Title
Health, Human Capital, and Behavior Change: Essays in Development Microeconomics

Author
Kirk, Angeli Elise

Publication Date
2016

Peer reviewed|Thesis/dissertation

Health, Human Capital, and Behavior Change:
Essays in Development Microeconomics

By

Angeli Elise Kirk

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Agricultural and Resource Economics

and the Designated Emphasis

in

Development Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Elisabeth Sadoulet, Chair
Professor Alain de Janvry
Professor David Levine
Professor Jeremy Magruder

Spring 2016

Health, Human Capital, and Behavior Change:
Essays in Development Microeconomics

by

Angeli Elise Kirk

Abstract

Health, Human Capital, and Behavior Change: Essays in Development Microeconomics
by
Angeli Elise Kirk

Doctor of Philosophy in Agricultural and Resource Economics
with Designated Emphasis in Development Engineering
University of California, Berkeley

Professor Elisabeth Sadoulet, Chair

This dissertation combines three empirical studies of household behaviors as they relate to investment in health and human capital in developing countries. The first explores how changes in children's nutrition in Uganda correspond to composition of a household's income. The second studies measurement activities in a cookstove intervention in Darfur, Sudan, with insights into what may be missed in traditional evaluation approaches as well as how technology adoption may benefit from an unintended "nudge." The third evaluates the impacts of a conditional cash transfer program in El Salvador, with a focus on how program compliance and benefits change time allocations among household members.

Chapter 1 explores the relationship between a household's income source (e.g. wage vs. farm) and children's nutrition in Uganda, in a joint work with Talip Kilic and Calogero Carletto. The analysis uses the three annual waves of the Uganda National Panel Survey and features a series of panel regressions for child height-for-age z-scores (HAZ) under age 5. We control for time-invariant child-level heterogeneity and other time-variant observable characteristics using fixed effects. The analysis finds no impact of short-term changes in total gross income on height scores overall. Sector-differentiated analyses indicate that compared to wage earnings, only the share of income originating from non-farm self-employment exerts positive effects on HAZ, while agriculture is more negative. Within agriculture, the income shares from (i) household's consumption of own crop production and (ii) low-protein crop production appear to underlie a negative effect seen from the share of income originating from crop production. We see that results are driven by the older and poorer cohorts, whose diets may be more influenced by shifts in income and production. Overall, any effects are small, given that coefficients represent a change from 100 percent wage income to 100 percent of the other source in a context where many households experience limited changes from year to year. We also cannot say that these relationships are causal, given that observed changes in income likely reflect changes in endogenous livelihood decisions from year to year. Still, the results suggest the possibility of stickiness of crop production to own consumption. While this may be nutrition-supporting in some contexts, it is possible that income growth in the production of low-protein crops in Uganda, which is known for a low-protein diet, may crowd out consumption of other goods and services that have the potential to serve as better nutritional investments. These results suggest a need for more information about how children's diets or childcare patterns accompany income changes.

Chapter 2 studies fuel-efficient cookstove adoption in Darfur, in a joint work with Daniel Wilson, Jeremy Coyle, Javier Rosa, Omnia Abbas, Mohammed Idris Adam, and Ashok Gadgil. In this study, we used sensors and surveys to measure objective versus self-reported adoption of freely-distributed cookstoves. Our data offer insights for how effective measurement and promotion of adoption, especially in a humanitarian crisis. With sensors, we measured a 71% initial adoption rate compared to a 95% rate reported during surveys. No line of survey questioning, whether direct or indirect, predicted sensor-measured usage. For participants who rarely or never used their cookstoves after initial dissemination ("non-users"), we find significant increases in adoption after a simple followup survey (p = 0.001). The followup converted 83% of prior "non-users" to "users" with average daily adoption of 1.7 cooking hours over 2.2 meals. This increased adoption, which we posit resulted from cookstove familiarization and social conformity, was sustained for a 2-week observation period post intervention. Given that most dissemination programs do not employ objective measurement of adoption to inform design, marketing, and dissemination practices, our findings suggest that self-report information may lead programs to over-estimate impacts. A lack of reliable data is likely to prevents insights and may contribute to consistently low adoption rates. Our findings also suggest a potential role for low-cost followup actions that may facilitate learning for a subset of the target population that could benefit from the new technology.

In Chapter 3, I use panel data from El Salvador to examine short-term responses in time use to the Comunidades Solidarias Rurales conditional cash transfer program during 2007/2008, applying difference-in-differences and regression discontinuity methods. Th program was introduced in stages based on observable municipality traits that precluded household-level influence over eligibility. This design allows for a selection-on-observables estimation approach. Because baseline analysis shows significant differences in a few characteristics between earlier and later phases, I use fixed effects specifications to control for time-invariant differences between groups. With only one baseline period, however, I cannot provide evidence against differences in time-variant trends. For each specification, I present results using two bandwidths from the treatment cutoff. To address the small number of municipalities in the sample, I apply wild cluster bootstrapping and present the resulting p-values along those obtained from clustered standard errors as typically applied for larger samples, and show that standard methods would lead to over-rejection of the null hypothesis in multiple instances. I use clustering at the municipality level in both cases. Overall, many of my results are small and somewhat variable across alterative specifications, potentially due to measurement error, a small number of clusters, or simply a small response in the short run to a program offering a relatively small sum of $15-20 a month. Despite these caveats, my findings suggest that for children 6-12, the program appears to have increased school attendance for girls by a small amount relative to boys. There were no gains in enrollment in most specifications, though this may not be surprising in a context where primary school enrollment is already around 90 percent. At the household level, the program may result in a slight reduction of household labor (defined to exclude housework or time allocated to program compliance) for wealthier households relative to poorer households, but a more important change seems to be the shift of productive labor from adult females toward men. Given the total number of statistical tests, however, multiple inference penalties reduce confidence in these few findings.

Acknowledgements

I want to thank my advisors Betty Sadoulet and Alain de Janvry for their guidance, support, and deep patience throughout my studies. I would also like to thank the other members of my committee, Jeremy Magruder and David Levine, for so much helpful advice. My co-authors in this work, too, have contributed endless insights and good humor, and I am happy to call them friends as well as colleagues.

I am thankful to so many other members of Berkeley's department of Agricultural and Resource Economics. I have benefitted frequently from the patient teaching and listening ear of Leo Simon. Carmen Karahalios and Diana Lazo were always available to help solve logistical challenges. My classmates have shared the highs and lows of the last six years, both sharpening my work and making the process much less lonely: in particular, I thank Daley Kutzman, Eddie Montoya, Walter Graf, and Elliott Collins, though there are many others.

My friends and family have also played a key role throughout my doctoral studies. I am eternally grateful to my parents Ron and Ann, who have offered me only unconditional support since long before this part of the journey began and whose encouragement has carried me through the toughest moments along the way. More recently, I have depended deeply on love, support, and laughter from my soon-to-be husband Ben, who has helped me see this work to the end as we start off on a new adventure together.

Chapter 1

Composition of Household Income and Child Nutrition Outcomes Evidence from Uganda

with Talip Kilic and Calogero Carletto[1]

Abstract

This study attempts to contribute to the empirical space between the cross-country analyses that explore the links between income and nutrition without insights on micro-level determinants, and the numerous microeconomic studies that suggest mechanisms of impact but are hindered by some combination of small sample size and incomplete data. The analysis uses the three annual waves of the Uganda National Panel Survey, and features a series of panel regressions of child height-for-age z-scores (HAZ) under age 5, controlling for time-invariant child-level heterogeneity and other time-variant observable characteristics using child fixed effects. We start by showing very little correlation between HAZ and short-term changes in rural household income. We then explore differences by sector of income, first between agricultural and non-agricultural sectors, and by type of agriculture. The analysis finds no impact of short-term changes in total gross income on height-related measures overall but documents very small positive correlations among the youngest children. Sector-differentiated analyses indicate that compared to wage earnings, only the share of income originating from non-farm self-employment exerts positive effects on HAZ, while agriculture is more negative. Within agriculture, the income shares pertaining to (i) consumption of own crop production and (ii) low-protein crop production appear to underlie a negative effect seen from the share of income originating from crop production. Dividing the sample into subsets, we see that the results appear to be driven by the older and poorer cohorts, whose diets potentially may be more influenced by shifts in income and production. Overall, any effects are relatively small, given that coefficients represent a change from 100 percent wage income to 100 percent of the other source in a context where many households depend on a diversified portfolio. We also cannot say that these relationships are causal, given that observed changes in income likely reflect changes in endogenous livelihood decisions from year to year, but the results suggest the possibility of stickiness of crop production to own consumption. While this may be nutrition-supporting in some contexts, it is possible that income growth in the production of low-protein crops in Uganda, which is known for a low-protein diet, may crowd out consumption of other goods and services that have the potential to serve as better nutritional investments. These results are also likely to depend heavily on the agricultural and dietary profile of Uganda and caution against uniform policies to support one sector over another without further information specifically about how children's diets or childcare patterns accompany income changes.

[1]Kilic and Carletto: The World Bank

Introduction

In the quest for widespread and sustainable welfare gains, not all income may have equal effects. Growth within some sectors or accruing to certain individuals within a population may be relatively more effective at reducing poverty and improving specific welfare outcomes in developing countries. Child under-nutrition, targeted directly by the first of the Millennium Development Goals and related to others, is an aspect of poverty that is often argued to be sensitive to growth in the agricultural sector, with potential for both gains and losses. In recent years, there has been a growing movement to pull together evidence on the links among agriculture, income, nutrition and health for the design of multi-sectoral interventions that target nutritional deficiencies.[2]

All income has the potential to benefit children's nutrition, and if household consumption choices depend on production outcomes only via total earnings, income from any source or sector will be equally beneficial. Empirically observed deviations from this theoretical case may originate from multiple sources: distribution of poverty across sectors, relative food production and consumption prices due to markups and transaction costs, risk preferences, and intra-household bargaining outcomes, to name a few. If such deviations occur, the direction and relative weights of these channels of impact would lead to very different prescriptions for policymaking and allocation of scarce resources meant to boost nutrition-supporting growth. Empirically, however, validation of the claims regarding whether and how household sectoral involvement and gains in productivity can contribute to changes in nutritional status and health has been hindered by data limitations and by methodological concerns.

A large collection of microeconomic studies attempting to determine the income links to nutrition through specific mechanisms provide mixed and often conflicting results. The investigated mechanisms include (i) commercialization (reviewed by DeWalt, 1993; Kennedy, Bouis, & Braun, 1992, von Braun & Kennedy, 1994), (ii) gender dynamics (reviewed by Kurtz & Johnson-Welch, 2007; Peña, Webb, & Haddad, 1996; Quisumbing, Brown, Feldstein, Haddad, & Peña, 1995; Quisumbing & Maluccio, 2000), and (iii) nutrition-sensitive production and education interventions (reviewed by Berti, Krasevec, & Fitzgerald, 2004; Gillespie & Mason, 1994; Leroy & Frongillo, 2007; Masset et al., 2011[3]; Ruel, 2001; Soleri, Cleveland, & Frankenberger, 1991). While some differences could be due to context-specific dynamics, numerous reviews in recent years express concerns regarding (i) the validity of the empirical methods used for impact estimation, and (ii) the inconsistency in the types of data used across studies which often lack information on income and have information on only consumption or anthropometry but not both (Arimond et al., 2011; World Bank, 2007;

[2]Some examples include (i) the International Food Policy Research Institute (IFPRI) "2020 Conference: Leveraging Agriculture for Improving Nutrition and Health" that was held in New Delhi, India in February 2011, and (ii) the United States Agency for International Development (USAID)'s review of nutrition and food security impacts of agriculture projects, which can be found here: http://pdf.usaid.gov/pdf_docs/PNADY253.pdf.

[3]asset et al. (2011) narrow down their focus on interventions with an explicit goal of improved child nutrition. The 2011 online version of the publication offer additional details on counterfactual analysis, power, intermediate outcomes, and heterogeneity of impacts, and can be found here: http://r4d.dfid.gov.uk/PDF/Outputs/SystematicReviews/Masset_etal_agriculture_and_nutrition.pdf.

Leroy et al., 2008).

Despite these challenges, the sheer number of studies conducted over the last few decades speaks to the long-standing and urgent demand for insights on how to effectively leverage growth for nutritional improvement. While researchers and key policy players overwhelmingly assert that there is a strong potential for agricultural development to support nutrition and health, they also lament the lack of insight into the specific conditions necessary and sufficient to achieve improved nutritional outcomes efficiently and at broad scale. Herforth (2013) synthesizes the current state of knowledge cites general consensus on many best practices for improving nutrition through agriculture but highlights two questions that are yet to be settled: (i) what are the relative nutritional impacts of agricultural production for own consumption vis-à-vis agricultural production for sales? and (ii) what agricultural products households should focus on, for example staple crops vs. animal-source foods? To this list, we add a third, overarching question that stems from the literature: Even if agricultural growth can be leveraged effectively for nutrition, is it more effective than non-agricultural growth at micro level?

With these questions in mind, we take advantage of the three waves of the household survey data from the Uganda National Panel Survey in an attempt to fill the knowledge gap between the cross-country analyses that explore the links between income and nutrition but cannot explore determinants at a micro level and the numerous smaller microeconomic studies that point to mechanisms of impact but are often hindered by some combination of sample size, data incompleteness, and other methodological considerations. We start by looking at how child nutritional outcomes correlate with short-term changes (1-2 years) in household income regardless of source. Subsequently, we explore heterogeneity by source of income, first between crop cultivation[4] and non-crop sources and then further within type of crop cultivation, according to the priorities set previously in the literature.

There are three key findings. First, we document no detectable impact of short-term changes in total gross income on height-for-age overall, though there may be a very small gain for the youngest children. Second, sector-differentiated analyses indicate that only the share of income originating from self-employment exerts positive and statistically significant effects on height relative to other sectors. Third, the income shares pertaining to (i) a household's consumption of own crop production and (ii) low-protein crop production, rather than crop production alone, appear to be driving the negative effect of the share of income originating from crop production. All of these relationships are small relative to typical year-on-year changes in income composition. The remainder of the paper is structured as follows. Section 2 discusses the theoretical mechanisms through which income growth and sector and subsector of growth can influence nutrition in the context of the existing body literature. Sections 3 describes our data sources; Section 4, empirical strategy and results. Section 5 concludes.

[4]While "agriculture" can refer to both crop cultivation and animal rearing, we focus primarily on crops and not animals.

Linking Income and Agriculture to Nutrition: Theory and Literature

The factors that are commonly understood to interact to that hinder nutrition are 1) household food insecurity, which encompasses food availability as well as quality, 2) inadequate care, and 3) unhealthy environment (UNICEF, 1990; Behrman & Deolalikar, 1988).[5] The direction of these biologically-based impacts is well established in the literature, and we take them as given: any positive or negative impacts of agriculture on nutrition must act through these channels. Descriptively, we offer a health production function for nutritional outcomes:

$$H_i = H(f_i, n_i, s_i, X_i),$$

which over time accumulate as:

$$H_{it} = H(f_i t, n_i t, s_i t, X_i, H_{it-1})$$

where time t-indexed food consumption f_{it}, care/nurturing n_{it}, and sanitary environment s_{it} as well as a vector of individual or household characteristics X_i and previous nutritional health outcomes H_{it-1}. Lack of any factor, such as food, care, sanitation, may be sufficient to induce under-nutrition, and the provision of each is expected to complement the others in producing health (while competing through the budget constraint), so we would expect the true production function will contain interactions of these terms, likely with non-linearities and minimal subsistence terms.

Connecting the dots conceptually from income to nutrition, households may value health directly or may value consuming inputs that contribute to health (food, care, sanitation) as well as other consumption c_{it} and leisure l_{it}, according to household characteristics X_{it}:

$$U_{it} = U(f_{it}, n_{it}, s_{it}, c_{it}, l_{it}, X_{it}).$$

The household wants to maximize utility subject to a budget constraint such as

$$p_f f_{it} + p_s s_{it} + p_c c_{it} - w(n_{it} + l_{it}) \leq I_{it},$$

where p_f, p_s, p_c, w are the prices of food, sanitation, other consumption, and the wage rate; and income I_{it} comprises represents farm profits, non-agricultural enterprise profits, and the value of household labor and land endowments. [6]

Under basic household models, income only affects these nutrition-inducing consumption choices by setting the budget constraint, with no other characteristic of income having influence. By relaxing the budget constraint, increases in income from any source may lead to

[5] There is a large literature establishing the importance of all three factors for nutritional outcomes, which we take as given, though a precise nutrition production function is widely absent, given measurement difficulties and identification challenges that arise from reliance on observational data.

[6] Given the limitations of nutritional science to define the biological relationships more precisely, and acknowledging the admonitions of Behrman and Deolalikar (1988), we do not detail a functional form of utility on health or of health on the relevant inputs. We will rely on multiple specifications and robustness checks rather than claiming a structural form.

greater food consumption; nutritional gains may be further facilitated by higher marginal consumption of food among the poor (Engel's Law) especially in terms of consumption of calories and essential micronutrients (Skoufias, Tiwari, & Hassan, 2012; Strauss & Thomas, 1995; Subramanian & Deaton, 1996). At the same time, income gains enable greater consumption of complementary health inputs such as sanitation improvements and healthcare services, and the income elasticity of health and sanitation expenditures can remain quite high throughout the income distribution (von Braun et al., 1991). Income can be used for childcare services or otherwise improve the quality of care given as well. For example, higher expenditure on education allocated to girls as a result of increased income eventually translates into higher maternal education, shown to improve child nutritional outcomes (Behrman & Wolfe, 1984; Umapathi, 2008; Webb & Block, 2004), though this can take years or decades to materialize.

