliver the proper amount of cash in a timely manner to beneficiaries.

- Timeliness and regularity of payments: according to payment data, the ministry was slow to roll out the program to all beneficiary CWACs with over 70 percent not receiving payments in the first year of implementation. Additionally, the ministry delayed payments over 40 percent of the time, sometimes delaying several consecutive payments, leaveing beneficiaries without any payment for up to six months. These delayed payments often occurred during the lean season when recipients are most vulnerable due to food shortages.

- Access to Payment: A majority of recipients (70 percent) reported that there travel to pay point locations is very easy or easy. More impressively, over 99 percent of beneficiaries reported that they incur no financial cost to receive their cash payments. These results suggest that the ministry has successfully designed and implemented the cash transfer program in Monze so that beneficiaries can easily access their funds.

- Beneficiaries understanding of the program's policies: We find that over two-thirds of beneficiaries have a strong understanding of program conditions, demonstrating that the Ministry has educated the people about the program.

Limitations

There are a number of limitations in this study that affect our ability to estimate impacts and challenge the validity of results. Although we acknowledge these limitations, we have taken steps to address them through a strong methodological design and believe they pose little threat to the internal validity of our findings. Perhaps the greatest limitation to this study is the small sample size that resulted from the baseline occurring before final selection of beneficiaries in each CWAC. Initial sample selection for the study was based on a preliminary list of potentially eligible households. Only about half of these were actually selected for inclusion in the program, drastically reducing the effective sample size of the study from that reported in the baseline data collection. Thus, this study is underpowered for estimating impacts on many of the outcomes of interests. We were unable to determine this reduced sample size and resulting limited power until conducting the follow up study and learning who actually benefitted from the program and who did not. We indicate in the study where we believe the reduced sample size might have affected our ability to find statistically significant results.

We believe that data collection instrument is not ideal for measuring impacts for this study. The survey instrument itself is strong in measuring productive activity and weak in expenditure. It is strong in young child health but there are very few children 0-5 because of the nature of the target population; it is weak in measuring health of everyone else in the household. AIR did not conduct the baseline data collection nor did we design the survey; however, we had to use the same survey from baseline (with some additions) in order to maintain the benefits of longitudinal data and measure change from baseline to follow up.

It was also necessary to employ a quasi-experimental approach for defining a comparison group to measure treatment effects due to the selection process that occurred in the treatment CWACs but no in the control CWACs. This situation resulted because the baseline data collection happened before the final selection of beneficiaries in either treatment or control CWACs. The need to model selection in control CWACs weakens our ability to make causal inferences because we cannot account for unobserved differences between treatment and control samples. We implement a PSM approach within the context of the DD, which has been shown to perform extremely well at replicating the experimental benchmark in social experiments (Heckman, Todd, Ichimura). Therefore, we believe that we can isolate the effects of the cash transfer program on beneficiaries and have a strong argument for making causal claims.

Future Research

We find a number of exciting results in this study that demonstrate in one area the potential of cash transfer programs to help the extremely poor, vulnerable, and labor constrained. However, the reduced sample size limited the power of the study to estimate impacts on all outcomes of interest. The positive point estimates for some results that are not statistically significant suggest that they might also be impacted from the program, but our insufficient sample size prevents us from making assertions of definite impacts. Additionally, the data collection instrument is too thin to properly measure some indicators such as health and expenditure. Therefore, we suggest pursuing another study with a greater sample size and expanded instrument to further investigate these potential impacts.

As far as we know, this study represents the first time anyone has investigated the effects of cash transfer programs for psychological indicators around future expectations, quality of life, and time preferences. We included these measures in the follow up instrument and found several significant and interesting results. Next steps would be to expand this section of the instrument to more deeply investigate these phenomena and include them on both the baseline and follow up instruments to measure change over time.

Last, our findings about program implementation suggest that the MCDSS has successfully delivered the cash transfers in a challenging environment. We suggest following up this preliminary investigation with qualitative research such as interviews and focus groups to learn how they achieved such success, what they learned, and how they would improve upon their delivery.

Appendix A: Description of Propensity Score Matching

The original sample selection for the Monze evaluation was based on the initial listing of households from the first stage of the targeting process. However not all households on the initial listing were eventually selected to participate in the program. Indeed only about half of the households on the initial list reported being in the program according to the follow-up household survey. As shown in Section 2 of the main text, these households are different from the full list of households (smaller households with more elderly, and with heads that have less education and are more likely to be female). The selected households therefore also have slightly different characteristics from the full sample of control households so we employ propensity score matching (PSM) to select, from among the full control group, a group of households that is most similar to the households actually selected to receive the program. The PSM technique proceeds in two stages. First, for each household in the truly treated and control samples, a propensity or balancing score is constructed which indicates the likelihood of receiving treatment (the higher the score, the greater the likelihood). This score is predicted from the coefficient estimates of a multivariate regression that contains variables that are thought to be associated with selection into treatment. Table A1 shows results of this estimation, and indicates that age of head (positive), household wealth score (negative), orphans (positive)and distance to clinic (farther away, more likely of selection) are the most important characteristics associated with selection into treatment status. The overall fit of the regression is low due to the fact that these households are generally quite similar, thus making it difficult to precisely predict selection.