Empirical studies using pooled cross-sectional data provide evidence that nutritional outcomes do improve alongside long-run, aggregate economic growth (Cole, 2003; Haddad & Smith, 2002; Headey, 2013;[7] Webb & Block, 2010). Yet this relationship is not guaranteed, depending on duration and distribution of growth. Under the permanent income hypothesis and consumption smoothing, short-term income fluctuations may be less likely to induce consumption of food or sanitation when compared to longer-term gains (Hall & Mishkin, 1982). Clearly, a household must be able to participate when there is aggregate growth in order to benefit from it. Looking at "nutritional episodes" with an average duration of 4.7 years, Heltberg (2009) looks at income growth across countries but finds less improvement in child stunting rates compared to longer term studies, with nutrition improving less in more unequal societies. Relatedly, Webb and Block (2010), with data largely drawn from Sub-Saharan Africa, find that growth from structural transformation fails to support nutrition for the rural poor in the short run but point to agriculture effectively lowering stunting by reaching the rural poor. Headey (2013) finds that once India is excluded in cross-country regressions, agricultural growth corresponds to a stronger reduction in stunting than non-agricultural growth in the medium term. Again, even if a household is able to participate in income growth, conversion to nutrition through the mechanisms of food, care, and sanitation may take time.

Agriculture for income

Is agricultural growth the most effective way forward to support nutrition? These income results for nutrition above need not be specific to agricultural income. Yet since many of the world's rural poor are dependent on agriculture as their main livelihood, growth in agriculture has the potential to be relatively more effective in reducing income poverty (Chen & Ravallion, 2007; de Janvry & Sadoulet, 2010), a strong determinant of under-nutrition. At the macro level, Ligon & Sadoulet (2007) find that agricultural income growth exerts a particularly beneficial effect on expenditures among the poorest and that non-agricultural growth boosts expenditures in a more modest fashion among these households. Also using cross-country studies, Heltberg (2009) suggests the explanation that non-agriculture growth

[7] Headey (2013) is able to point to greater food production, though not consumption in particular.

associated with structural transformation tends to be geographically exclusive of the rural poor, and Loayza and Raddatz (2010) provide evidence that the gains arise through agriculture providing labor-intensive income opportunities to the unskilled. Christiaensen et al. (2011) find that the benefits from agricultural growth are more concentrated among the extreme poor (less than $1 a day) than among the better-off poor. Extending from poverty outcomes to nutrition outcomes, if agriculture is more accessible to the poor, then agricultural income could have more potential to improve nutrition-supporting consumption.

Agricultural sub-sectors: commercialization vs. own consumption, and crop choice

Within agriculture, too, the type of agricultural growth may have important implications for pass-through to nutrition. Agricultural commercialization is often favored for its ability to facilitate specialization, technological growth, and higher expected returns, thus allowing households to convert in-kind income to cash income, which can in turn be used to purchase greater food security and other health-supporting goods and services (Kennedy, 1994; Pingali, 1997; Pingali & Rosegrant, 1995; Romer, 1993, 1994; Timmer, 1997; von Braun, 1995). Yet findings from various studies suggest that increased income through commercialization haven't always yielded nutritional improvements and sometimes have been associated with nutritional declines among farming households (DeWalt, 1993; Dewey, 1981; Fleuret & Fleuret, 1980; Kennedy, Bouis, & von Braun, 1992; von Braun & Kennedy, 1986, 1994).[8] Theoretical mechanisms for this possibility reflect that cash income can facilitate substitution toward non-food consumption or toward consumption of less nutritious foods through changing preferences or shifting of resources and/or control among household members with different expenditure preferences (Bouis & Haddad, 1990; von Braun et al., 1991; von Braun, 1995). Other studies offer an alternative explanation in which labor inputs necessary for commercialization may in some cases detract from health-supporting efforts in the home (e.g. breastfeeding or other childcare) (Abbi, Christian, Gujral, & Gopaldas, 1991; Kennedy & Cogill, 1987; Popkin, 1980) or increase exposure to hazardous chemical inputs or zoonotic disease (Mullins, Wahome, Tsangari, & Maarse, 1996). Access to commercialization may also offer household investment opportunities that increase the opportunity costs of current consumption, potentially suppressing food expenditures in the short run.

By contrast, agricultural production for own consumption (subsistence agriculture) has been viewed traditionally as a last-resort, low-productivity option for those who face high transaction costs and missing markets or who are highly risk averse (Timmer, 1997). Yet there is a growing momentum for promoting own production as a direct support of food security, dietary diversity, and nutrient-dense consumption. An implicit assumption in these interventions is that food production income will be more likely to "stick" as food consumption relative to other kinds of income. For example, von Braun et al. (1991) found that even after controlling for total income level, households with higher ratios of subsistence food production as a proportion of total income show higher food consumption. Designed

[8] See Timmer (1997) and Strasberg et al. (1999) for a general discussion and a discussion on Kenya, respectively.

with this stylized fact in mind, interventions that encourage dietary diversity and protein or micronutrient consumption through home-production channels – home gardens, biofortified varieties, and animal-sourced foods – do appear to successfully effect improvements in relevant biomarkers in some cases, with the caveats mentioned above (Masset et al., 2011).

The "stickiness" of gains in own food production may bear out in part through price effects and risk aversion.[9] The Food Price Crisis of 2007-08 has served as a reminder that the production of food crops can help insure vulnerable groups' consumption against food price risk, since rising food prices also raise the income value of the crop at the same time (de Janvry & Sadoulet, 1995; Headey, 2013). Especially in rural areas, where households face shallow markets, seasons of high and geographically correlated production will lower relative food prices, inducing substitution toward food consumption. Price risk aversion and transaction costs can further increase consumption of own production by driving a wedge between the effective sale and purchase prices, again making consumption of own food relatively more attractive (de Janvry et al., 1991, Jensen, 2010; Key et al., 2000; Svensson & Yagaizawa, 2009). Completely missing markets for the purchase of nutritious foods represents the extreme case of transaction costs, in which the only means of acquiring necessary micronutrients and achieving dietary diversity is own production.

In a more mechanical sense similar to the general argument for income, improved productivity in food cropping for own consumption may be differentially good at boosting food consumption and then nutritional outcomes because it is often the very poor and women who engage in subsistence agriculture and who may be most likely to convert gains into increased food intake. Aside from the distinction between commercialization and own production, the nutritional qualities of the particular crop (or animal) associated with income growth may also be relevant. At the macro level, Headey (2013) goes further than previous cross-country analyses to show that the nutritional gains are strongest where agricultural growth manifests as increased food production and in countries whose food production was low initially.[10] And to explain part of India's failure to convert economic growth to nutrition, pooled cross-sectional studies point to non-food agricultural production and price effects that shift consumption from more protein-rich pulses toward cheaper and less nutrient-rich grain (Deaton & Dreze, 2009; Headey, Chiu, & Kadiyala, 2012). Given the level of geographic aggregation, however, none of these studies are able to offer insight on whether for the individual or the households, it is important to produce for one's own consumption, or whether in the presence of sufficiently deep markets for nutritious foods, households may be better of maximizing the income value rather than the nutritional value of their agricultural portfolios. Clearly, the nutritional benefit derived from consumption from own crop production will depend on the nutritional quality of crops being produced, and the benefit from other sources of income must depend on the nutritional value of food being purchased.

[9] Many interventions also incorporate educational components to try to increase preference for nutrition-supporting consumption, a mechanism outside of what we can test using the current dataset.

[10] It is unclear whether these changes arise through agricultural growth among the poor or through falling food prices economy-wide.

Data

To explore the links between income and child nutrition, we make use of three rounds of the nationally-representative Uganda National Panel Survey (UNPS) collected in 2009/10, 2010/11 and 2011/12. The UNPS is implemented by the Uganda Bureau of Statistics with financial and technical support from the World Bank Living Standard Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) program.[11] In each round, the survey collects anthropometric measures (height and weight) for children under five years of age and detailed information on household consumption, income, and agricultural activities. Having individual-level anthropometric measures and household-level income, including detailed information about agricultural production and consumption from own production, offers the opportunity to explore potential pass-through (or not) from income to nutrition via consumption.[12] Specific to the UNPS, the panel nature (at the household and individual levels) and the short period between the survey rounds allow us to conduct within-household analyses over time, and to control in our estimations for unobserved time-invariant child attributes that may be missed in cross-sectional or pooled cross-sectional investigations. In fact, it is this feature that provides the basis for the present study.

Given the different income and consumption patterns between urban and rural populations, the specific focus on agricultural income, and the relatively smaller sample size for the urban population, we focus on the rural subsample of households in each round and on children that appear in at least two of the three survey rounds. The latter restriction allows for the inclusion of child-specific fixed effects in our estimations, and leads to 748, 924, and 653 child observations in 2009/10, 2010/11, and 2011/12, respectively, to be part of the analysis sample. Table 1a provides summary statistics for the sample across the three panel waves. Within this sample, the median household has approximately seven members, including four children under the age of 15 and two children less than five years of age. 53 percent of the children in the sample are male. As is typically found in Uganda, there are much higher levels of stunting (35-38 percent across three years) than underweight (11-16 percent).

Table 1b presents income and consumption statistics for the sample, deflated to the first survey round in 2009/10 and converted to US dollars. While rural households derive their income from multiple sources, nearly all households participate in crop farming in each round, and the vast majority also engages in livestock activities. Approximately half report non-agricultural self-employment, with average self-employment (for the whole population) approximately equal to average crop income. Agricultural and non-agricultural wage employment each count one-quarter of the households, with greater income coming from non-agricultural wages.

[11] The UNPS data and documentation are publicly available on www.worldbank.org/lsms.

[12] Income measures have been constructed following the cross-country comparable Rural Income Generating Activities (RIGA) income aggregate methodology. More information on RIGA is available on http://www.fao.org/economic/riga/rural-income-generating-activities/en/.

Table 1a: Summary Statistics for UNPS 2009/10 - 2011/12
Demographics, Assets, and Child Characteristics
Subset: rural children with at least two observations in the three waves

	2009/10 (748 obs)			2010/11 (924 obs)			2011/12 (653 obs)		
Variable	mean	sd	median	mean	sd	median	mean	sd	median
REGION									
Central w/o Kampala	0.235	0.424	0	0.256	0.437	0	0.27	0.444	0
Eastern	0.32	0.467	0	0.318	0.466	0	0.291	0.455	0
Northern	0.287	0.453	0	0.302	0.459	0	0.309	0.463	0
Western	0.158	0.365	0	0.123	0.329	0	0.13	0.337	0
DEMOGRAPHICS									
Household Size	7.13	2.78	7	7.04	2.56	7	7.11	2.56	7
Number of children <15yrs	4.33	1.93	4	4.37	1.97	4	4.36	1.96	4
Number of children <5yrs	1.95	0.809	2	1.94	0.835	2	1.94	0.886	2
Number of adults 15+	2.81	1.43	2	2.67	1.21	2	2.75	1.24	2
Number of males in household	1.26	0.944	1	1.19	0.804	1	1.26	0.821	1
%Female-headed household	0.168	0.375	0	0.18	0.384	0	0.176	0.381	0
Dependency ratio	1.83	0.992	1.59	1.95	1.12	1.67	1.9	1.12	1.67
Head's years of school	5.61	4.32	6	6.28	4.63	6	5.75	4.32	6
Spouse's years of education	3.88	3.77	4	4.51	4.24	4	4.26	3.93	4
Average years of education for members>21	4.54	3.15	4	5.01	3.45	4.4	4.6	3.16	4
Highest years of education in household	7.17	4.26	7	7.99	4.64	7	7.38	4.2	7
ASSETS									
Has Improved Roof	0.523	0.5	1	0.527	0.5	1	0.545	0.498	1
Has Improved Walls	0.634	0.482	1	0.642	0.48	1	0.688	0.464	1
Has Improved Floor	0.148	0.356	0	0.162	0.369	0	0.164	0.37	0
Treats Water	0.357	0.479	0	0.339	0.474	0	0.315	0.465	0
Has Improved Water Source	0.686	0.464	1	0.697	0.46	1	0.741	0.438	1
Has Improved Toilet Facility	0.893	0.309	1	0.878	0.328	1	0.888	0.315	1
Has Hand Washing Station	0.098	0.297	0	0.042	0.201	0	0.075	0.264	0
CHILD									
Age of Child (in months)	25.2	11.5	25	33.3	14.3	33	40.5	11.1	41
Gender: 1 = Male	0.528	0.5	1	0.532	0.499	1	0.531	0.499	1
Height, cm	81.4	9.33	81.4	86.8	10.7	87	91.4	8.19	91
Weight, kg	10.8	2.56	10.6	12.3	2.92	12.4	13.5	2.43	13.4
Height for Age Score	-1.55	1.48	-1.59	-1.58	1.4	-1.56	-1.71	1.25	-1.69
Weight for Age Score	-0.908	1.17	-0.855	-0.792	1.1	-0.775	-0.817	1.02	-0.81
Weight for Height Score	-0.078	1.18	0.015	0.138	1.13	0.14	0.216	1.09	0.21
% Stunted (HAZ < -2)	0.368	0.482	0	0.354	0.478	0	0.38	0.486	0
% Wasted (WHZ < -2)	0.159	0.366	0	0.116	0.32	0	0.104	0.306	0
% Underweight (WAZ < -2)	0.059	0.236	0	0.031	0.175	0	0.02	0.14	0

Notes: Sample includes only rural households with children whose birthdate was consisent across panel waves. Z-scores calculated from date of birth rather than reported age in month when not in agreement.

Table 1b: Summary Statistics for UNPS 2009/10 - 2011/12
Income and Consumption
Subset: rural children with at least two observations in the three waves

	2009/10 (748 obs)			2010/11 (924 obs)			2011/12 (653 obs)		
Variable	mean	sd	median	mean	sd	median	mean	sd	median
INCOME									
Net Total Income (1)	892	1,198	578	968	1,085	623	1,105	1,416	703
Net Total Income (2)	1,084	1,215	775	1,075	1,100	764	1,230	1,420	893
Gross Total Income (1)	1,399	1,679	870	1,459	1,578	954	1,548	1,891	988
Gross Total Income (2)	1,589	1,705	1,103	1,569	1,587	1,076	1,673	1,884	1,153
PARTICIPATION									
% Agriculture Wage Employment > 0	0.286	0.452	0	0.237	0.425	0	0.193	0.395	0
% Non-Ag Wage Employment > 0	0.242	0.429	0	0.215	0.411	0	0.225	0.418	0
% Crop Production (1) > 0	0.939	0.24	1	0.926	0.261	1	0.928	0.259	1
% Crop Production (2) >0	0.949	0.22	1	0.944	0.231	1	0.939	0.24	1
% Livestock Production > 0	0.762	0.426	1	0.728	0.445	1	0.689	0.463	1
% Non-Ag Self Employment > 0	0.537	0.499	1	0.522	0.5	1	0.489	0.5	0
% Transfers > 0	0	0	0	0	0	0	0.006	0.078	0
SECTOR INCOME									
Self-employ, Net	192	585	3.55	264	641	1.84	292	762	0
Self-employ, Gross	494	1,242	22.8	559	1,215	26.2	604	1,399	0
Crop Income (1), Net	282	416	171	279	349	178	373	506	234
Crop Income (2), Net	358	456	226	357	381	252	441	546	275
Crop Income (2), Gross	474	466	386	386	389	295	499	524	395
Crop Income (2), Gross	548	502	442	467	416	361	565	559	447
Livestock Income, Net	115	237	53.2	127	252	63	91.7	188	29.8
Livestock Income, Gross	242	281	149	239	290	146	150	210	74.5
Wages:	294	886	0	287	753	0	332	1,056	0
Ag Wages	84.6	263	0	61.3	176	0	71.3	256	0
Non-ag Wages	210	820	0	226	746	0	260	1,034	0
CROP SHARES									
% of Gross Income from Crops	0.355	0.285	0.289	0.360	0.295	0.292	0.401	0.313	0.345
% Gross Crop Income, Low Protein Crop	0.306	0.302	0.237	0.333	0.294	0.29	0.34	0.309	0.31
% Gross Income, Low Protein Crop	0.121	0.168	0.051	0.136	0.177	0.069	0.16	0.193	0.087
% Gross Crop Income, Other Food Crop	0.575	0.342	0.604	0.529	0.336	0.527	0.531	0.34	0.539
% Gross Income, Other Food Crop	0.21	0.217	0.139	0.195	0.206	0.131	0.215	0.218	0.137
% Gross Crop Inc, Nonfood Crop	0.057	0.141	0	0.067	0.146	0	0.057	0.136	0
% Gross Income, Nonfood Crop	0.024	0.068	0	0.028	0.07	0	0.026	0.068	0
CONSUMPTION									
Total Annual Food Consumption	694	509	579	835	619	685	804	640	639
Total Annual Crop Food Consumption	246	196	196	261	221	194	256	236	201
Total Annual Low Protein Food Consumption	27.6	58.5	0	25.7	61.8	0	36.1	74.5	0
Total Annual Livestock/Byproduct Consumption	97.5	118	51	131	145	90.3	135	149	96.8
Total Annual Food Consumption, Purchases	333	251	272	400	307	321	380	310	308
Crop Income, Own Consumption (1)	170	176	112	180	175	131	229	233	153
Crop Income, Own Consumption (2)	360	266	313	294	246	239	360	280	329
Low Protein Crop Income, Own Consumption (1)	90.6	127	43	99.4	125	55.8	131	163	71.7
Low Protein Crop Income, Own Consumption (2)	126	134	89.2	119	160	71.8	126	147	67.8
Livestock Income, Own Consumption (2)	43.1	93.2	0	52	112	0	44.5	102	0

Notes: Sample includes only rural households with children whose birthdate was consisent across panel waves. Total income and crop income (1) with value of consumption from own production calculated from agricultural module. Total income and crop income (2) calculated from conumption module. All income and consumption values are deflated to 2009/2010 and converted to USD using March 1, 2010. "Low protein crops" are cassava and plantain, two staple food crops in Uganda.