Table A1: Logit Estimates of Odds of Receiving treatment

```
Logistic regression                                     Number of obs   =       1544
                                                        LR chi2(10)     =      91.09
                                                        Prob > chi2     =     0.0000
Log likelihood = -932.54756                             Pseudo R2       =     0.0466

------------------------------------------------------------------------------
        true |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      wealth |  -.4142414   .0961207    -4.31   0.000    -.6026345   -.2258483
      lnsize |   .0136932   .1448765     0.09   0.925    -.2702595    .2976458
      lnfood |  -.0934523   .0717227    -1.30   0.193    -.2340262    .0471215
    demo_dep |  -.0536663   .0432201    -1.24   0.214    -.1383762    .0310437
         fhh |   -.073021    .139713    -0.52   0.601    -.3468536    .2008115
    head_age |   .0138464   .0044974     3.08   0.002     .0050316    .0226612
   hhorphans |    .091899   .0478315     1.92   0.055    -.0018489    .1856469
    head_edu |  -.0245427   .0220237    -1.11   0.265    -.0677083    .0186229
 lnssch_time |  -.0259855   .0528666    -0.49   0.623    -.1296021    .0776312
lnclinic_t~e |   .1941962   .0654338     2.97   0.003     .0659484    .3224441
       _cons |  -1.323803   .9042987    -1.46   0.143    -3.096196    .4485901
------------------------------------------------------------------------------
```

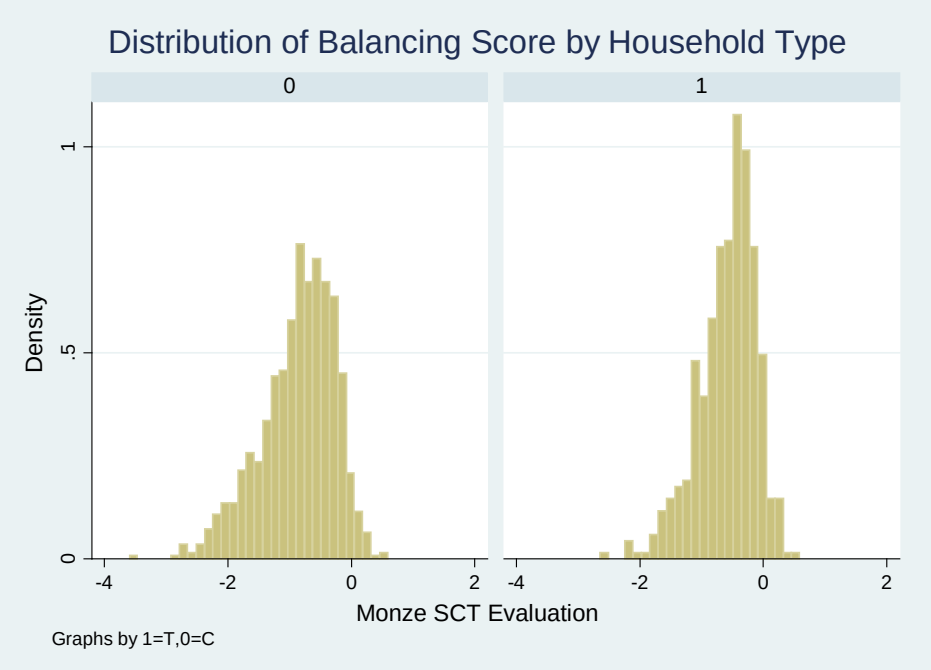
Table A2 shows the distribution of values of the generated log of the odds ratios derived from the logistic regression (the propensity or balancing scores). AS we would expect, the median value is higher for the truly treated group (-0.499) as is the lower bound and the 25th and 75th percentile values, though the upper values are roughly the same for both samples. Figure A1 shows full densities for the scores by sample. The density for the truly treated is clearly centered to the right of the density for the control group with a mass just below 0; the control group in contrast has a much thicker left tail. Despite these differences there is sufficient overlap (or

common support) between the distributions to identify good matches—households in the control group with identical or similar scores to those in the truly treated group. Indeed there are only 8 treated households with scores that are 'off the common support'—that is, with scores that are higher than the highest score of any control households. In our application of PSM we impose one-to-one nearest neighbor matching without replacement. That is, for each truly treated household, we choose the control household with the closest propensity score as its 'match' and we allow control households to only be matched once (they cannot be replaced in the pool of potential matches). This technique allows us to generate a matched comparison group of equal size as the truly treated group. We then use this matched comparison group in our analysis of treatment effects.

Table A2: Distribution of Log Odds Ratios by Sample

1=T,0=C	min(lo)	p25(lo)	med(lo)	p75(lo)	max(lo)
0	-3.6116376	-1.2196105	-.77836585	-.42142292	.56103142
1	-2.6467134	-.83460166	-.49883724	-.25666099	.58751872

Figure A1: Distribution of Balancing Score by Household Type



PSM Results for Spillover Effects

Our analysis of spillover follows the same approach as described above. In this scenario, the 'treated' households are those who were initially listed but later not-selected into the program. These households reside in program CWACS but do not actually receive the program. The comparison group for these households is selected from the full (original) control group—households that reside in control areas. The regression results for calculating the propensity score are shown in Table X3, the distribution of scores at different percentiles shown in Table X4, and the full densities depicted in Figure X2. Once again we see that the treated group (in this case households who were initially listed but later rejected) have a density of scores that is slightly to the right, but there is considerable region of overlap indicating no problem with finding good matches. Interestingly, we find that roughly one-third of matched households in this procedure were also selected as part of the comparison group for the truly treated households (above). This again highlights the fact that there is not much difference between these households in general, though the comparison of means in Section 2 demonstrates that despite this fact, PSM is still able to select a comparison group that is more similar on observable characteristics.

Table A3: Logit Estimates of Odds of Being Rejected for Treatment

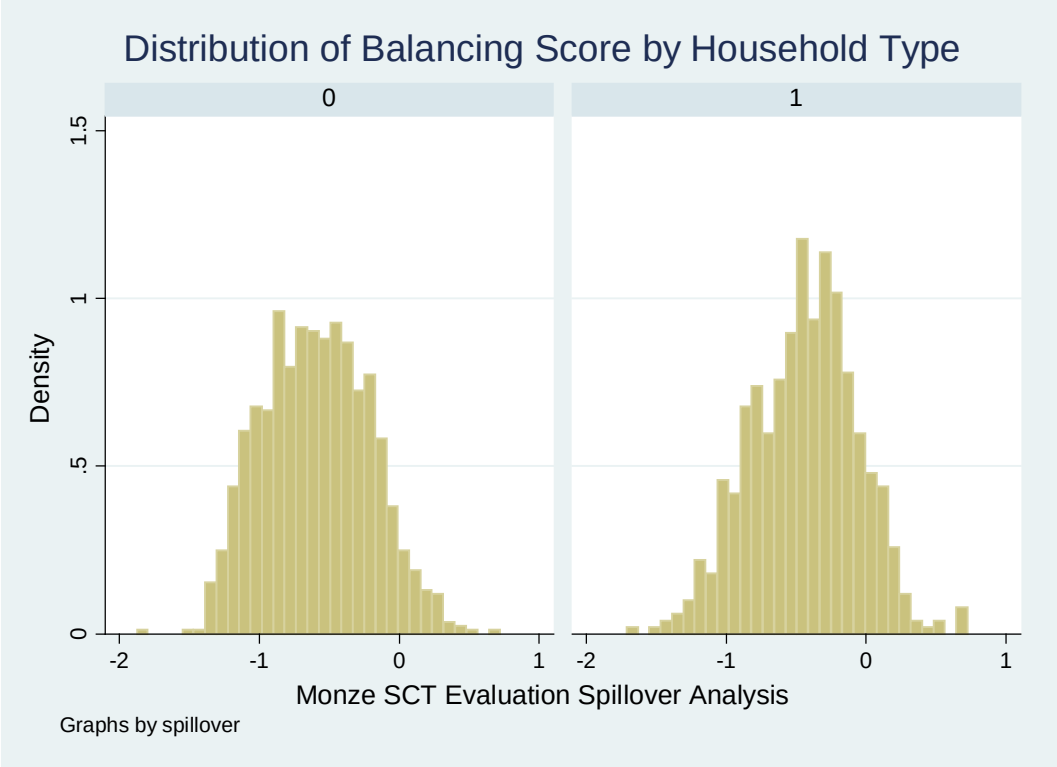
```
Logistic regression                                     Number of obs   =       1653
                                                        LR chi2(10)     =      55.55
                                                        Prob > chi2     =     0.0000
Log likelihood = -1064.3113                             Pseudo R2       =     0.0254

------------------------------------------------------------------------------
       spill |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      wealth |   -.064517   .0736909    -0.88   0.381    -.2089485    .0799145
      lnsize |   .2803403   .1358091     2.06   0.039     .0141594    .5465212
      lnfood |   .0375786    .076944     0.49   0.625     -.113229    .1883861
    demo_dep |  -.0021735   .0370399    -0.06   0.953    -.0747704    .0704234
         fhh |  -.2292248   .1307717    -1.75   0.080    -.4855326    .0270831
    head_age |  -.0096198   .0037345    -2.58   0.010    -.0169393   -.0023003
   hhorphans |  -.0272597   .0379692    -0.72   0.473    -.1016779    .0471586
    head_edu |   .0189406    .018485     1.02   0.306    -.0172892    .0551705
 lnssch_time |  -.2106569   .0495881    -4.25   0.000    -.3078478    -.113466
lnclinic_t~e |    .103735   .0615166     1.69   0.092    -.0168353    .2243053
       _cons |   .0334026   .9433441     0.04   0.972    -1.815518    1.882323
------------------------------------------------------------------------------
```