We report initially two alternative crop income calculations that estimate the value of crop production consumed at home either from the agriculture questionnaire (1) or the food consumption section of the household questionnaire (2). The second methodology generates lower total and crop income in the second wave compared to the other rounds: this reflects local and international food price fluctuations that occurred during the span of the survey periods more than changes in production. The second methodology also relies on consumption data from a single week, while the first methodology uses reported consumption of own produce for the entire years. To minimize the impact of seasonality in food consumption reporting on the calculation of income measures, we elect to focus on the income variables derived using the first methodology for the remainder of this study.[13]More than one-third of gross income comes from crops, with approximately one-third of crop income coming from two "low-protein" crops, namely cassava and plantain varieties. We do not classify other crops according to nutritional status but focus on cassava and plantain both as the top two starchy staples consumed in Uganda as well as two major crops that are particularly low in protein as well as many other important nutrients (FAO, 1990), in a context where diets are recognized to be largely deficient in protein and vitamin A and zinc, among other micronutrients (FANTA/USAID, 2010). These two crops account for 12-16 percent of total gross income. Non-food crops, which include coffee, comprise only a small fraction of the gross income portfolio. The decrease in livestock income in the third year is tied to the decline in the sales, births, and production of byproducts, which may partially be underlined by the outbreak of food and mouth disease during the reporting period (FEWS NET, 2011).[14]To examine the implications of restricting the sample for fixed effects to include only children who appear in at least two of the three survey rounds,[15] the in-sample and out-of-sample means were compared for the children in the first survey round (2009/2010) for each of the variables reported in Tables 1a and 1b. Table 1c reports only the outcomes for which the differences were statistically significant at the 0.10 level or lower. The included sample is more heavily representative of the Eastern Region, and less representative of the Western Region. The spouse of the head of household tends to have completed a half a year more of schooling. The included sample is slightly more representative of boys, at 53 percent versus 49. Unsurprisingly, the included sample is an average of 13 months younger in the first survey round. In fact, approximately 40 percent of the sample was over the age of 48 months during the first round; most of these would be too old to be measured in any following round and thus drop from the sample. Due the age difference, weight and height are also lower among the included sample; however the respective z-scores are not statistically different.

[13]Livestock income is the exception to this. Due to changes in the livestock modules of the agriculture questionnaire during the three waves of the UNPS, the estimated value of livestock and livestock by-product production consumed at home is derived from the food consumption section of the household questionnaire.

[14]Anecdotally, there is also some concern that the livestock modules may have suffered differential attrition relative to other modules over the course of the panel series, perhaps related to survey fatigue and the fact that the livestock modules were administered after several other modules.

[15]The out-of-sample population additionally includes (i) children whose birthdates did not match between survey rounds and could not be reconciled using reported age in months, (ii) a small number of observations who were missing household income data, and (iii) eight children who change households within the UNPS sample across survey rounds.

Many of the remaining differences suggest that more heavily agricultural households were more likely to be resurveyed in future rounds, potentially related to greater permanence of residence and lower attrition.

Table 1c: Comparative Statistics
In Sample vs Out of Sample, 2009/2010 Survey Wave
Subset: rural children

	Means		
	In Sample	Out of Sample	Difference
REGION			
Eastern	0.32	0.268	.052**
	(0.02)	(0.02)	(0.02)
Western	0.157	0.256	-.098***
	(0.01)	(0.02)	(0.02)
DEMOGRAPHICS			
Spouse's years of education	3.89	3.41	.477**
	(0.14)	(0.13)	(0.19)
CHILD			
Age of Child (in months)	25.10	38.00	-12.9***
	(0.42)	(0.58)	(0.71)
Gender: 1 = Male	0.53	0.49	.041*
	(0.02)	(0.02)	(0.03)
Weight, kg	10.80	13.00	-2.13***
	(0.09)	(0.11)	(0.14)
Height, cm	81.30	89.50	-8.20***
	(0.35)	(0.45)	(0.59)

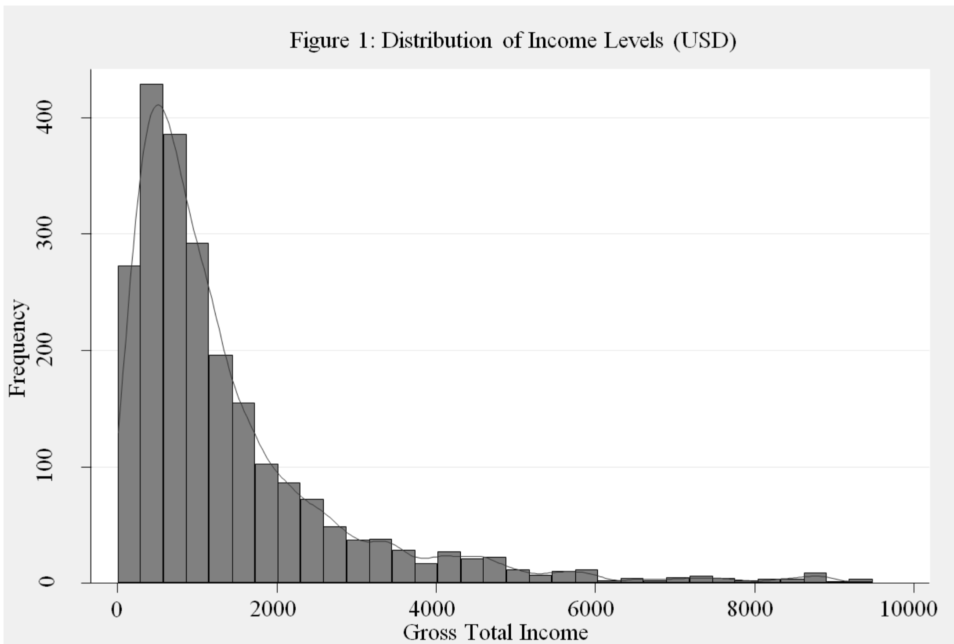
	Means		
	In Sample	Out of Sample	Difference
ASSETS			
Has Improved Roof	0.522	0.598	-.075***
	(0.02)	(0.02)	(0.02)
Has Improved Walls	0.63	0.55	.078***
	(0.02)	(0.02)	(0.02)
PARTICIPATION			
% Crop Production (1) > 0	0.94	0.88	.064***
	(0.01)	(0.01)	(0.01)
% Crop Production (2) >0	0.95	0.90	.046***
	(0.01)	(0.01)	(0.01)
% Livestock Production > 0	0.76	0.72	.042*
	(0.02)	(0.02)	(0.02)
SECTORINCOME			
Livestock Income, Net	116.00	96.20	20.0*
	(8.69)	(8.17)	(12.00)
Ag Wages	84.00	60.60	23.4**
	(9.51)	(6.60)	(11.30)
% Gross Crop Income, Other Food Crop	0.57	0.52	.052***
	(0.01)	(0.01)	(0.02)

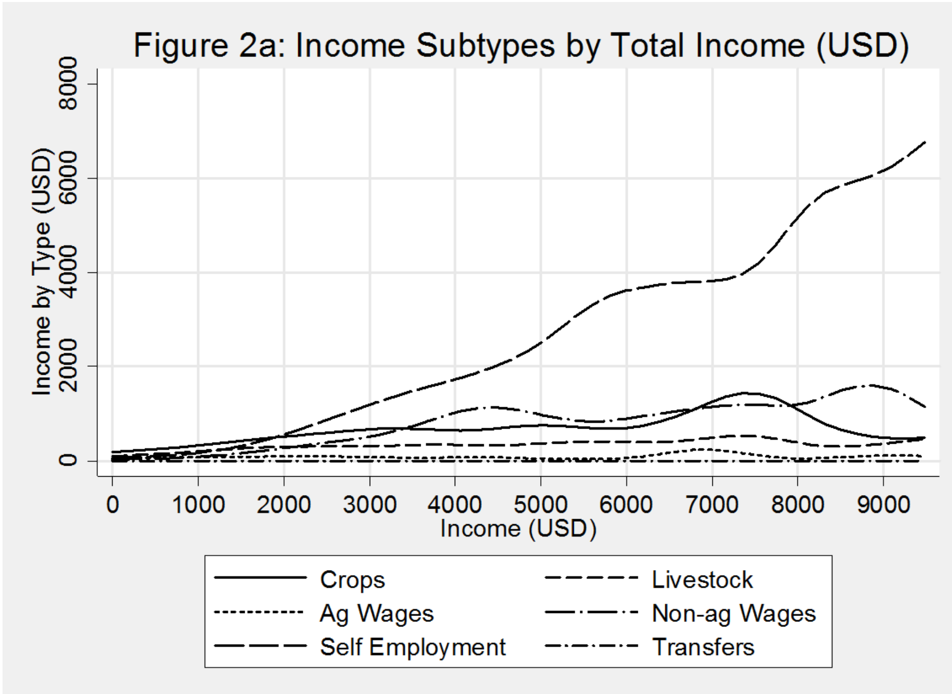
Only significant differences are presented, from comparisons of all variables presented in Tables 1a and 1b. Std errors in parentheses. In Sample includes children in at least two rounds, with matching or reconcilable birthdates and income data. Sample size: In Sample, 756 obs; Out of Sample 895 obs.

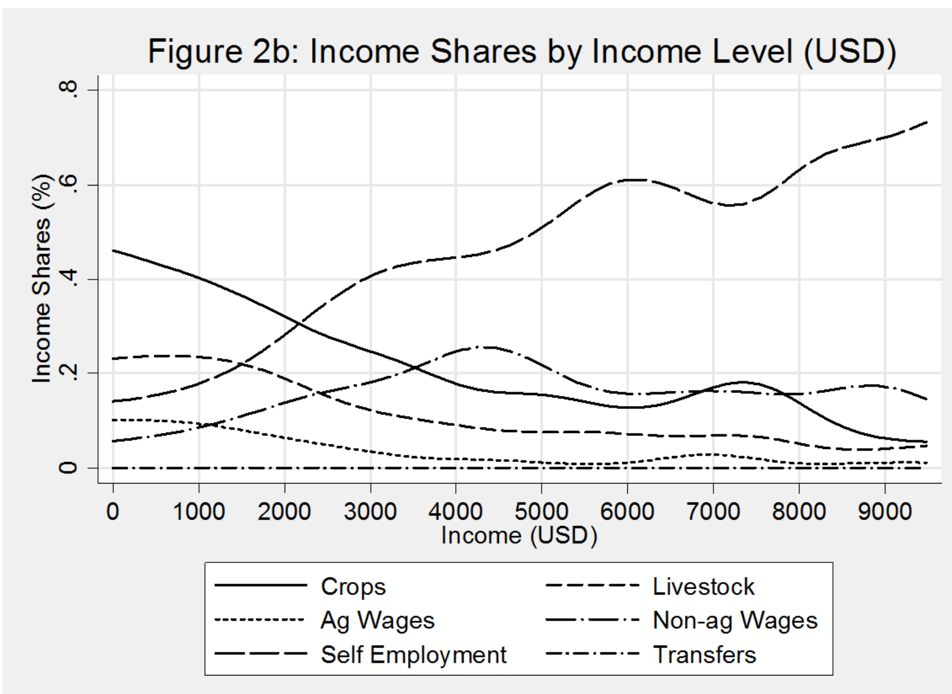
Figure 1 gives the distribution of gross total income for all three year combined.[16]Figures 2a and 2b give a bit more insight into the distribution of income by source. Though income from any particular source does not decrease in levels, as overall income increases, there is

[16]We choose to focus on gross income for a number of reasons, including differences in the ways that expenses were collected across years, small numbers of households reporting any expenditure, and the difficulty of shares in the presence of negative numbers. Gross income may also better represent intensity of income activity by sector.

a clear pattern of agricultural income (crops, livestock, and agricultural wages) falling as a share of total income, strongly in favor of self-employment income. Except for the top percentage of earners, nonagricultural wage also increases its contribution as total income increases. Breaking down crop income into three types, low-protein (cassava and plantain varieties), non-food (cotton, tobacco, coffee), and other food crops, Figure 3 shows a trend that low-protein crops and non-food crops constitute an increasing percentage of crop income among higher crop income earners, though the sample of farmers who grow nonfood crops at any income level is small.







Nutritional outcomes

For our outcomes of interest, we use children's anthropometric measurements to reflect their nutritional status. Because we use differences in income and anthropometry over one-year periods, we focus specifically on height, which is considered to be the best measure of long-term growth. According to the World Health Organization guidelines, we use height-for-age (HAZ) z-scores to normalize height measures by age in order to allow for useful comparisons across children of various ages.[17] [18]

Figure 4a presents kernel regressions of HAZ and WAZ (weight for age) on age in months for children between the ages of 6 and 59 months in the rural UNPS sample. Nutritional challenges are readily apparent: average height for age plunges quite steeply during the first 18 months and remains low. While both WAZ and HAZ are below international norms, it is striking that HAZ in particular is more than 1.5 standard deviations below international norms throughout childhood suggesting a diet that is less energy-deficient (captured by WAZ) and more nutrient deficient. Given the strong nonlinearities over time, in many specifications we will opt to include flexible controls for age (Cummins 2015).

To begin looking at the static relationship between income and nutritional status, Figure 4b presents kernel regressions of HAZ and WAZ on age in months split by median income.[19] The top two kernel density plots show weight for age z-scores, which stay relatively flat across ages. The higher plot marks children whose households are above the median income for the sample; the poorer half the sample tracks the same relatively flat trajectory but at a lower score. For height for age, both groups decline during the first two years, but the higher income half declines somewhat less dramatically. In the absence of omitted variables, a first glance would lead us to expect a strong correlation between income and nutritional status. Figure 5a shows HAZ by income shares by source. Households move toward the right on any curve if they specialize in that sector. The highest z-scores are among those most specialized in livestock, with non-agricultural self-employment and non-agricultural wage also looking favorable relative to crops and agricultural wages. These do not speak to changes in income nor total levels of income but might inform initial priors. Similarly, Figures 5c and 5d show z-scores by shares of crops by subtype as a proportion of gross total income. The lowest average z-scores are among those with the highest shares of income from low-protein crops. These figures, however, only describe anthropometric trends based on a static income profile and do not show that increases in low-protein crop income would lower z-scores.

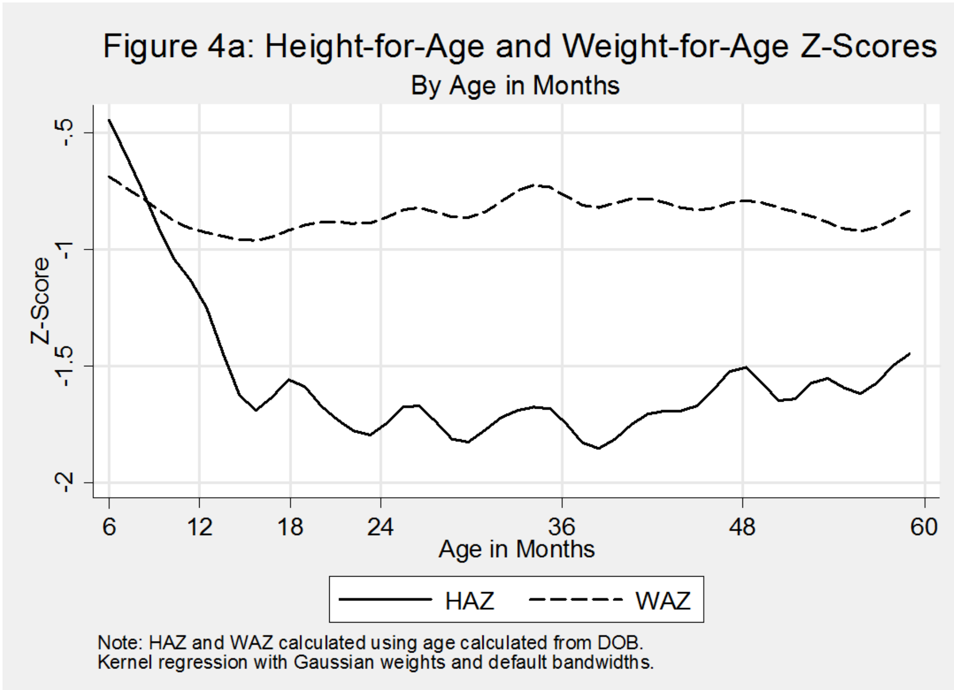
While our dataset is very rich in a number of ways, there are some missing elements that might allow for deeper analysis than what we are able to conduct presently. First, while there is a consumption module for the household, there is no individual-level consumption data. Given that young children are frequently fed a different mix of foods from the rest of the household, it is difficult to infer children's consumption from the household data. For this reason, data collected specifically for research on children's nutrition will collect information on children's diet specifically as well as on the frequency of feeding. There is also no module

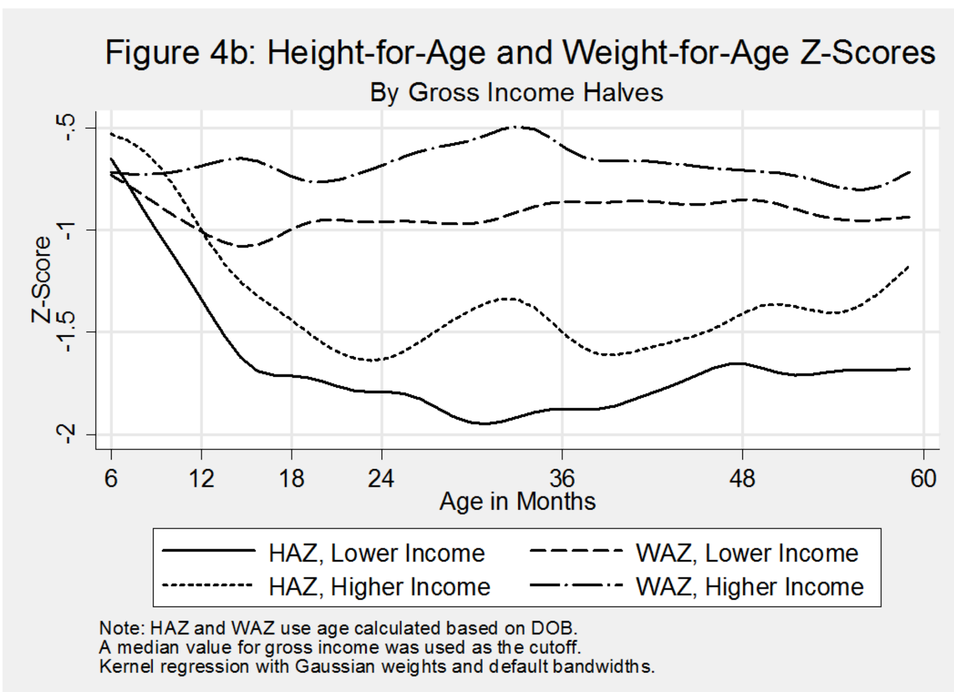
[17] According to WHO guidelines, we exclude children whose scores fall outside recommended cutoffs for plausible values: HAZ below -6 or above +6, WAZ below -6 or +5, WHZ below -5 or above +5. Source: http://www.who.int/nutgrowthdb/software/Differences_NCHS_WHO.pdf

[18] In all reported results, HAZ was calculated using reported date of birth when birth date and reported age in months did not correspond. HAZ using reported age in months generally yields similar results.

[19] Child age has been calculated taking into account the date of birth and the interview date, rather than the reported age in months, when these two measures did not match in the data.

on time use or other information that might be able to link income or labor activities more carefully to childcare tasks, another theoretical mechanism through which certain types of income could lead to differential outcomes for children.





Empirical Strategy

Using the subset of households in the UNPS with children under the age of five described above, we test the a set of hypotheses suggested by the theory laid out in the first sections, taking advantage of the child-level panel data. We show our most basic results on income without and then with child fixed effects, and then proceed with the preferred specification with child fixed effects.

We first look for evidence on the central question of whether short-term changes in income correspond to observable changes in nutritional outcomes. As a benchmark, we begin with the estimation most available in the literature, treating our panel dataset as a set of repeated cross sections. Thus, we estimate:

$$H_{it} = \beta_Y \log Y_{it} + \epsilon_{it},$$

where H_{it} represents the health measure (height for age (HAZ) z-score, height in centimeters, weight for age (WAZ) z-score, and weight in kilograms); Y_{it} is natural log of income, which can be specified various ways; and ϵ_{it} is the error term. We use the log of total gross income for interpretation in percentages and to accommodate diminishing marginal returns, after finding qualitatively similar results with a combination of level and square root of gross income. With this simple specification, there is a possibility of omitted variable bias from observable and unobservable characteristics that may influence both income and the anthropometric measures of children (parental education as one likely candidate). A vector of additional covariates Z_i may be added to capture such observable characteristics along with a survey round fixed effect η_t and seasonal (month) fixed effect s_t to absorb unobservable characteristics that are common to the sample in a particular survey round or a certain month of the year:

$$H_{it} = \beta_Y \log Y_{it} + Z_i\gamma + \eta_t + s_t + \epsilon_{it}.$$

Still, there are likely other unobservable (or simply unobserved) attributes that may bias the estimation results. One example might be tall parents whose height and strength increase wage earnings but also genetically predispose a child to attain greater height or weight than average. Another example could be related to parental intelligence, which can be used for earnings and for providing better care for children. In the absence of a set of convincing instruments for income variables, including a child-specific fixed effect in the specification allows us to control for time-invariant child, household, and community characteristics that might otherwise jointly determine income and child nutrition outcomes.

Thus, to test whether we observe changes in short-term changes in children's nutrition corresponding to short-term total income corresponding to short-term changes in children's nutrition, we estimate:

$$H_{it} = \beta_Y \log Y_{it} + Z_i\gamma + \eta_t + s_t + \nu_i + \epsilon_{it},$$

where H_{it} represents the health measure (height for age (HAZ) z-score, height in centimeters, weight for age (WAZ) z-score, and weight in kilograms), $\log Y_{it}$ is the log of total gross income

income, η_t is a survey round fixed effect, and ν_i is a child fixed effect, and ϵ_i is the error term. Here the vector of time-varying covariates Z_i is limited to include child age in months (with square and square root to accommodate the age trends with age fixed effects that allow a flexible specification by month [Cummins 2015]) and, female headship, number of children, and total household size. Even with child fixed effects, however, there remains a possibility of time-variant unobservable factors that may bias these estimates. Therefore, we see our estimation as a useful and informative diagnostic exercise but caution overly strong confidence in causal interpretations.

Income by sector

The next step is to test whether changes in height are related to the sector of income.[20] We conduct this analysis by looking for differential impacts of income from sectors vis-à-vis total income. We start by adding sector share of total income for the each of the four major income sectors individually: crops, livestock, nonagricultural self-employment, and wage income:

$$H_{it} = \beta_s(Y_{it}^s/Y_{it}) + \eta_t + \nu_i + \epsilon_i$$

where (Y_{it}^s/Y_{it}) is the share of gross income coming from sector as and $\log Y_{it}$ is log of total gross income. If consumption and nutrition are only related to income except for by through the budget constraint, we would expect to see no statistically significant coefficients for the sector indicators.

This specification may introduce omitted variable bias if changes different income shares are correlated to changes in total income, so we include a term for total income:

$$H_{it} = \beta_s(Y_{it}^s/Y_{it}) + \beta_Y \log Y_{it} + \eta_t + \nu_i + \epsilon_i.$$

We prefer this specification as the cleanest and most straightforward to interpret, as share of each sector compared to the sum of all other sectors. For robustness, however, we also consider an alternate specification that includes sector share for three of the four major income sectors at the same time: crops, livestock, nonagricultural self-employment. Using wage labor as the omitted (fourth) category, we estimate:

$$H_{it} = \Sigma_s\beta_s(Y_{it}^s/Y_{it}) + \beta_Y \log Y_{it} + \eta_t + \nu_i + \epsilon_i.$$

This specification will show different coefficients if changes in income are especially correlated between certain sectors more than others.

[20]Shares were chosen to prevent over-weighting of large-income households, who are likely to have larger nominal year-on-year fluctuations of reported income. Typically, logs would be used for this purpose, but shares allow us to keep households who do not have positive values for all income categories, where they are undefined as log(0).

Income by crop type

We test in the Ugandan context the idea commonly seen in the literature that type of crop income may influence nutrition. To do so, we break crops into three two broad categories based on nutrient availability: 1) low-protein, low-nutrient food crops and, 2) other food crops, and 3) non-food crops.[21] We adopt the same shares approach used for sector income, looking at the share of each crop type a in total income:

$$H_{it} = \beta_a(Y_{it}^a/Y_{it}) + \beta_Y \log Y_{it} + \eta_t + \nu_i + \epsilon_i$$

Again, we add controls for total income. For robustness checks, we also include specifications that add total crop share. It is important to note that the two crops we categorize as low-protein are banana/plantain and cassava – the top two staple crops in Uganda (and grown by more than 70% of our sample) with lower protein availability than other staples such as cereals and sweet potatoes. Uganda is recognized as maintaining a low-protein diet. As such, income gains that come in the form of additional low-protein food might be less likely to benefit nutrition that other forms of income unless the produce is sold to fund other nutrition-supporting purchases.[22]

Finally, we attempt to address the question of whether agriculture may be impactful through production alone or more specifically through own consumption. We do so by comparing the above specifications on shares of crop production to specifications that include shares of own consumption of production, represented by

$$H_{it} = \beta_c(C_{it}^a/Y_{it}) + \beta_Y \log Y_{it} + \eta_t + \nu_i + \epsilon_i,$$

with C_{it}^a representing the value of consumption originating from own production.

Results

Table 2 shows the results for overall income for height-for-age (HAZ) and prevalence of stunting (HAZ<-2), respectively. The first two columns each present HAZ z-scores as dependent variables, and the last two present stunting. Columns (1) and (3) show the standard pooled-cross sectional results with additional time-varying covariates and with standard errors clustered at the household level. These estimates represent the status quo for observational data where the same children cannot be followed over time. An income coefficient of 0.108 for HAZ is statistically significant but small, since we can interpret the coefficient to be the change in HAZ if household income doubles. For our sample, this only translates into approximately 0.08 standard deviations for a 100 percent gain in income. For the same change

[21] We started with three categories: 1) low-protein food crops, 2) other food crops, and 3) non-food crops, but found too few observations and too little variation for non-food crops to treat them separately

[22] While plantain is a source of Vitamin A and other micro-nutrients, it is are particularly low in protein, a necessary component for growth and development and most biological processes. Plantain and cassava have only a fraction of the protein content in cereals, which are also considered limited in their ability to meet preschoolers' protein needs without complementary higher-protein foods (FAO 1997, Ch. 7, available at http://www.fao.org/docrep/w0078e/w0078e08.htm)

in income, we predict that the prevalence of stunting would fall by 3.8 percentage points, or approximately 10 percent. Coefficients for the additional covariates are not reported, for brevity, but these variables include time indicators for Round 2 and Round 3, child gender (only relevant for the pooled cross-sectional estimations), age-in-months fixed effects to accommodate the common non-linear fall of z-scores over time (as observed in Figure 4), interview month fixed effects to reduce seasonally-based statistical noise in income reporting or child health, household size, number of children under 5 and under 15, household head years of education, and an identifier for and female headship.

The preferred fixed effects specifications are presented in columns (2) and (4) of Table 2 with the same set of controls. In contrast to the pooled cross-sectional results, the coefficients on income in fixed effects estimations fall by more than half for HAZ and to nearly zero for stunting. Though the cross-sectional effects were fairly small, we can infer that unobservable characteristics drive some of the relationship between income and anthropometric outcomes that we are able to control for by comparing children to themselves over time. These results hold for subsamples below the sample median income in 2009/10 (i.e. the first survey round), but for the subsample below 24 months of age, the income coefficient holds fairly steady for HAZ. Again, this coefficient is quite small, but it may suggest that short-term income gains may have a small nutrition-supporting role for younger children, who are still in the most critical period of nutritional development (UNICEF 1998, p. 21-23). Still, this effect is too small to suggest policy dependence on income growth to boost child growth measures. After this, all reported findings are based on regressions that control for child fixed effects and the aforementioned time-varying observables.

Table 2: Height-for-Age and Stunting on Total Income & Controls

Whole Sample (2325 observations in 757 clusters)

	HAZ		% Stunting	
	(1)	(2)	(3)	(4)
Ln Gross Total Income	0.108***	0.0473	-0.0376***	0.00243
	(0.0389)	(0.0352)	(0.0131)	(0.0113)
Child Fixed Effects	NO	YES	NO	YES
R-Sq	0.0926	0.229	0.0626	0.132
Adj R-Sq	0.0631	0.204	0.0322	0.104
Sub-sample: Under 24 Months (1069 obs in 496 clusters)				
	HAZ		% Stunted	
	(1)	(2)		
Ln Gross Total Income	0.119**	0.0908*	-0.0418***	-0.00129
	(0.0470)	(0.0541)	(0.0157)	(0.0163)
Child Fixed Effects	NO	YES	NO	YES
R-Sq	0.130	0.334	0.0846	0.162
Adj R-Sq	0.0909	0.304	0.0430	0.123
Sub-sample: Below Median Income in the First Year (1574 obs in 577 clusters)				
	HAZ		% Stunted	
	(1)	(2)	(3)	(4)
Ln Gross Total Income	0.0880*	0.0374	-0.0228	0.00601
	(0.0491)	(0.0362)	(0.0178)	(0.0132)
Child Fixed Effects	NO	YES	NO	YES
R-Sq	0.0930	0.310	0.0639	0.170
Adj R-Sq	0.0484	0.276	0.0179	0.129

Notes: Standard errors in parentheses, clustered at the household level; * p<0.1, ** p<0.05, *** p<.01; All regressions include the following time-varying controls: indicators for the survey rounds, household size, number of children under 15 and under 5, identifier for female headed household, head's education, survey month fixed effects, and age-in–months fixed effects.

Income by sector

Table 3a presents the results for income shares by sector with HAZ as the dependent variable across the whole sample. The crop income share without controlling for gross total income exerts a negative and significant impact, which declines in magnitude about 16 percent and becomes only marginally significant once gross total income is included. These differences seem to reflect the fact that crop income levels are fairly flat over the total income distribution, while crop income as a share of total income falls with total income. We can interpret the coefficients as being the expected change in HAZ score if a household went from no crop income to 100 percent crop income. In the more realistic change of 10-20 percent change in income share from year to year, this would translate to approximately 0.02 to 0.04 points of HAZ. On the other hand, the coefficients for self-employment in the same table are persistently positive and significant, and slightly larger in magnitude than the crop shares, potentially reflecting self-employment enabling different consumption or care habits, and meriting further investigation in the future. When controlling for income, both livestock and wage shares have coefficients near zero.

Table 3a: HAZ on Income Shares
(Shares of Gross Sector Income & Child Fixed Effects)
Whole Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% Crop	-0.256**	-0.215*						
	(0.114)	(0.118)						
% Livestock			-0.116	-0.0279				
			(0.129)	(0.128)				
% Self Employ					0.289**	0.255**		
					(0.127)	(0.124)		
% Wage							0.0797	-0.0133
							(0.118)	(0.128)
Ln Gross Total Income		0.0210		0.0463		0.0236		0.0481
		(0.0383)		(0.0358)		(0.0335)		(0.0389)
Joint P-value		0.0706		0.388		0.0906		0.371
R-Sq	0.186	0.231	0.182	0.229	0.186	0.232	0.182	0.229
Adj R-Sq	0.166	0.206	0.162	0.204	0.166	0.207	0.162	0.204

Notes: Standard errors in parentheses, clustered at the household level; * p<0.1, ** p<0.05, *** p<.01; 2325 observations in 757 clusters; Joint P-value is associated with the test of joint significance of the income shares and log gross total income; All regressions include the child fixed effects and the following time-varying controls: indicators for the survey rounds, household size, number of children under 15 and under 5, identifier for female headed household, head's education, survey month fixed effects, and age-in-months fixed effect.

Income by crop type

Table 3b present the results for crop production and consumption of own crop production individually as shares of gross total income for the whole sample. We present total crop production and consumption and then the breakdowns by crop category, with and without controlling for log of total income. The coefficients can be interpreted as the predicted change in z-score from a change from no income coming from that source to 100 percent coming from that source.

We see again that increased share of income coming from crop production corresponds to lower HAZ scores, with the coefficient dropping slightly and becoming only marginally statistically significant once total income is included. However, the coefficients for HAZ on the consumption of own crop production are and statistically significant and nearly twice the magnitude as for production. Thus, consumption of own production rather than production alone seems to drive the negative crop result in the context of Uganda. In fact, in the joint specification that includes both share of crops in income and share of consumption of own production (Table 3c, column 2), consumption of own crops share becomes more negative, while the crop production share changes signs and becomes small and positive and marginally significant. While this result needs more information about other mechanisms (food choice, care patterns) that accompany a shift toward more own-consumption, it suggests the possibility that households in rural Uganda could be a little bit better off converting produce into cash for nutrition-supporting purchases, or that there may be interventions that could help households who tend to consume their own production to protect young children from any associated nutritional disadvantages.

Production of the low protein crops shows a significant negative coefficient similar to consumption of own crop production, at 0.44 (column 6) predicted decrease from a shift from 0 to 100 percent plantain or cassava production, and the coefficient for consumption of own-produced low-protein crops is even more negative at -0.62 (column 8), suggesting that on average in the sample, crops may better serve long-term nutrition when converted to cash than through direct consumption. Supporting these results, Table 3c includes multiple share types. Column 4 combines crop share and low protein crop share. The magnitude of low protein crops holds but falls to marginal significance (joint p-value 0.056), and the magnitude coefficient for crop share shrinks to close to zero. With crop share, low protein production, and low protein consumption (column 6), crop share remains small and negative, low protein production share becomes positive, and low protein consumption share becomes even more negative at -0.72 (though not statistically different) than when included alone. For the extreme case of a child whose household's whole income comes from production of low-protein crops, this result predicts approximately 0.5 standard deviations difference in HAZ between consuming all of that produce and selling all of it.

Underlying these results, there is a 0.78 versus 0.45 correlation between consumption and production of low protein versus other food crops (not shown), respectively, pointing to income growth in the form of cassava and plantain production being particularly unlikely to convert into consumption of other foods or nonfoods. These anthropometric score results suggest that this "stickiness" of crop production to own consumption that may, in a context with low-protein staple crops, may potentially render agricultural growth less beneficial for nutrition than other types of growth, and that in Uganda, all else equal, there are likely to be nutritional gains from shifting toward more nutrient-rich crop production.[23]

[23] There is some concerns that high shares of own-produced starches would be indicate household poverty, and that the results could be driven by nonlinearities in responsiveness to income. Because we changes in shares, however, we capture a whole range of baseline shares. Also, in choosing log income, we tested a number of more flexible forms, but none appeared to offer improvements over log income.

Table 3b: HAZ - Crop Production and Own Consumption
(Shares of Gross Total Income & Child Fixed Effects)
Whole Sample (2325 observations in 757 clusters)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% Income, Crops	-0.256**	-0.215*						
	(0.114)	(0.118)						
% Income, Consumption of Own Crops			-0.434***	-0.440**				
			(0.154)	(0.171)				
% Income, Low Protein Crop Production					-0.421***	-0.442**		
					(0.193)	(0.189)		
% Income, Consumption of Low Protein Crop Production							-0.611***	-0.619***
							(0.225)	(0.225)
Ln Gross Total Income		0.0210		0.00187		0.0304		0.0214
		(0.0383)		(0.0415)		(0.0363)		(0.0371)
Joint P-value		0.0706		0.00943		0.0238		0.00765
R-Sq	0.186	0.231	0.188	0.235	0.186	0.234	0.189	0.236
Adj R-Sq	0.166	0.206	0.169	0.209	0.166	0.209	0.169	0.211

Notes: Standard errors in parentheses, clustered at the household level; * p<0.1, ** p<0.05, *** p<.01; Joint P-value is associated with the test of joint significance of the income shares and log gross total income; All regressions include the child fixed effects and the following time-varying controls: indicators for the survey rounds, household size, number of children under 15 and under 5, identifier for female headed household, head's education, survey month fixed effects, and age-in-months fixed effects.