Table A4: Distribution of Log Odds Ratios by Sample

```
-------------------------------------------------------------------------
spillover |    min(lo)     p25(lo)     med(lo)     p75(lo)     max(lo)
----------+--------------------------------------------------------------
        0 | -1.8747517   -.8728117  -.59668743  -.32100992   .70583385
        1 | -1.7077499  -.71764197  -.43069844  -.19360101   .72446504
-------------------------------------------------------------------------
```

Figure A2: Distribution of Balancing Score by Household Type



Appendix B: Impact Estimates by Household Size

Table B.1: Spending Impacts by Household Size

	Small		Large	
Variables	Coefficient	R-squared	Coefficient	R-squared
Total Food (Kw)	7,575.17	0.018	-3,861.27	0.037
	(1.43)		(1.39)	
Food Shares				
Cereal	-0.031	0.033	0.009	0.028
	(1.07)		(0.26)	
Tubers	-0.016	0.018	-0.010	0.051
	(1.08)		(0.72)	
Legumes	0.018	0.015	0.000	0.007
	(1.12)		(0.01)	
Meats	0.032	0.026	0.019	0.017
	(1.30)		(0.72)	
Dairy	-0.003	0.01	-0.01	0.028
	(0.39)		(1.73)	
Veg	-0.013	0.005	-0.023	0.006
	(0.41)		(0.99)	
Fruits	-0.007	0.006	0.004	0.007
	(1.88)		(1.10)	
Sugars	0.012	0.024	0.003	0.009
	(1.03)		(0.45)	
Oils/Fats	0.004	0.018	0.012	0.023
	(0.36)		(1.19)	
Beverages	0.002	0.007	-0.004	0.01
	(0.33)		(0.77)	
Non-Food Spending (Kw)				
Wood, Carcoal, Grinding	528.957	0.012	-237.821	0.039
	(0.92)		(0.51)	
Fertilizer*	0.057		0.095	
	(1.72)		(1.92)	
Land Purchases	80.973	0.005	-122.963	0.008
	(0.17)		(0.99)	
Land Rent	295.704	0.019	278.723	0.025
	(1.41)		(1.48)	
Observations	992		1032	

Table B.2: Impacts on demographics by Household Size

	Small		large	
Variable	Coefficient	R-squared	Coefficient	R-squared
Number of Elders	0.000	0.57	0.001	0.573
	(0.39)		(0.58)	
Number of Orphans	0.134	0.106	**0.443**	0.15
	(1.14)		(1.98)	
Number of Female Orphans	0.110	0.109	**0.458**	0.167
	(0.92)		(2.03)	
Number of Male Orphans	0.101	0.055	**0.352**	0.106
	(0.97)		(2.24)	
Dependency Ratio	-0.161	0.268	0.263	0.421
	(1.15)		(1.70)	
Number of Deaths	-0.022		-0.041	
	(1.01)		(1.28)	
Obs	996		1040	

Table B.3: Impacts on Livestock and Small Tools by Household Size

	Small	Large
Variables	Marginal Effect	Marginal Effect
Chicken	0.076	0.101
	(0.98)	(1.87)
Pig	**0.095**	0.001
	(2.94)	(0.02)
Goat	**0.457**	0.123
	(8.19)	(1.70)
Cattle	-0.017	-0.029
	(0.86)	(0.89)
Jerrycan	0.014	-0.013
	(0.14)	(0.13)
Radio	0.058	-0.071
	(1.80)	(1.29)
Machete	0.038	0.005
	(1.75)	(0.23)
Pick	-0.015	-0.011
	(0.82)	(0.40)
Axe	0.051	0.036
	(0.95)	(0.64)
Hoe	0.021	0.013
	(0.79)	(0.57)
Oxplough	0.059	-0.021
	(1.65)	(0.69)
Bicycle	-0.013	-0.023
	(0.45)	(0.52)
Observations	996	1040

Table B.4: Outcomes for Children 6-16 by Household Size

	Small		Large	
Variables	Marginal Effect	Observations	Marginal Effect	Observations
Whether Enrolled	0.045	822	**0.079**	2622
	(0.89)		(2.15)	
Enrolled Ages 6-8	**0.291**	217	0.176	702
	(1.96)		(1.79)	
Enrolled Ages 15-16	-0.129	155	**0.132**	448
	(1.13)		(2.61)	
Missed 2+ days last week	0.045	513	-0.047	1468
	(0.69)		(0.75)	
<2 meals in last day	-0.011	821	0.065	2644
	(0.21)		(1.64)	