Table 3c: HAZ - Crop Production and Own Consumption
(Shares of Gross Total Income & Child Fixed Effects)
Whole Sample (2325 observations in 757 clusters)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% Income, Crops	0.0415	0.0899*	-0.134	-0.0284	-0.154	-0.0673	-0.0475	0.0348
	(0.180)	(0.180)	(0.139)	(0.142)	(0.139)	(0.141)	(0.191)	(0.186)
% Income, Consumption of Own Crops	-0.478*	-0.530**					-0.226	-0.244
	(0.245)	(0.265)					(0.261)	(0.274)
% Income, Low Protein Crop Production			-0.287	-0.416*	0.352	0.159	0.335	0.156
			(0.238)	(0.232)	(0.318)	(0.314)	(0.314)	(0.310)
% Income, Consumption of Low Protein Crop Production					-0.810***	-0.721***	-0.678**	-0.590*
					(0.337)	(0.343)	(0.344)	(0.335)
Ln Gross Total Income		0.00357		0.0279		0.0304		0.00771
		(0.0411)		(0.0383)		(0.0363)		(0.0418)
Joint P-value		0.0239		0.0562		0.0449		0.0570
R-Sq	0.188	0.235	0.187	0.234	0.190	0.237	0.191	0.237
Adj R-Sq	0.168	0.209	0.167	0.208	0.170	0.211	0.170	0.211

Notes: Standard errors in parentheses, clustered at the household level; * p<0.1, ** p<0.05, *** p<.01; Joint P-value is associated with the test of joint significance of the income shares and log gross total income; All regressions include the child fixed effects and the following time-varying controls: indicators for the survey rounds, household size, number of children under 15 and under 5, identifier for female headed household, head's education, survey month fixed effects, and age-in-months fixed effects

Alternative Specifications for Robustness

Tables 3c and 3d provides alternate specifications, including multiple types of shares. As discussed above, Table 3c combines multiple types of crop shares, and Table 3d includes sector shares and crop-type shares, with wage income always as the omitted category. We see that the results with these combined shares are largely consistent with the individual specification in magnitude, though significance levels fluctuate in some specifications. Overall, it does not appear that underlying correlations between share types are driving our results.

Table 3d: HAZ on Income Shares
(Shares of Gross Sector Income & Child Fixed Effects)
Whole Sample (2325 observations in 757 clusters)

	(1)	(2)	(3)	(4)	(5)	(6)
% Crop	-0.179	-0.166	0.000390	0.0265	-0.0121	-0.0101
	(0.119)	(0.141)	(0.151)	(0.170)	(0.149)	(0.168)
% Livestock	-0.0430	-0.0350	-0.0521	-0.0375	-0.0366	-0.0355
	(0.153)	(0.123)	(0.152)	(0.162)	(0.151)	(0.162)
% Self Employ	0.193	0.193	0.203	0.199	0.206	0.206
	(0.156)	(0.156)	(0.160)	(0.157)	(0.160)	(0.157)
% Crop: Low-protein			-0.419*	-0.428*	0.162	0.159
			(0.233)	(0.232)	(0.310)	(0.314)
% Crop: Low-protein Home Consumption					-0.738**	-0.736**
					(0.336)	(0.345)
Ln Gross Total Income		0.00781		0.0144		0.00108
		(0.0391)		(0.0393)		(0.0402)
Joint P-value		0.111		0.0798		0.0519
Joint P-value (shares)	0.0606	0.0987	0.0432	0.0779	0.0294	0.0541
R-Sq	0.229	0.229	0.236	0.236	0.239	0.239
Adj R-Sq	0.205	0.204	0.210	0.210	0.213	0.212

Notes: Standard errors in parentheses, clustered at the household level; * p<0.1, ** p<0.05, *** p<.01; Joint P-value is associated with the test of joint significance of the income shares and log gross total income; Joint P-value (shares) is associated with the test of joint significance of the income shares only. All regressions include the child fixed effects and the following time-varying controls: indicators for the survey rounds, household size, number of children under 15 and under 5, identifier for female headed household, head's education, survey month fixed effects, and age-in–months fixed effects.

In order to explore price versus quantity underlying these results, I also add a joint specifications using crop income reconstructed using the same prices across all three years instead of letting them vary by year, so that changes in the various income shares are driven by changes in quantities rather than prices. The original approach (represented in Table 3d), valuing crop production and own consumption at market prices, was employed because it captures the opportunity cost of any own consumption. A limitation of the original specification, however, is that by applying changing prices to the portion of production that is consumed by the household, calculated income can increase even when the physical quantity consumed stays the same or even falls. In this case, changes in the income value assigned to own consumption may be divorced from the nutritional value of own consumption, because the assigned value of consumption can increase with quantity and/or price. In particular, we are looking to see whether the alternative price-constant specification gives non-negative coefficients for crop income and subcategories of crop income. If production or consumption prove neutral or beneficial for child height, then we may infer that the negative correspondence in the original specification is likely driven by higher prices masking lower real consumption.

and then the revised (3e) below. The first two columns that only use the broad sectors (no subsectors of crops) with and without log total income are very similar between the two approaches, suggesting that overall, it does not matter much whether crop income changes because of changes in income or in price in out context. Columns 3 and 4, which add the share of income coming from low-protein crops, are again fairly similar in magnitude, but low-protein-specific share is a little less negative and overall crop share is a little more negative (not statistically different). The shares in columns 2 and 4 are jointly marginally statistically significant at 10%. Columns 5 and 6, which add the share of own-consumed low-protein crops, deviate the most from the original specification, though qualitatively they still reflect the original results. The magnitude of the coefficients on low-protein own consumption falling by half from -0.7 to -0.3 but still remain negative, and the overall coefficient for low-protein share again shrinks once consumption is included. The summed negative effect for own-consumption of low-protein does not show a big change, but now a little more of the negative correlation remains in the overall crop and overall low protein production. The joint p-value rises from 0.05 to 0.11.

Overall, the changes between these results, or general lack thereoff, suggests that the original results are not primarily driven by higher prices masking lower consumption. That the negative coefficient from primarily own consumption of low-protein crops is partially recaptured by overall production and low-protein production is perhaps a little surprising and may point to a limitation of the second approach. Applying the same set of prices to all the years better captures the changes in quantities produced and consumed between years. It is also limited, however, in that it assumes that changing quantities continue to have the same market value across years. This assumption ignores real income fluctations realized through crop sales. In general, if prices tend to fall when quantities rise, then by holding prices constant we may overestimate crop income (via sales) and total income in quantity-abundant years. If higher crop shares bias calculated income upward, we might expect to see slightly more negative coefficients for crops that include sales, which may help reconcile the two sets of results.[24]

[24]In practice, I cannot offer evidence about prices falling under abundant supply, even though this is a guiding principle in markets. I don't see that relationship very well in the data, but there are a lot of unobserved quality and market mechanisms that may be at play.

Table 3e: HAZ on Income Shares with Fixed Crop Prices
(Shares of Gross Sector Income & Child Fixed Effects)
Whole Sample (2325 observations in 757 clusters)

	(1)	(2)	(3)	(4)	(5)	(6)
% Crop	-0.223*	-0.202	-0.104	-0.0682	-0.107	-0.0803
	(0.119)	(0.136)	(0.142)	(0.158)	(0.142)	(0.159)
% Livestock	-0.0607	-0.0469	-0.0671	-0.0479	-0.0599	-0.0463
	(0.153)	(0.163)	(0.153)	(0.163)	(0.153)	(0.163)
% Self Employ	0.156	0.153	0.161	0.157	0.162	0.159
	(0.160)	(0.157)	(0.160)	(0.157)	(0.160)	(0.157)
% Crop: Low-protein			-0.313	-0.326	-0.0564	-0.0905
			(0.218)	(0.217)	(0.348)	(0.350)
% Crop: Low-protein Home Consumption					-0.323	-0.293
					(0.355)	(0.360)
Ln Gross Total Income		0.0139		0.0195		0.0146
		(0.0390)		(0.0392)		(0.0399)
Joint P-value		0.0833		0.0736		0.112
Joint P-value (shares)	0.0430	0.0877	0.0398	0.0860	0.0673	0.135
R-Sq	0.236	0.237	0.238	0.238	0.238	0.238
Adj R-Sq	0.211	0.211	0.212	0.212	0.212	0.212

Notes: Standard errors in parentheses, clustered at the household level; * p<0.1, ** p<0.05, *** p<.01; Joint P-value is associated with the test of joint significance of the income shares and log gross total income; Joint P-value (shares) is associated with the test of joint significance of the income shares only. All regressions include the child fixed effects and the following time-varying controls: indicators for the survey rounds, household size, number of children under 15 and under 5, identifier for female headed household, head's education, survey month fixed effects, and age-in–months fixed effects.

Subsamples by age and median income

Tables 4a-c, 5a-c, and 6a-c present the results for the same equations as 3a-c, this time broken down by subsamples of under 24 months of age in the first year, over 24 months in first year, and below the median income at baseline, respectively. When the results are broken down by age group, we can that the magnitudes for the crop shares are larger and more significant for the older group (Tables 5b and 5c) – this appears to be the group that drives the results from the full sample. This may be surprising, since the largest losses in height are expected to occur during the first years that are considered more critical for nutrition. On the other hand, we may infer that effects on the younger children may be partially mitigated by breastfeeding or by different levels of care. We do see that the younger group continues to have a slightly larger magnitude for the coefficient on total income (still not significant). One possible explanation is that because a portion of the younger group is still breastfeeding, complementary feeding with low-protein foods or any kind of food contributes needed supplementary carbohydrates, while for older children, consumption of low-nutrient foods displaces more nutrient-dense consumption. Breastfeeding could also affect level of adult supervision, which may influence total amounts consumed. Given the lack of data on children's individual consumption or care, however, it is not possible to distinguish these or other mechanisms. For children whose household income was below the median value in the first year, the results generally follow the overall sample results with a few differences. Across sectors, there appears to be little premium for self-employment but a larger gain in wage labor. Looking at the two specifications for wage labor with and without total income, the magnitude falls once total income is included. We speculate that for poorer households, self-employment may more often be an option of last resort or may require costly inputs to be more profitable, and wage income may provide a more available or more predictable income during lean periods. For crops, the "penalty" for own consumption, especially of the low-protein crops, appears to grow in magnitude relative to production for sale. It is difficult to say without more information, but it is possible that this subset of children has a less nutritious diet at the start, such that marginal food consumption could be even more important.

Table 4a: HAZ on Income Shares
(Shares of Gross Sector Income & Child Fixed Effects)
Sub-sample: Under 24 Months

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% Crop	-0.306	-0.153						
	(0.215)	(0.221)						
% Livestock			-0.343	-0.226				
			(0.243)	(0.218)				
% Self Employ					0.394*	0.346*		
					(0.223)	(0.194)		
% Wage							0.159	-0.0317
							(0.193)	(0.222)
Ln Gross Total Income		0.0652		0.0790		0.0501		0.0878
		(0.00760)		(0.0660)		(0.0605)		(0.0728)
Joint P-value		0.269		0.216		0.154		0.340
R-Sq	0.286	0.394	0.286	0.394	0.289	0.397	0.283	0.393
Adj R-Sq	0.259	0.359	0.258	0.359	0.261	0.362	0.256	0.358

Notes: Standard errors in parentheses, clustered at the household level; * $p<0.1$, ** $p<0.05$, *** $p<.01$; 1069 observations in 496 clusters; Joint P-value is associated with the test of joint significance of the income shares and log gross total income; All regressions include the child fixed effects and the following time-varying controls: indicators for the survey rounds, household size, number of children under 15 and under 5, identifier for female headed household, head's education, survey month fixed effects, and age-in-months fixed effects.

Table 4b: HAZ - Crop Production and Own Consumption
(Shares of Gross Total Income & Child Fixed Effects)
Sub-sample: Under 24 Months in First Year (1069 observations in 496 clusters)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% Income, Crops	-0.306	-0.153						
	(0.215)	(0.221)						
% Income, Consumption of Own Crops			-0.353	-0.156				
			(0.270)	(0.303)				
% Income, Low Protein Crop Production					-0.270	-0.320		
					(0.335)	(0.313)		
% Income, Consumption of Low Protein Crop Production							-0.441	-0.368
							(0.365)	(0.350)
Ln Gross Total Income		0.0652		0.0683		0.0721		0.0669
		(0.00760)		(0.0810)		(0.0676)		(0.0694)
Joint P-value		0.269		0.280		0.224		0.206
R-Sq	0.286	0.394	0.286	0.393	0.284	0.395	0.286	0.395
Adj R-Sq	0.259	0.359	0.258	0.359	0.256	0.360	0.258	0.360

Notes: Standard errors in parentheses, clustered at the household level; * p<0.1, ** p<0.05, *** p<.01; Joint P-value is associated with the test of joint significance of the income shares and log gross total income; All regressions include the child fixed effects and the following time-varying controls: indicators for the survey rounds, household size, number of children under 15 and under 5, identifier for female headed household, head's education, survey month fixed effects, and age-in-months fixed effects.

Table 4c: HAZ - Crop Production and Own Consumption
(Shares of Gross Total Income & Child Fixed Effects)
Sub-sample: Under 24 Months (1069 observations in 496 clusters)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% Income, Crops	-0.209	-0.133	0.0360	0.0215	-0.347	-0.00137	-0.355	-0.138
	(0.315)	(0.291)	(0.421)	(0.249)	(0.264)	(0.253)	(0.338)	(0.314)
% Income, Consumption of Own Crops	-0.143	-0.0312					0.0163	0.330
	(0.401)	(0.414)					(0.423)	(0.435)
% Income, Low Protein Crop Production			-0.257	-0.338	0.764	-0.0966	0.763	-0.149
			(0.338)	(0.366)	(0.474)	(0.415)	(0.477)	(0.417)
% Income, Consumption of Low Protein Crop Production					-0.861*	-0.277	-0.868*	-0.397
					(0.343)	(0.417)	(0.458)	(0.424)
Ln Gross Total Income		0.0643		0.0743		0.0673		0.0777
		(0.0806)		(0.0762)		(0.0796)		(0.0847)
Joint P-value		0.445		0.391		0.522		0.726
R-Sq	0.286	0.394	0.286	0.395	0.289	0.395	0.289	0.396
Adj R-Sq	0.258	0.358	0.258	0.359	0.260	0.359	0.260	0.395

Notes: Standard errors in parentheses, clustered at the household level; * p<0.1, ** p<0.05, *** p<.01; Joint P-value is associated with the test of joint significance of the income shares and log gross total income; All regressions include the child fixed effects and the following time-varying controls: indicators for the survey rounds, household size, number of children under 15 and under 5, identifier for female headed household, head's education, survey month fixed effects, and age-in-months fixed effects.

Table 5a: HAZ on Income Shares
(Shares of Gross Sector Income & Child Fixed Effects)
Sub-sample: Over 24 Months in First Year (1421observations in 600 clusters)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% Crop	-0.161	-0.113						
	(0.113)	(0.121)						
% Livestock			-0.0404	-0.0227				
			(0.125)	(0.129)				
% Self Employ					0.188	0.140		
					(0.128)	(0.131)		
% Wage							0.0355	-0.0361
							(0.108)	(0.111)
Ln Gross Total Income		0.0146		0.0278		0.0164		0.0295
		(0.0345)		(0.0318)		(0.0329)		(0.0327)
Joint P-value		0.418		0.681		0.370		0.666
R-Sq	0.0750	0.129	0.0723	0.127	0.0755	0.129	0.0722	0.128
Adj R-Sq	0.0503	0.0933	0.0474	0.0.0916	0.0508	0.0933	0.0474	0.0917

Notes: Standard errors in parentheses, clustered at the household level; * p<0.1, ** p<0.05, *** p<.01; 1069 observations in 496 clusters; Joint P-value is associated with the test of joint significance of the income shares and log gross total income; All regressions include the child fixed effects and the following time-varying controls: indicators for the survey rounds, household size, number of children under 15 and under 5, identifier for female headed household, head's education, survey month fixed effects, and age-in-months fixed effects.

Table 5b: HAZ - Crop Production and Own Consumption
(Shares of Gross Total Income & Child Fixed Effects)
Sub-sample: Over 24 Months in First Year (1421observations in 600 clusters)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	HAZ	HAZ	HAZ	HAZ	HAZ	HAZ	HAZ	HAZ
% Income, Crops	-0.161	-0.113						
	(0.113)	(0.121)						
% Income, Consumption of Own Crops			-0.386**	-0.378**				
			(0.151)	(0.176)				
% Income, Low Protein Crop Production					-0.400**	-0.389**		
					(0.191)	(0.184)		
% Income, Consumption of Low Protein Crop Production							-0.613***	-0.615***
							(0.206)	(0.210)
Ln Gross Total Income		0.0146		-0.00966		0.0141		0.00535
		(0.0345)		(0.0383)		(0.0317)		(0.0324)
Joint P-value		0.418		0.0479		0.0653		0.00751
R-Sq	0.0750	0.129	0.0815	0.134	0.0793	0.134	0.0841	0.0138
Adj R-Sq	0.0503	0.093	0.0569	0.0989	0.0547	0.0982	0.0596	0.103

Notes: Standard errors in parentheses, clustered at the household level; * p<0.1, ** p<0.05, *** p<.01; Joint P-value is associated with the test of joint significance of the income shares and log gross total income; All regressions include the child fixed effects and the following time-varying controls: indicators for the survey rounds, household size, number of children under 15 and under 5, identifier for female headed household, head's education, survey month fixed effects, and age-in-months fixed effects.