Table B.5: Outcomes for Children 0-5 Years by Household Size

	Small		Large	
	Marginal Effect	Obs	Marginal Effect	Obs
Under-5 Health Card	-0.258	112	-0.064	509
	(-1.15)		(-0.57)	
Under-5 Visit (Well baby checkup)	0.083	112	-0.014	508
	(-0.61)		(-0.14)	
Morbidity (Whether ill in the last 4 weeks)	0.096	136	0.102	586
	(-0.57)		(-1.07)	
Curative Care	0.114	136	-0.002	586
	(-0.84)		(-0.02)	
Had <2 meals in last 24 hours	**0.998**	21	-0.037	574
	(-15.87)		(-0.73)	

Appendix C: Assessment of Anthropometric Data

There are a total of 1142 children 0-72 months that are in households contained in both waves. Of these children, 863 or 76 percent were actually measured. For the baseline sample, valid z-scores, that is, scores between -5 and +5, were obtained for 553 (height-for age), 450 (weight-for-height) and 645 (weight-for-age) respectively. In other words roughly one-third of the measurements were excluded because they implied a z-score that was outside the valid range (see Table). Moreover the standard deviation for these valid scores is quite large, and in one case (HAZ) is over 2, indicating that data are of poor quality. A comparison of children that were measured versus all children under 73 months indicates that measured children are about 2 months older on average. Furthermore, a comparison of children measured versus those with valid z-scores reveals that those with valid z-scores are 3 months older on average (and thus about 5 months older than the full sample of children 0-72 months of age). This indicates that enumerators had difficulty measuring very young children, a common challenge when taking anthropometric measurements in large-scale field surveys such as this.

Table C1

Variable	Obs	Mean	Std. Dev.	Min	Max
haz	553	-1.317207	2.165223	-4.998241	4.976833
whz	450	.1580743	1.817002	-4.805031	4.742358
waz	645	-.5096661	1.778023	-4.983737	4.988968

The three figures below describe the relationship between measured z-scores and age over the two survey rounds (these are local linear or lowess regressions). The HAZ figure depicts the familiar pattern of steeply declining z-scores about to about 2 years, then a small increase which should be followed by a a flattening of the curve by age 4 years though in these data this flattening is not apparent. Somewhat worrying however is the fact that there appears to be a constant difference in the two curves at all age levels, indicating a systematic measurement error between the two surveys. The relationship between WAZ and age shows the same constant difference between the two survey rounds at all ages, though this time the 2010 graph lies below the 2007 graph. Thus it appears as if weight is over-measured in 2007 and height is under-measured in 2007 and as a result we expect WHZ to be higher in 2007 than in 2010, which is exactly what is shown in the third figure depicting the relationship between WHZ and age in months by survey round. For these reasons we do not feel it appropriate to analyze the anthropometric measurements contained in the survey, though we do provide graphs comparing treatment and comparison group children over time for the three indicators.