Table 5c: HAZ - Crop Production and Own Consumption
(Shares of Gross Total Income & Child Fixed Effects)
Sub-sample: Over 24 Months in First Year (1421observations in 600 clusters)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% Income, Crops	0.188	0.211	0.00621	0.0935	-0.0072	-0.0675	0.154	0.198
	(0.210)	(0.211)	(0.132)	(0.148)	(0.132)	(0.147)	(0.214)	(0.213)
% Income, Consumption of Own Crops	-0.585**	-0.594**					-0.338	-0.316
	(0.280)	(0.299)					(0.318)	(0.340)
% Income, Low Protein Crop Production			-0.406*	-0.479**	0.132	0.0598	0.0867	0.0345
			(0.225)	(0.224)	(0.414)	(0.414)	(0.409)	(0.409)
% Income, Consumption of Low Protein Crop Production					-0.742*	-0.735*	-0.520	-0.543
					(0.431)	(0.437)	(0.434)	(0.434)
Ln Gross Total Income		0.00786		0.0213		0.0104		0.0000
		(0.0380)		(0.0341)		(0.0351)		(0.0387)
Joint P-value		0.0766		0.112		0.522	0.0498	0.0640
R-Sq	0.0829	0.136	0.286	0.134	0.0843	0.139	0.0861	0.140
Adj R-Sq	0.0577	0.100	0.258	0.0980	0.0584	0.102	0.0596	0.103

Notes: Standard errors in parentheses, clustered at the household level; * p<0.1, ** p<0.05, *** p<.01; Joint P-value is associated with the test of joint significance of the income shares and log gross total income; All regressions include the child fixed effects and the following time-varying controls: indicators for the survey rounds, household size, number of children under 15 and under 5, identifier for female headed household, head's education, survey month fixed effects, and age-in-months fixed effects.

Table 6a: HAZ on Income Shares
(Shares of Gross Sector Income & Child Fixed Effects)
Sub-sample: Below Median Income in the First Year (1574 observations in 577 clusters)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% Crop	-0.191*	-0.101						
	(0.137)	(0.133)						
% Livestock			-0.222	-0.151				
			(0.157)	(0.150)				
% Self Employ					0.0746	0.00294		
					(0.135)	(0.136)		
% Wage							0.405***	0.275**
							(0.143)	(0.131)
Ln Gross Total Income		0.0226		0.0356		0.0371		0.0181
		(0.0417)		(0.0365)		(0.0375)		(0.0381)
Joint P-value		0.420		0.371		0.588		0.0532
R-Sq	0.206	0.311	0.205	0.311	0.203	0.310	0.211	0.313
Adj R-Sq	0.176	0.276	0.176	0.277	0.174	0.276	0.182	0.279

Notes: Standard errors in parentheses, clustered at the household level; * p<0.1, ** p<0.05, *** p<.01; Joint P-value is associated with the test of joint significance of the income shares and log gross total income; All regressions include the child fixed effects and the following time-varying controls: indicators for the survey rounds, household size, number of children under 15 and under 5, identifier for female headed household, head's education, survey month fixed effects, and age in months fixed effect.

Table 6b: HAZ - Crop Production and Own Consumption
(Shares of Gross Total Income & Child Fixed Effects)
Sub-sample: Below Median Income in the First Year (1574 observations in 577 clusters)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% Income, Crops	-0.191*	-0.101						
	(0.137)	(0.133)						
% Income, Consumption of Own Crops			-0.382**	-0.338**				
			(0.178)	(0.189)				
% Income, Low Protein Crop Production					-0.611	-0.211		
					(0.211)	(0.184)		
% Income, Consumption of Low Protein Crop Production							-0.451**	-0.338*
							(0.241)	(0.18218)
Ln Gross Total Income		0.0226		0.00367		0.0295		0.0226
		(0.0417)		(0.0433)		(0.0372)		(0.0376)
Joint P-value		0.420		0.103		0.291		0.131
R-Sq	0.206	0.311	0.210	0.314	0.205	0.311	0.208	0.313
Adj R-Sq	0.176	0.276	0.181	0.280	0.176	0.277	0.179	0.279

Notes: Standard errors in parentheses, clustered at the household level; * p<0.1, ** p<0.05, *** p<.01; Joint P-value is associated with the test of joint significance of the income shares and log gross total income; All regressions include the child fixed effects and the following time-varying controls: indicators for the survey rounds, household size, number of children under 15 and under 5, identifier for female headed household, head's education, survey month fixed effects, and age in months fixed effect.

Table 6c: HAZ - Crop Production and Own Consumption
(Shares of Gross Total Income & Child Fixed Effects)
Sub-sample: Below Median Income in the First Year (1574 observations in 577 clusters)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	HAZ	HAZ	HAZ	HAZ	HAZ	HAZ	HAZ	HAZ
% Income, Crops	0.202	0.271	-0.133	-0.0046	-0.148	-0.0396	0.0901	0.187
	(0.219)	(0.209)	(0.157)	(0.156)	(0.156)	(0.158)	(0.237)	(0.225)
% Income, Consumption of Own Crops	-0.590**	-0.603*					-0.458	-0.507
	(0.291)	(0.296)					(0.339)	(0.352)
% Income, Low Protein Crop Production			-0.138	-0.207	0.684*	0.456	0.645*	0.457
			(0.244)	(0.218)	(0.361)	(0.343)	(0.349)	(0.332)
% Income, Consumption of Low Protein Crop Production					-0.997***	-0.790**	-0.744*	-0.545
					(0.381)	(0.373)	(0.399)	(0.386)
Ln Gross Total Income		0.00353		0.0290		0.0168		-0.0018
		(0.0433)		(0.0418)		(0.0426)		(0.0446)
Joint P-value		0.131		0.480		0.204		0.171
R-Sq	0.210	0.315	0.206	0.311	0.211	0.314	0.213	0.317
Adj R-Sq	0.181	0.281	0.176	0.277	0.181	0.279	0.183	0.281

Notes: Standard errors in parentheses, clustered at the household level; * p<0.1, ** p<0.05, *** p<.01; Joint P-value is associated with the test of joint significance of the income shares and log gross total income; All regressions include the child fixed effects and the following time-varying controls: indicators for the survey rounds, household size, number of children under 15 and under 5, identifier for female headed household, head's education, survey month fixed effects, and age-in-months fixed effects.

Discussion

In this study, we have used panel data to explore the potential for short-term income gains to improve children's nutrition in Uganda. The high frequency in the UNPS—three rounds in three years—allow for fixed effect estimations that eliminate time-invariant unobservable factors that may typically confound similar studies. Without convincing instrumental variables for income, one still must take care with causal interpretations, but such exercises can serve as informative diagnostic tools for policy and future research.

Our results show very little relationship between changes in income levels and height, considered a good marker of long-term nutrition. Contrary to the benchmark case of in which income generation might be uniformly good for nutrition, however, increased self-employment income appears correlate relatively more with positive nutritional outcomes than other sectors. Future studies distinguishing the mechanisms behind this trend, potentially dietary or related to proximity to home for childcare, would be worthwhile.

Specific to Uganda, we find the potentially less-expected result that agricultural income appears to be more nutrition-negative than others, seemingly through the production of low-nutrient crops and specifically through own consumption. Our results suggest the possibility of stickiness of crop production to own consumption; while this may be a nutrition-supporting feature in other contexts, income growth in the production of low-nutrient crops may crowd out consumption of other goods and services that could serve as better nutritional investments. These results appear to be concentrated among the older and poorer subset of children in our sample. Still, any effects appear to be relatively small in magnitude, especially given that year-on-year changes in income for households usually fall within limited bounds. These results are also likely to depend heavily on the agricultural and dietary profile of Uganda and caution against uniform policies to support one sector over another without further information specifically about how children's diets or childcare patterns accompany income changes. When data are available, similar diagnostic techniques may be useful for identifying which sectors and interventions may be most nutrition-supporting in other contexts, but would be best complemented with more details about household practices for child feeding and care.

Acknowledgements

The authors would like to thank Luc Christiaensen, Will Martin, Emmanuel Skoufias, Eduardo Montoya, Stefania di Giuseppe, Aminata Bakouan, Jonathan Kaminsky, a set of anonymous reviewers. and members of the development workshop in UCB's department of Agricultural and Resource Economics for their comments. All remaining errors are our own.

References

Abbi, R., Christian, P., Gujral, S., and Gopaldas, T. 1991. "The impact of maternal work on the nutrition and health status of children." Food and Nutrition Bulletin, 13.1, pp. 20 –24.

Arimond, M., Hawkes, C., Ruel, M.T., Sifri, Z., Berti, P.R., Leroy, J.L., Low, J.W., Brown, L.R. and Frongillo, E.A. 2011. "Agricultural interventions and nutrition: lessons from the past and new evidence." In B. Thompson, and L. Amoroso (Eds.), Combating micronutrient deficiencies: food-based approaches, Rome: Food and Agriculture Organization of the United Nations (FAO).

Behrman, J. R., and Wolfe, B. L. 1984. “More evidence on nutrition demand: Income seems overrated and women’s schooling underemphasized.” Journal of Development Economics, 14.1, pp. 105–128.

Behrman, J.R. and Deolalikar, A.B. 1988. “Health and Nutrition.” In H. Chenery and T. N. Srinivasan (Eds.), Handbook of development economics. Amsterdam: North Holland.

Berti, P.R., Krasevec, J. and Fitzgerald, S. 2004. “A review of the effectiveness of agriculture interventions in improving nutrition outcomes.” Public Health Nutrition, 7, pp. 599–609.

Bouis, H. and Haddad, L. 1990. “Effects of agricultural commercialization on land tenure, household resource allocation, and nutrition in the Philippines.” International Food Policy Research Institute (IFPRI) Research Report 79.

Chen, S., and Ravallion, M. 2007. “China’s (uneven) progress against poverty.” Journal of Development Economics, 82.1, pp. 1–42.

Christiaensen, L., Demery, L., Kuhl, J. 2011. “The (evolving) role of agriculture in poverty reduction—an empirical perspective.” Journal of Development Economics, 96.2, pp. 239–254.

Cole, T.J. 2003. “The secular trend in human physical growth: a biological view.” Economics and Human Biology, 1, pp. 161–168.

Cummins, J. 2015. “On the Use and Misuse of Child Height-for-Age Z-score in the Demographic and Health Surveys.” Working paper.

de Janvry, A., Fafchamps, M. and Sadoulet, E. 1991. ”Peasant household behavior with missing markets: some paradoxes explained.” The Economic Journal, 101, pp. 1400–1417.

de Janvry, A., and Sadoulet, E. 1995. Quantitative development policy analysis. Baltimore, MD: The John Hopkins University Press.

de Janvry, A., and Sadoulet, E. 2010. “Agricultural growth and poverty reduction: additional evidence.” The World Bank Observer, 25.1, pp. 1–20.

Deaton, A. and Drèze, J. 2009. “Food and nutrition in India: facts and interpretations.” Economic and Political Weekly, 44.7, pp. 42–65.

DeWalt, K. 1993. “Nutrition and the commercialization of agriculture: ten years later.” Social Science & Medicine, 36, pp. 1407–1416.

Dewey, K. G. 1981. “Nutritional consequences of the transformation from subsistence to commercial agriculture in Tabasco, Mexico.” Human Ecology, 9, pp. 151–187.

FANTA/USAID 2010. “The analysis of the nutrition situation in Uganda.” Retrieved from http://www.fantaproject.org/sites/default/files/resources/Uganda_NSA_May2010.pdf.

FAO 1990. “Roots, tubers, plantains and bananas in human nutrition.” Retrieved from http://www.fao.org/docrep/t0207e/t0207e05.htm.

FAO 1997. Agriculture, food, and nutrition for Africa – A resource book for teachers of agriculture. Retrieved from http://www.fao.org/docrep/w0078e/w0078e00.HTM.

FEWS NET. 2011. “Uganda food security outlook update.” Retrieved from:
http://www.fews.net/sites/default/files/documents/reports/Uganda_FSOU_February_2011_02_final.pdf

Fleuret, P. and Fleuret, A. 1980. “Nutrition, consumption, and agricultural change.” Human Organization, 39, pp. 250–260.

Gillespie, S., and Mason, J. 1994. “Controlling vitamin A deficiency.” UN Administrative Committee on Coordination Subcommittee on Nutrition, Nutrition Policy Discussion Paper No. 14.

Haddad, L. and Smith, L.C. 2002. “How potent is economic growth in reducing undernutrition? What are the pathways of impact? New cross-country evidence.” Economic Development and Cultural Change, 51, pp. 55-76.

Hall, R., and Mishkin, F. 1982. “The sensitivity of consumption to transitory income: estimates from panel data on households.” NBER Working Paper No. 505.

Headey, D. 2013. “Developmental drivers of nutritional change: a cross-country analysis.” World Development, 42.C, pp. 76–88.

Headey, D., Chiu, A., and Kadiyala, S. 2012. “Agriculture’s role in the Indian enigma: help or hindrance to the malnutrition crisis?” Food Security, 4.1, pp. 87–102.

Heltberg, R. 2009. “Malnutrition, poverty, and economic growth.” Health Economics, 18.1, pp. 77–88.

Herforth, A. 2013. Synthesis of Guiding Principles on Agriculture Programming for Nutrition. Rome: FAO.

Jensen, R. 2010. “Information, efficiency and welfare in agricultural markets.” Agricultural Economics, 41.S1, pp. 203–216.

Kennedy, E. 1994. “Health and nutrition effects of commercialization of agriculture.” In J. von Braun and E. Kennedy (Eds.), Agricultural commercialization, economic development, and nutrition (pp. 79-99). Baltimore, Maryland: Johns Hopkins University Press.

Kennedy, E., Bouis, H., and von Braun, J. 1992. “Health and nutrition effects of cash-crop production in developing countries: a comparative analysis.” Social Science & Medicine, 35, pp. 689–697.

Kennedy, E. and Cogill, B. 1987. “Income and nutritional effects of the commercialization of agriculture in southwestern Kenya.” IFPRI Research Report No. 63.

Key, N., Sadoulet, E., and de Janvry, A. 2000. “Transaction costs and agricultural household supply response.” American Journal of Agricultural Economics, 82.2, pp. 245–259.

Kurz, K. and Johnson-Welch, C. 2007. “Enhancing women’s contributions to improving family food consumption and nutrition.” Food and Nutrition Bulletin, 22, pp. 443–453.

Leroy, J. L. and Frongillo, E. A. 2007. “Can interventions to promote animal production ameliorate undernutrition?” Journal of Nutrition, 137, pp. 2311–2316.

Leroy, J.L., Ruel, M., Verhofstadt, E., and Olney D. 2008. “The micronutrient impact of programs addressing both direct and underlying determinants of malnutrition.” Report for the Innocenti Micronutrient Program Meeting: Strengthening the Evidence-base for Programs that Reduce Micronutrient Deficiencies and Improve Health and Development.

Ligon, E. and Sadoulet, E. 2007. “ Estimating the effects of aggregate agricultural growth on the distribution of expenditures.” Background paper for the World Bank World Development Report 2008. Loayza, N. V., and Raddatz, C. 2010. “The composition of growth matters for poverty alleviation.” Journal of Development Economics, 93.1, pp. 137–151.

Masset, E., Haddad, L., Cornelius, A., and Isaza-Castro, J. 2011. “A systematic review of agricultural interventions that aim to improve nutritional status of children.” London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Mullins, G., Wahome, L., Tsangari, P., and Maarse L. 1996. “Impacts of intensive dairy production on smallholder farm women in coastal Kenya.” Human Ecology, 24, pp. 231–53.

Peña, C., Webb, P., and Haddad, L. 1996. “Women’s economic advancement through agricultural change: a review of donor experience.” IFPRI Food Consumption and Nutrition Division Discussion Paper No. 10.

Pingali P. 1997. “From subsistence to commercial production System: the transformation of Asian agriculture.” American Journal of Agricultural Economics. 79, 2, pp. 628–634.

Pingali, L. P., and Rosegrant, M. W. 1995. “Agricultural commercialization and diversification: Process and polices.” Food Policy, 20.3, pp. 171–185.

Popkin, B. M. 1980. “Time allocation of the mother and child nutrition.” Ecology of Food and Nutrition, 9, pp. 1–14.

Quisumbing, A., Brown, L., Feldstein, H., Haddad, L. and Peña, C. 1995. “Women: the key to food security.” IFPRI Food Policy Report.

Quisumbing, A., and Maluccio, J. 2000. “Intrahousehold allocation and gender relations: new empirical evidence from four developing countries.” IFPRI Food Consumption and Nutrition Division Discussion Paper No. 84. Romer, P. 1993. “Idea gaps and object gaps in economic development.” Journal of Monetary Economics, 32.3, pp. 543–573.

Romer, P. 1994. “New goods, old theory and the welfare cost of trade restrictions.” Journal of Development Economics, 43.1, pp. 5–38.

Ruel, M. 2001. Can food-based strategies help reduce vitamin A and iron deficiencies? a review of recent evidence. Washington, DC: IFPRI.

Skoufias, E., Tiwari, S., and Zaman, H. 2012. “Crises, food prices, and the income elasticity of micronutrients: estimates from Indonesia.” World Bank Economic Review, 26.3, pp. 415–442.

Soleri, D., Cleveland, D. A., and Frankenberger, T. R. 1991. “Gardens and vitamin A: a review of recent literature.” Retrieved from
http://www.es.ucsb.edu/faculty/cleveland/CV/1991_Gardens&VitA.pdf

Strasberg, P. J., Jayne, T. S., Yamano, T., Nyoro, J., Karanja, D., and Strauss, J. 1999. “Effects of agricultural commercialization on food crop input use and productivity in Kenya.” Michigan State University International Development Working Paper No. 71.

Strauss, J., and Thomas, D. 1995. “Human resources: empirical modeling of household and family decisions.” In J. Behrman, and T. N. Srinivasan (Eds.), Handbook of Development Economics (pp. 1884-2023), Amsterdam: North-Holland.

Subramanian, S., and Deaton, A. 1996. “The demand for food and calories.” Journal of Political Economy, 104.1, pp. 133–162.

Svensson, J., and Yanagizawa, D. 2009. “Getting prices right: the impact of the market information service in Uganda.” Journal of the European Economic Association, 7.23, pp. 435-445.