Figure C1: Height-for-Age Z-Score by Age: Children 0-72 Months

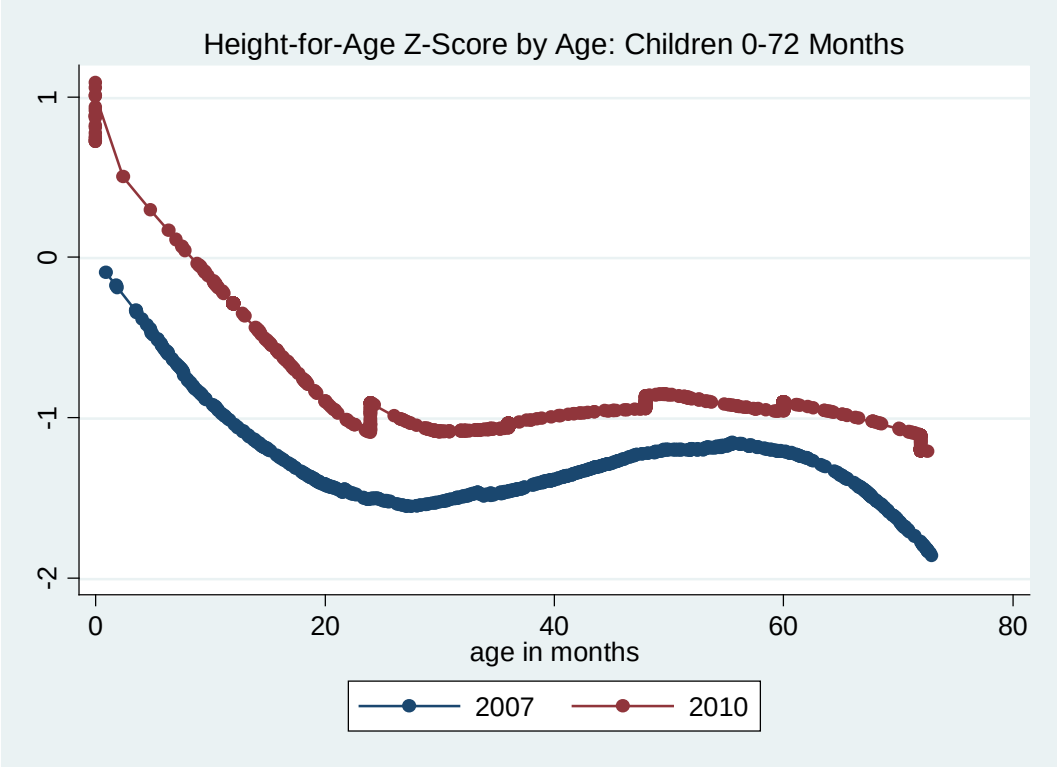


Figure C2: Weight-for-Age Z Score by Age: Children 0-72 Months

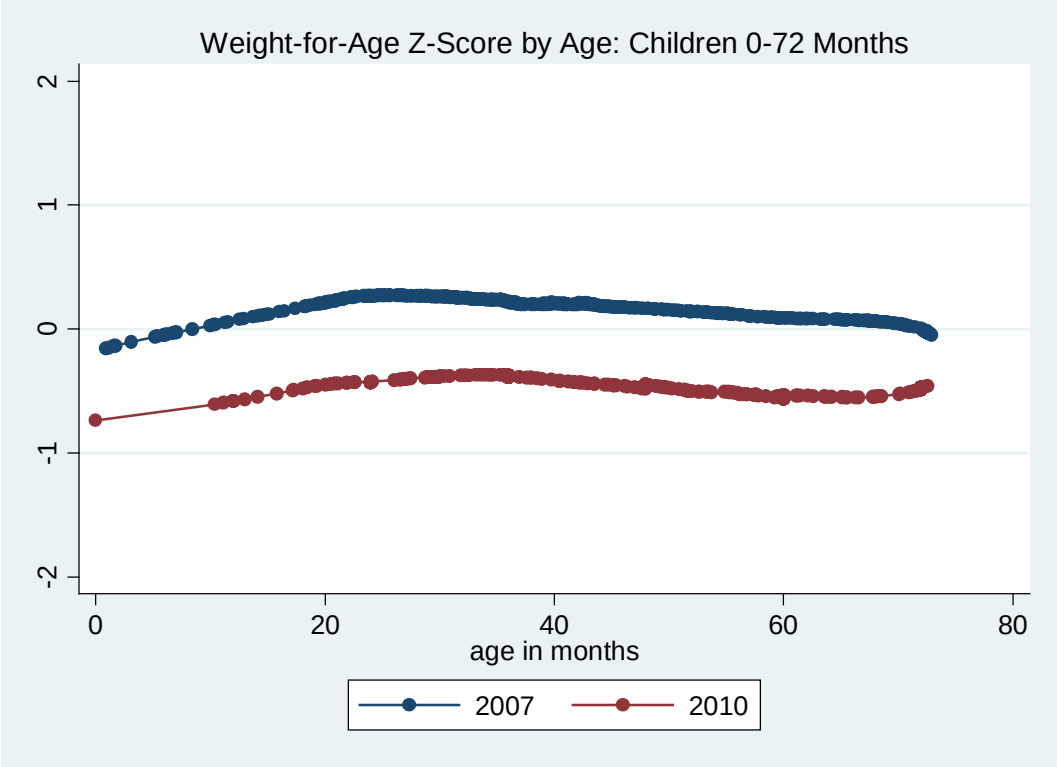
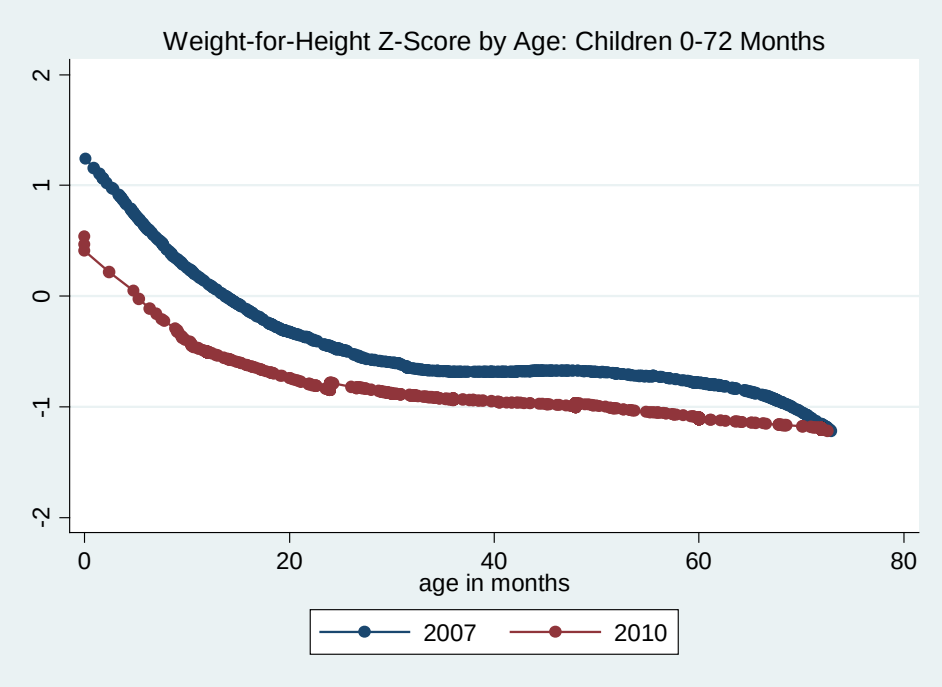