Timmer, P. 1997. “Farmers and markets: political economy of new paradigms.” American Journal of Agricultural Economics, 79.2, pp. p. 621-627.

Umapathi, N. 2008. “Maternal education, child-care and nutritional status: lessons from a nutritional program.” Job Market Paper, Institute for Fiscal Studies and Centre for Microdata Methods and Practice, University College London.

UNICEF. 1990. Strategy for improved nutrition of children and women in developing countries. New York, NY: UNICEF.

Von Braun, J., de Hean, H. and Blanken, J. 1991. “Commercialization of agriculture under population pressure: effects on production, consumption, and nutrition in Rwanda.” IFPRI Research Report No. 85.

Von Braun, J. and Kennedy, E. 1986. “Commercialization of subsistence agriculture: income and nutritional effects in developing countries.” IFPRI Working Paper on Commercialization, Agriculture and Nutrition No. 1.

Von Braun, J. and Kennedy, E. 1994. Agricultural commercialization, economic development, and nutrition. Baltimore, Maryland: Johns Hopkins University Press.

Von Braun, J. 1995. “Agricultural commercialization: Impacts on income and nutrition and implications for policy.” Food Policy, 20.3, pp. 187–202.

Webb, P. and Block, S. A. 2004. “Nutrition information and formal schooling as inputs to child nutrition.” Economic Development and Cultural Change, 52.4, pp. 801–820.

Webb, P. and Block, S. A. 2010. “Support for agriculture during economic transformation: Impacts on poverty and undernutrition.” Proceedings of the National Academy of Sciences, pp. 1–6.

World Bank. 2007. “From agriculture to nutrition: pathways, synergies, and outcomes.” World Bank Report No. 40196-GLB.

Chapter 2

Measuring and Increasing Adoption Rates of Cookstoves in a Humanitarian Crisis

with Daniel Wilson, Jeremy Coyle, Javier Rosa, Omnia Abbas, Mohammed Idris Adam, and Ashok Gadgil[25]

Abstract

Traditional smoky cooking fires are today's greatest environmental threat to human life. These fires, used by half the global population, cause 3.9 million annual premature deaths. "Clean cookstoves" have potential to improve this situation; however, most dissemination programs do not employ objective measurement of adoption to inform design, marketing, and dissemination practices. Lack of data prevents insights and may contribute to consistently low adoption rates. In this study, we used sensors and surveys to measure objective versus self-reported adoption of freely-distributed cookstoves in Darfur, Sudan. Our data insights demonstrate how to effectively measure and promote adoption, especially in a humanitarian crisis. With sensors, we measured a 71% initial adoption rate compared to a 95% rate reported during surveys. No line of survey questioning, whether direct or indirect, predicted sensor-measured usage. For participants who rarely or never used their cookstoves after initial dissemination ("non-users"), we found significant increases in adoption after a simple followup survey ($p = 0.001$). The followup converted 83% of prior "non-users" to "users" with average daily adoption of 1.7 cooking hours over 2.2 meals. This increased adoption, which we posit resulted from cookstove familiarization and social conformity, was sustained for a 2-week observation period post intervention.

Introduction

Since the beginning of the modern Darfur conflict in 2003, violence has forced Darfuri families from their homes. Many displaced families have emigrated from their homelands to large Internally Displaced People's (IDP) camps; current UN figures estimate 2.5 million IDPs in Darfur (UNOCHA 2015). In 2005, The University of California, Berkeley and Lawrence Berkeley National Laboratory began a joint effort to design an efficient cookstove for use in Darfuri IDP camps. The impetus for the Berkeley-Darfur Stove (BDS) was to reduce the burden and danger IDP women face when acquiring fuel in and around the camps. The BDS's improved thermal efficiency allows users to cook food using less fuel of a traditional three-stone fire (TSF) style cookstove (locally known as a "ladaya") (Preble et al. 2014). As

[25]Wilson, Coyle, and Rosa: University of California, Berkeley. Abbas: Potential Energy. Adam: Al-Fashir University. Gadgil: University of California, Berkeley and Lawrence Berkeley National Laboratory.

of June 2015, 42,000 BDSs have been distributed to households in Darfur. With an average IDP household size of 7, the BDS has reached roughly 294,000 individuals.[26]

Objective monitoring and evaluation is a major barrier to quantifying the impacts of "clean cookstoves" like the BDS. For decades, clean cookstoves have promised to reduce the global burden of disease and drudgery attributable to traditional cooking. Air pollution from traditional cooking methods is the world's largest environmental health risk – traditional biomass-fueled cooking is linked with 3.9 million annual premature deaths (Lim et al. 2013, Smith et al. 2014). Clean cookstoves' promise is to displace traditional smoky biomass fires (used by almost half the world's population) with cleaner combustors (World Bank 2011). However, positive outcomes of clean cookstove interventions are rarely significant or sustained (Smith et al. 2015) because clean cookstoves have not been widely adopted or sufficiently displaced traditional cookstoves (i.e. via "stove stacking") (Pillarisetti et al. 2014).

Stove Use Monitor (SUM) sensors have the potential to objectively inform implementation agencies, policy makers, and analysts about field performance and adoption of cookstoves (Thomas et al. 2013, Pillarisetti et al. 2014, Ram et al. 2010, Ruiz et al. 2011, Ruiz et al. 2013, Wilson et al. 2015. This information can improve cooks' health and economic outcomes. Many stakeholder agencies are eager to understand which cookstoves, training programs, and marketing methods are effective.

However, most cookstove adoption studies use unreliable survey data subject to three problematic sources of error (Lewis and Pattanayak 2012, Burwen 2011, Pillarisetti et al. 2014). First, respondents may over-report "ideal" behaviors, thus generating "social desirability" or "courtesy" bias (Nunnally 1978, Edwards 1957). Second, respondents may struggle to recall and accurately aggregate data on their stove use over long recall periods. Finally, frequent visits by enumeration staff are known to induce observation bias via the "Hawthorne Effect," (Parsons 1974). Even when recall data is unbiased on average, measurement error in dependent variables (say, expenditures or frequency of symptoms) reduces the precision of estimates, giving reduced statistical significance. Unbiased measurement error in independent variables (for example, usage rates) leads to "attenuation bias" that pushes impact estimates toward zero (Das et al. 2012). This effect only worsens when users systematically overstate adoption. Unlike surveys, sensors are unbiased, discreet, and long-lasting. Objective sensor data leads to improvements in our understanding of cookstove adoption and enables insights about cookstove designs, training, or marketing techniques that may increase utilization.

In this study, we add critical information to the small pool of studies objectively measuring cookstove adoption with sensors (BAMG 2012, Burwen et al. 2012, Pillarisetti et al.

[26]Qualitative reports suggest that the BDS is well-liked. Owing to its speedy cooking and the difficulty Arabic speakers have pronouncing "The Berkeley-Darfur Stove," the BDS has earned its own Arabic nickname that translates to "Five-Minute Stove". Most BDS units have been distributed free of charge into large IDP camps such as Al-Salam, Abu Shouk, and Zam Zam in North Darfur. Approximately 5,000 cookstoves have been sold in sales trials within villages near these camps. Large-scale dissemination of the BDS is enabled by three key parties: Potential Energy is a USA-based non profit that manages logistics and fund raising, Shri Hari Industries manufactures BDS kits in Mumbai, India and ships them to Sudan, and Sustainable Action Group, a Sudanese Non-Governmental Organization, assembles kits into complete BDSs and distributes them in Darfur.

2014, Ruiz et al. 2011, Ruiz et al. 2013, Wilson et al. 2015). To our knowledge, other groups have not analyzed adoption of cookstove technologies in an ongoing humanitarian crisis. Evaluation of technologies and techniques that improve living conditions and environmental conditions in humanitarian crises is important but inadequately addressed in the scientific literature. Crises are important case studies because they are regrettably common, are a frequent target market for cookstove dissemination programs, and represent unique social and economic contexts that make non-crises insights potentially non-transferable. In a prior study, sensor data has been correlated with survey data (Thomas et al., 2013), but we extend the literature by testing whether multiple surveying techniques can be combined to better predict sensor-measured behavior. Additionally, we present a novel framework for categorizing stove recipients as "users" or "non-users" and demonstrate the value of this delineation in revealing data insights. Unlike prior work, this study rigorously evaluates causes of sensor damage and loss, and we perform sensitivity analysis for lost data. We also present the first study to our knowledge quantifying the impacts of social pressure on cookstove adoption ("courtesy use"), and we show an example of how sensors revealed user-generated innovations. Lastly, and most importantly, this study adds meaningful data to the literature by assessing the impact of enumeration activities—or the anticipation thereof—on cookstove adoption behavior, and we reveal the potential for low-cost strategies to dramatically increase cookstove adoption.

Design and Methods

A detailed description of the design and methods of this experiment is discussed in prior work (Wilson et al. 2015) and in the supplement. For clarity, a brief overview follows.

This work took place in the Al-Salam IDP Camp outside Al-Fashir, North Darfur, Sudan. Sustainable Action Group (SAG), a Sudanese non-profit that assembles and distributes the BDS in IDP camps, selected participants for this study in the usual procedure for BDS dissemination: 180 participants were selected to receive free BDSs by chief camp administrators ("omdas") who selected study participants from a comprehensive list of inhabitants. Selected participants were limited to five of the camp's "Administrative Units" representing the geographical and cultural emigration origin of IDPs. In compliance with the University of California, Berkeley's Institutional Review Board approval (CPHS #2013-03-5132), participants were told they would be taking place in a study of the BDS "to improve future BDSs" and that a temperature sensor would be attached to the BDS. However, information about the tracking of cooking behaviors was withheld. All five Units participated in a baseline and followup survey, and three units participated in an additional second followup.

The five Units, each with 36 participants, received their BDSs and took a baseline survey between July 29th and August 2nd of 2013. Serious security concerns precluded enumeration staff from travel and extended stay in the camps, so the baseline survey and all subsequent interactions took place at midday in a women's center in Al-Salam Camp. After four to twelve weeks, depending on the Unit, a followup survey was conducted. All women were supplied with the date, time, and location of their followup. Women were instructed to bring their cookstoves to the Women's Center for the followup survey. On the appointed day, a survey was conducted and SUMs were removed from cookstoves. Data were discreetly downloaded from SUMs using a laptop computer. For three of the five Units, SUMs were re-attached and a second followup was conducted two weeks after the first followup. The second followup also required bringing the cookstove to the Women's Center so SUMs could be removed, but no additional surveys were conducted. In all cases, women brought their BDSs home from the followup survey(s) and owned the stoves indefinitely thereafter. Additional details about scheduling can be found in the Table S1.

Table S1: The scheduling of baseline survey, 1^{st} followup, and 2^{nd} followup and the spectra of SUMs sampling periods used in this study.

Scheduling by Unit	Korma	Al-Fashir	Zaghawa	Jebel Si	Tawila
start date (MM/DD/2013)	07/29	07/30	07/31	08/01	08/02
participants	36	36	36	36	36
duration to 1^{st} followup (weeks)	4	6	8	10	12
1^{st} to 2^{nd} followup (weeks)	-	-	2	2	2
period of primary SUMs (min)	4.9	7.4	9.8	12.3	14.7
cookstoves by SUMs period					
primary^{a} + dummy	26	26	26	26	26
primary^{a} + 1-minute^{b}	4	4	4	4	4
4-minute^{c} + dummy	4	4	4	4	4

^{a}measured entire duration between baseline and 1^{st} followup
^{b}measured the last 8,200 minutes (5.7 days) before followup
^{c}measured the last 33,000 minutes (23 days) before followup

Building on the methods of others (BAMG 2012, Burwen 2012, Pillarisetti et al. 2014, Ruiz et al. 2013, Mukhopadhyay et al. 2012), we utilized Maxim's DS1922E-series iButtons as temperature data loggers, as seen in Figure S1. Figure S2 shows the mounting location of SUMs which was chosen by laboratory cooking experiments to maximize signal (temperature) while still preventing overheating of the sensor.

Surveys in this study were conducted by a team of two redundant enumerators. One enumerator administered the survey using paper and pen while the other used Open Data Kit (ODK) (a smartphone-based survey tool). Data from these two methods were tested against one another for quality control purposes. Data from SUMs and surveys were processed in Version 3 of the open-source statistical computing software R. A further discussion of the algorithm used to label cooking events can be found in Wilson et al. (2015).

Figure S1: Left: women gathered at Al-Salam's women's center to partake in baseline surveys and receive their BDS. Right: a woman holds her BDS and points to SUMs mounted near its base.

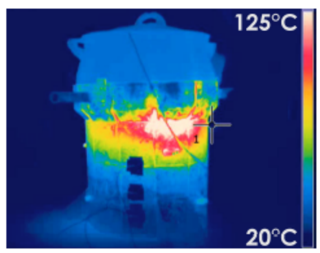


Figure S2: An infrared image of a wood-fired BDS with pot and lid being lab tested for sensor placement. Dark squares are low emissivity aluminum tape holding thermocouples on potential cookstove mounting locations. Tape appears cool because it is reflecting the low-temperature infrared of the cool surroundings.

Results

SUMs Loss and Bias

The SUMs in this experiment were vulnerable to failure, which may produce SUMs data bias. During the study we discovered that some women innovated by flipping the BDS upside down, filling the bottom with charcoal, and preparing drinks or small meals. The BDS was not designed for this mode of use. The SUM was mounted at the bottom of the BDS to avoid overheating from wood fires (see Figure S2 above), but the bottom of the BDS is precisely where charcoal fires would make the BDS hottest. Of 170 participants with SUMs-equipped cookstoves, 29 participants had thermally-damaged SUMs. In followup surveys, participants who reported using charcoal as a primary cooking fuel (for food or drink) were 2.8 times more likely ($p = 0.01$, Fisher test) to thermally damage their SUMs as participants who did not report using charcoal as a primary fuel. A summary of SUMs failures by Administrative Unit is presented in Table S2.

Table S1: The scheduling of baseline survey, 1st followup, and 2nd followup and the spectra of SUMs sampling periods used in this study.

Scheduling by Unit	Korma	Al-Fashir	Zaghawa	Jebel Si	Tawila
start date (MM/DD/2013)	07/29	07/30	07/31	08/01	08/02
participants	36	36	36	36	36
duration to 1st followup (weeks)	4	6	8	10	12
1st to 2nd followup (weeks)	-	-	2	2	2
period of primary SUMs (min)	4.9	7.4	9.8	12.3	14.7
cookstoves by SUMs period					
primary[a] + dummy	26	26	26	26	26
primary[a] + 1-minute[b]	4	4	4	4	4
4-minute[c] + dummy	4	4	4	4	4

[a]measured entire duration between baseline and 1st followup
[b]measured the last 8,200 minutes (5.7 days) before followup
[c]measured the last 33,000 minutes (23 days) before followup

Table S2: A summary of SUMs fates

SUMs fate by Unit	Korma	Al-Fashir	Zaghawa	Jebel Si	Tawila	Total
SUMs loss totals	8	7	14	12	18	59
SUMs loss before 1st followup	8	7	12	9	11	47
thermal damage	3	3	10	8	5	29
dead at 1st programming	3	0	0	0	1	4
lost before first programming	1	1	2	0	1	5
participant did not return	1	2	1	0	5	9
stolen stove	0	1	0	0	0	1
other	0	0	0	1	0	1
SUMs loss after 1st followup	-	-	1	3	6	10
thermal damage	-	-	1	3	0	4
participant did not return	-	-	0	0	6	6

Notes: There was only one followup for Korma and Al Fashir. Three stove users in Tawila and one in Zaghawa who did not return for the first followup returned for the second followup, making part of their missing pre-1st-followup data available.

Because data were unrecoverable from thermally-damaged SUMs, this study has possible bias. We posit that adopters of the BDS are more likely to damage SUMs and therefore adopters are underrepresented in surviving SUMs data. Put another way, it would be difficult to thermally damage a SUM mounted on a BDS that was never used for cooking. Therefore, data loss from thermal damage represents a non-random sampling bias in SUMs data and a probable downward bias on sensor-measured adoption rates. Accordingly, SUMs-derived data presented in this study are a conservative estimate of adoption throughout the entire experimental population. Other less prevalent causes of SUMs loss were observed, namely misplaced and faulty sensors (before distribution and baseline survey), one stolen stove, and a small number of women not returning for followup surveys. However, we assume these loss factors do not meaningfully bias data. Unless otherwise noted, quantitative SUMs data presented throughout this study are derived from surviving SUMs.

Defining "User" and "Non-User" Groups

To perform more meaningful analyses, we classified participants into two groups based on their pre-followup BDS adoption: "users" and "non-users." First, using SUMs data for each participant, we computed the proportion of cookstove ownership days where at least one cooking event was observed. The "pre-followup period" analyzed was defined from one day after the participant's baseline survey until two days before the participant's follow up survey (to avoid effects near the followup discussed later). This variable, termed "proportion of days use" is plotted as a cumulative distribution function in Figure 1. An arbitrary delineation was drawn at 10% of days used, and participants utilizing the BDS more than 10% BDS ownership days in this period were classified as "users." Although the quantitative delineation between users and non-users was arbitrary, it is useful to characterize study participants in terms of women who generally adopted the BDS before their follow up survey (users) and those who rarely or never used the BDS before the follow up survey (non-users). Unless otherwise specified, "user" and "non-user" classification refers to pre-followup behavior.

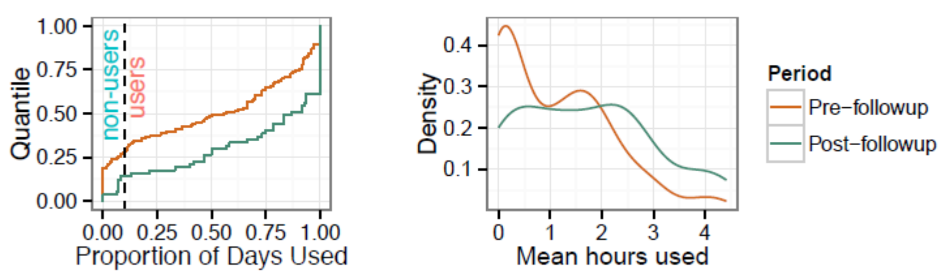


Figure 1: Left: a cumulative distribution function of the proportion of days used as measured by SUMs. Dashed vertical line indicates the 10% of use days delination between "non-users" (left of dashed line) and "users". Right: the probability densities of hours of daily cooking pre and post followup.