Figure C3: Weight-for-Height Z-Score by Age: Children 0-72 Months



The figures below compare treatment and comparison group children 0-72 months over time for the three anthropometric indicators.

Figure C4: Height-for-Age Z-Score Treatment and Comparison Groups

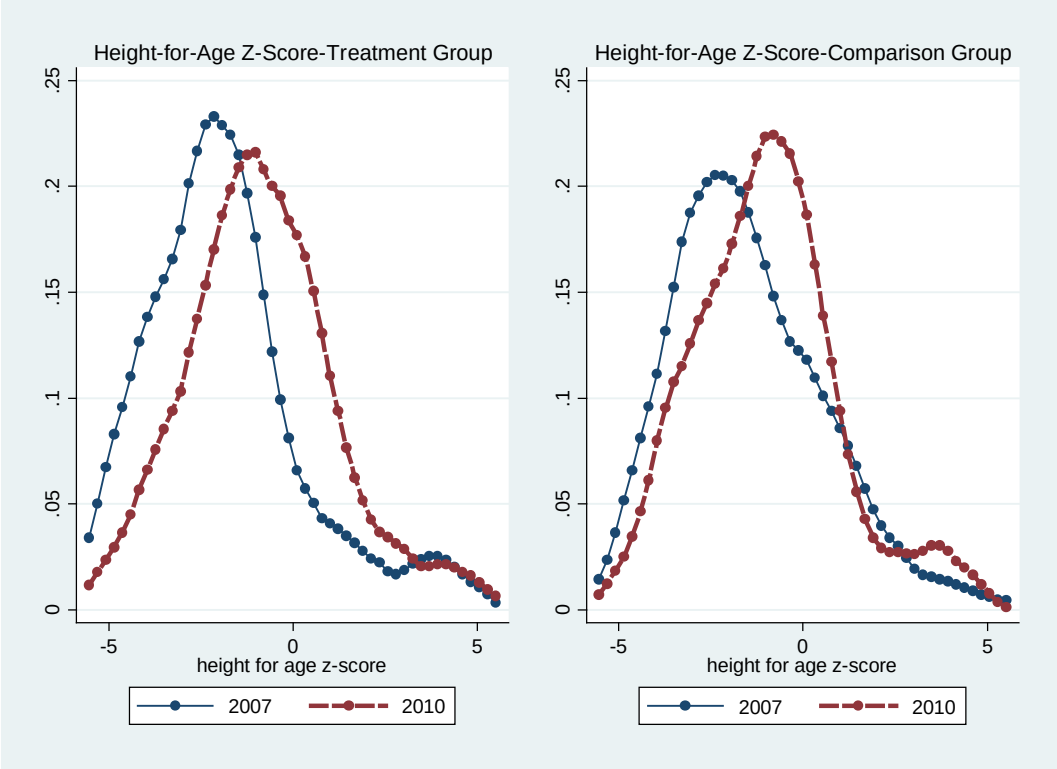
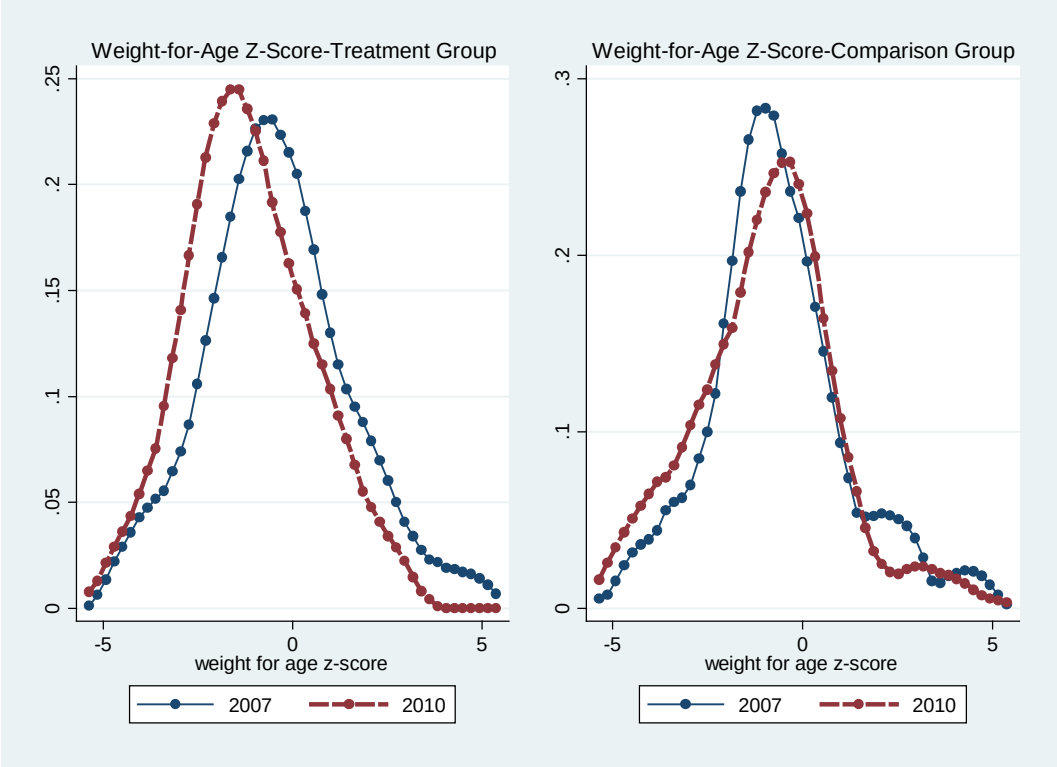


Figure C5: Weight-for-Height Z-Score Treatment and Comparison Groups

Figure C6: Weight-for-Age Z-Score Treatment and Comparison Groups



REFERENCES

Anglewicz, P., Adams, J., Onyango, F., Watkins, S., Kohler, H-P. (2009). The Malawi Diffusion and Ideational Change Project 2004-06: Data Collection, Data Quality, and Analysis of Attrition. *Demographic Research*, 20(21):503-540.

Bardhan, P. K., & Udry, C. (1999). *Development Microeconomics*. Oxford: Oxford University Press.

Barnett, T. 7 Weston, M. (2008). Wealth, Health, HIV and the Economics of Hope. *AIDS*, 22(Suppl 2): S27-S34.

Cardenas, J. C., & Carpenter, J. (2008). Behavioural development economics: Lessons from field labs in the developing world. *Journal of Development Studies*, *44*(3), 311-338.

Heckman, James, Hidehiko Ichimura, and Petra Todd (1997) "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program,." Review of Economic Studies. 64: 605-654.

Heckman, James, Hidehiko Ichimura, and Petra Todd (1998) "Matching as an Econometric Evaluation Estimator." Review of Economic Studies. 65: 261-294.

Juan Jose Diaz & Sudhanshu Handa, "An Assessment of Propensity Score Matching as a Non Experimental Impact Estimator: Evidence from Mexico's Progresa Program," Journal of Human Resources Vol. 41(2): 319-345, 2006.

Miller, C., Tsoka, M. G., & Reichert, K. (2011). Impacts on children of cash transfers in Malawi. In S. Handa, S. Devereux & D. Webb (Eds.), Social Protection for Africa's Children. New York: Routledge.

Rae, J. M. C. W. (1905). The Sociological Theory of Capital. New York: The Macmillan Co.

Sadoulet, Elizabeth & Alain de Janvry (1995) Quantitative Development Policy Analysis, Baltimore: Johns Hopkins University Press.

Samson, M., Heinrich, C., Kaniki, S., Regalia, F., Mac Quene, K., Muzondo, T., et al. (2011). Impacts of South Africa's Child Support Grant. In S. Handa, S. Devereux & D. Webb (Eds.), Social Protection for Africa's Children. New York: Routledge.

Samuelson, P. A. (1937). A note on measurement of utility. The Review of Economic Studies, 4(2), 155-161.

Sianjalika, B. (2010). Zambia in historic buper harvest. Times of Zambia. No 14,773. May 22, 2010.

Smith, A. (1776). An Inquiry into the Nature and Causes of the Wealth of Nations. London: W. Strahan and T. Cadell.

Sudhanshu Handa, Benjamin Davis, Marco Stampini & Paul Winters, "Heterogeneous Treatment Effects in Conditional Cash Transfer Programs: Assessing the impact of Progresa on Agricultural Households," Journal of Development Effectiveness, Vol.2(3): 320-335.

UNAIDS (2008). Report on the Global AIDS Epidemic. Geneva: UNAIDS.

UNICEF (2008) Social Protection in Eastern and Southern Africa: A Framework and Strategy for UNICEF, Nairobi: UNICEF ESARO.

Ward, P., Hurrell, A., Visram, A., Riemenschneider, N., Pellerano, L., MacAuslan, I. & Willis, J. (2010). Kenya CT-OVC Programme Operational and Impact Evaluation 2007-2009. Oxford Policy Management, Oxford, UK.