Pre-Followup Adoption Measured by SUMs and Surveys

In the pre-followup period, 87 of 122 (71%) participants with surviving SUMs were classified as users of the BDS. Remembering that SUMs thermal damage presents a downward bias on this study, if all thermally-damaged SUMs were presumed to belong to "users," the study-wide adoption rate would be 77% users.

The proportion of users varied widely and significantly by Administrative Unit (p=0.004; Fisher's exact test) with Al-Fashir Rural having the highest rate of users at 89% and Jebel Si having the lowest at 62%. Study-wide, a typical user utilized her BDS 1.54 (SD = 0.97) hours per day over 2.04 (SD = 1.30) cooking events. Including non-users, the study-wide average adoption rates were 1.10 (SD = 1.07) hours and 1.47 (SD = 1.43) events of daily cooking. A summary of SUMs-measured adoption is shown in Table 1. Although ownership periods were relatively short compared with other longitudinal studies (Pillarisetti et al. 2014, Ruiz et al. 2011), users showed no significant linear trend in average hours cooked per day over the pre-followup ownership period (estimated increase of 0.0002 hours/day; p=0.84).

Table 1: A summary of SUMs results. Note that "user" and "non-user" are defined in the context of the pre-followup period. Data format is: mean (standard deviation).

Results by Unit	Korma	Al-Fashir	Zaghawa	Jebel Si	Tawila	Total
pre 1st followup						
users	17	24	16	16	14	87
non-users	9	3	5	10	8	35
hours cooked per day	1.00 (0.99)	1.64 (1.01)	0.98 (0.91)	0.72 (0.84)	1.14 (1.38)	1.10 (1.07)
users	1.52 (0.85)	1.84 (0.87)	1.27 (0.84)	1.15 (0.82)	1.79 (1.37)	1.54 (0.97)
non-users	0.03 (0.07)	0.00 (0.00)	0.03 (0.06)	0.04 (0.09)	0.02 (0.02)	0.03 (0.06)
cooking events per day	1.58 (1.56)	2.10 (1.41)	1.36 (1.32)	0.90 (1.06)	1.33 (1.55)	1.47 (1.43)
users	2.39 (1.33)	2.36 (1.26)	1.77 (1.25)	1.44 (1.04)	2.08 (1.49)	2.04 (1.30)
non-users	0.05 (0.09)	0.00 (0.00)	0.04 (0.07)	0.04 (0.10)	0.03 (0.03)	0.04 (0.07)
post 1st followup						
users	-	-	16	14	11	41
non-users	-	-	5	8	6	19
hours cooked per day	-	-	2.01 (1.10)	1.50 (1.08)	1.67 (1.62)	1.73 (1.26)
users	-	-	1.91 (1.15)	1.51 (1.28)	1.77 (1.44)	1.74 (1.26)
non-users	-	-	2.32 (0.94)	1.48 (0.69)	1.50 (2.05)	1.70 (1.30)
cooking events per day	-	-	2.65 (1.57)	1.97 (1.32)	1.62 (1.83)	2.11 (1.60)
users	-	-	2.51 (1.68)	1.84 (1.45)	1.70 (1.74)	2.06 (1.62)
non-users	-	-	3.11 (1.15)	2.21 (1.12)	1.47 (2.15)	2.21 (1.57)

Participants reported high rates of BDS adoption regardless of their SUMs-measured usage: 95% of participants reported using the BDS as their "primary stove." As shown in Figure 3, nearly all users reported using the BDS three times a day on a "normal day." This is compared to a SUMs-measured daily cooking events of 2.04 events for users and 0.04 events for non-users. 77% of users and 86% of non-users over-report cooking events. Of the 62 participants who use the BDS an average of less than once per day, only five actually report doing so. Both users and non-users over-report cooking hours with 85% of users and 86% of non-users over-reporting. All told, over-reporting represents 1.2 hours and 1.3 events of daily cooking overestimation.

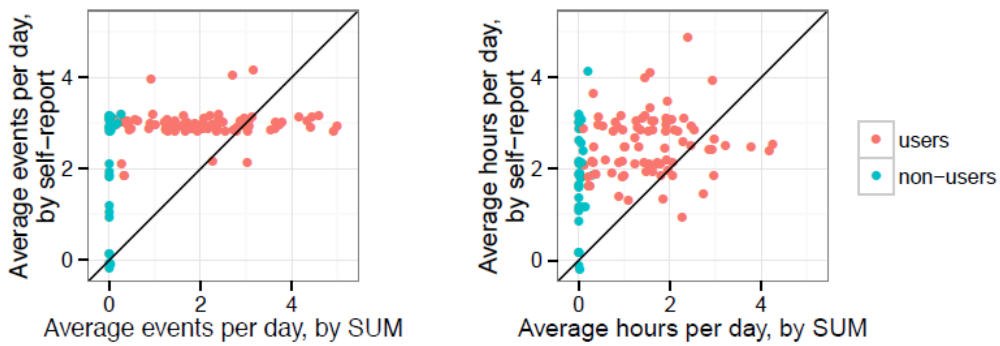
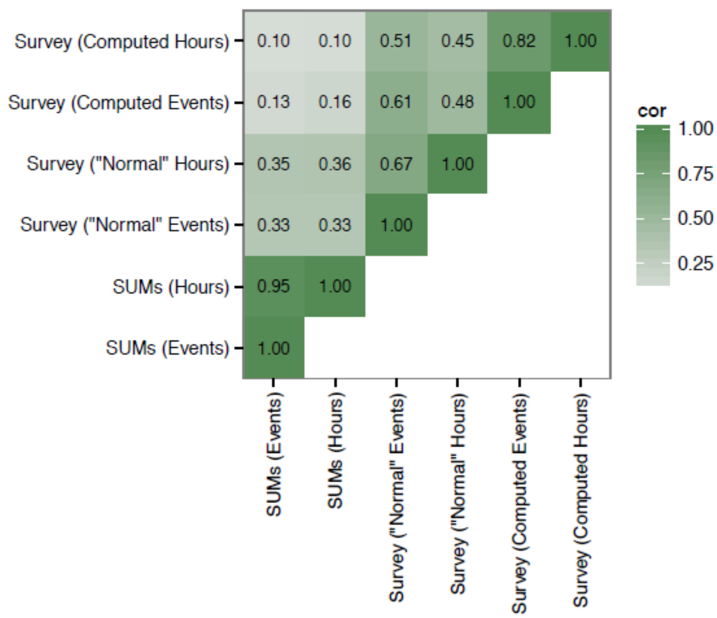


Figure 3: Daily events (left) and hours of BDS use (right) are shown as scatter plots of SUMs-measured versus self-reported data. The scatter plots include the 1:1 line that data would fall on if users' self reports perfectly agreed with the SUMs algorithm. To avoid over-plotting, plot points are "jittered."

Disagreement between SUMs and surveys was so extreme that we assumed participants had difficulty with the wording of the question "On a 'normal day,' how many times/hours do you use your BDS?" We posited that women may over-report because they consider behavior only on use days rather than average days; to answer the survey questions in a way that would perfectly correlate with SUMs, women would have to time-average usage and then report their average daily use. We believed this was challenging, so we checked whether SUMs data correlated with another way we asked about usage. For an exhaustive list of meal types, we asked how many times in a normal week she prepares that meal with the BDS and how long each meal takes to cook on the BDS. The sum of weekly occurrences is termed "Computed Events" and the sum of the product of meal durations and weekly occurrences is termed "Computed Hours." These data are shown in a correlation table with SUMs data in Table 2. Although these variables do not rely as heavily on cooks' mental math as the "normal day" questions, they correlate even more weakly with SUMs data.

Table 2: Correlation coefficients tabulated for SUMs and survey-based measures of adoption.



Impacts of Enumeration Activities on Adoption

Figure 4 illustrates the effect of enumeration activities on adoption for the 3 Units tracked after the first followup. Beginning roughly two days before the scheduled followup, non-users begin to adopt their BDSs. After followup, enumeration activities had a statistically significant positive impact on the non-user group, increasing hours of daily cooking by 1.6 hours ($p<0.001$, paired t-test) and no meaningful impact on the user group, increasing hours of daily cooking by only 10 minutes. Compared with the pre-followup period, non-users also increased adoption of the BDS in terms of events, with an average increase of 2.1 events per day. As a reminder, "users" and "non-users" are classified solely by their BDS adoption *before* the followup survey. A summary of results both pre- and post-followup is shown in Table 1.

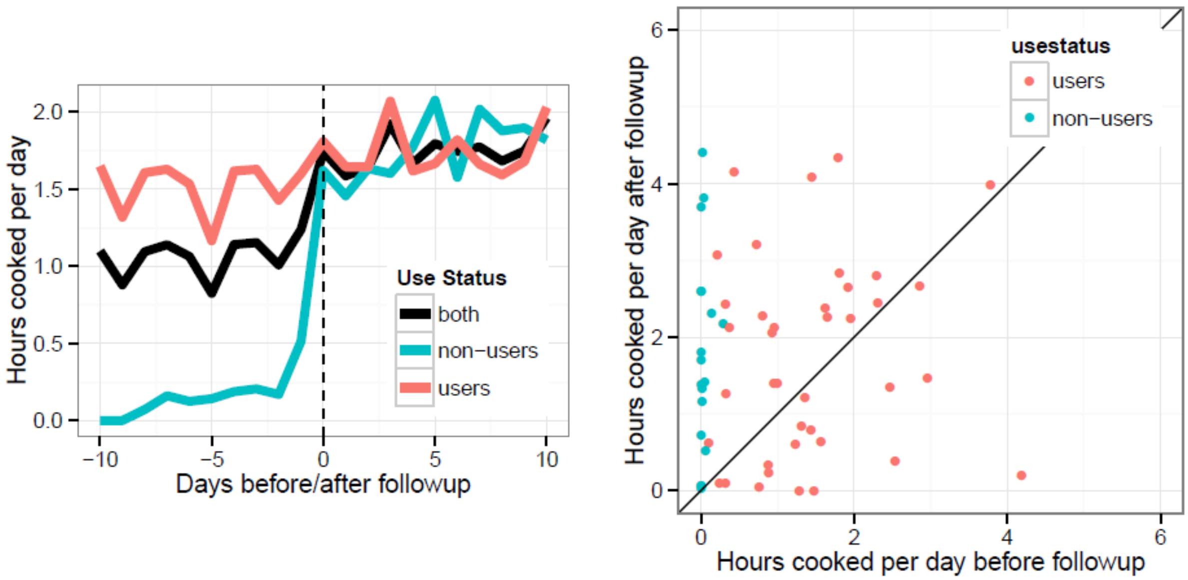


Figure 4: Left: average hours of daily cooking per day in the ten days preceding and tailing the followup. Right: post followup versus pre followup SUMs data of hours of cooking per day. Points falling along the 1:1 line represent participants whose behavior was unchanged by the followup. Participants above the 1:1 line use their stove more after the followup, and participants below the 1:1 line use their stove less after the followup.

Discussion

Although the BDSs were distributed free of charge, it was relatively well-adopted among recipients, with 71% of surviving SUMs classifying cooks as "users." The (downward) bias caused by thermally damaged SUMs resulting from inverted BDS charcoal fires means that up to 77% of participants were potential "users" in the pre-followup period. Among "non-users," pre-followup behavior is characterized by little or no BDS utilization whatsoever. There is little evidence that non-users try and then abandon the BDS. Rather, non-users seem to neglect using the BDS altogether until just days before the followup survey.

Other studies have found that socioeconomic and educational factors are the most important predictors of cookstove adoption (Lewis et al. 2012), and this trend likely holds in our study as well. Namely, Al-Fashir Rural Unit, which exhibited the highest rates of adoption, is comprised of inhabitants who have emigrated to Al-Salam IDP Camp from peri-urban settlements near North Darfur's capital. Although socioeconomic and educational factors were not measured explicitly in surveys, Al-Fashir Rural residents were likely to have been exposed to better educational and work opportunities than residents from other Units representing poorer rural parts of Darfur.

As found in other studies of health-related technologies (Arnold et al. 2009, Kremer et al. 2010, Thomas et al. 2013, Ram et al. 2010), study participants tended to over-report cookstove adoption. In this study, average self-reported used was roughly twice SUMs data in terms of hours and events of cooking per day. However, because there is little correlation between SUMs and survey data, it is somewhat misleading to think of reporting as a two-fold overestimation. For example, almost all women surveyed (83%) report using the BDS three times per day – every meal. In other words, it is incorrect to think that women inflate their adoption by two-fold; instead, nearly uniformly, women report using the BDS for all daily meals, but, population-wide, women actually use their BDSs for about half of all meals. Though we attempted to adjust for users' inability to average over large time periods by calculating adoption from other questions, no manner of questioning or reinterpretation of reporting periods correlated well with SUMs-measured behavior, leading us to believe that many participants intentionally misrepresent cookstove adoption when surveyed.

We were surprised by the strong influence of survey enumeration activities on the "non-user" group. As easily seen in Figure 4, non-users exhibit a strong up-tick in adoption starting two days before their scheduled followup survey. This effect is seen in all Units that were observed until a second followup, and this effect spanned a range of socio-economic levels and cultural or geographic origin (as indicated by Administrative Unit) and pre-followup ownership duration. We posit that non-users may feel social pressure to use their BDS before the followup survey; perhaps it would be embarrassing to bring a shiny, clean, unused, donated stove back to the women's center. Or, perhaps, non-users felt the need to educate themselves about the BDS before returning for the followup survey. Regardless of motivation, in the days leading up to the followup survey, the non-user group strongly exhibits what we refer to as "courtesy use." This spike in usage in reaction to (or anticipation of) direct observation is consistent with other studies in the developing and developed world (Ram et al. 2010, Pedersen et al. 1986, Russell et al., 1992).

However, what was not expected was the non-user group's sustained adoption in the two weeks after followup survey enumeration. Because enumeration activities were not instructive or coercive, one would expect women who did not adopt the BDS would continue to neglect it after the follow up survey. Quite to the contrary, upon returning home, non-users' cookstove utilization became indistinguishable (mean 1.77 hours, standard error 0.31) from their "user" peers (mean 1.74 hours, standard error 0.22). In fact, using the same definition of "user" as the pre-followup period (BDS use on >10% of ownership days), 83% of previous non-users transitioned to "post-followup users." Additionally, population wide, 86% of participants would be classified as users in the post-followup period compared with 71% in the pre-followup period.

We propose three conjectures that may explain the phenomenon of non-user conversion. First, non-users rarely or never used their BDSs until just before the followup, so this group may never have realized the benefits of the BDS. Perhaps, after finally trialing the BDS as a courtesy immediately before the followup, non-users realized they enjoyed the BDS and subsequently continued use after the followup. Second, it is possible that peer pressure at the followup survey influenced non-users; one can imagine non-users walking and talking with their user peers to the Women's Center. At the Women's Center, although surveys were private, women may have seen peers with well-used cookstoves. Non-users may have overheard others others talking, truthfully or not, about enjoying the BDS and thus felt more comfortable trying the BDS. Finally, SUMs vs. surveys suggest that many non-users told mistruths during the followup survey about how often they use the BDS.

This deceit, although untrue, could have built non-users' self-efficacy as a BDS adopter and helped non-users visualize themselves as BDS adopters, inducing adoption after returning home (Bandura 1977). Alternatively, deceit may generate unpleasant cognitive dissonance, which participants may resolve through adoption consistent with their self-reports (Aronson 1980, Dickerson 1992). These theories were not tested in this study, but suggest potential contributors to post-followup adoption.

In this study, we presented cookstove adoption and reporting behaviors for recipients of the BDS, joining a small set of studies that have been able to combine traditional self-reported data with objective sensor-based measurements. Our analysis is unusual in part because of its context: despite the distribution of tens of thousands of BDS cookstoves in IDP camps to date, the challenging operating environment that makes aid needed also lends itself to data scarcity and other monitoring challenges.

The relevance of our results is not limited to the IDP context. Free distribution of improved cookstoves is commonplace, and it is likely to be accompanied by a desire among recipients to report behaviors preferred by distributors. Indeed, this study contributes evidence of the discrepancies between self-reported and sensor-detected usage, and affirms the need for sensor-based inquiry when impact must be accurately measured.

Additionally, we were able to use sensors to show an example of how monitoring activities can themselves alter the behaviors being monitored: usage spiked just before a followup visit, and for many previously-non-users, the uptick was sustained for the following two weeks. This discovery may have useful implications for optimal followup after distribution and provide insight about strategies for inexpensive "light-touch" interventions to increase cookstove adoption. We also show nuances in how and when participants may interact with a new technology – exposure or usage may not be uniform during the reference period, which may in turn complicate self-reports of behavior.

Our findings do not suggest that sensors can or should replace self-reported data more generally. Sensors can be costly to implement and can only cover a small fraction of the types of data that may be relevant for analysis. Still, it is important to consider which types of data are most likely to be reliable and whether objective data sources may complement survey data in a given context. In this IDP context, it is apparent that surveys are extremely unreliable means of measuring technology adoption.

We believe this study has illuminated important insights about the value of objective data, the unreliable nature of surveys, and how sensors can reveal valuable insights to stakeholders interested in solving the crisis of cooking.

Acknowledgments

The authors would like to thank our enumeration and implementation team. Namely, we thank Nada Abdalla Mohammed, Afaf Adam Abdallha, Eissra Hamid Gamer, Abel Rahman Abdalla Addoma, Aziza Mohammed Tugod, Fatima Adam Ibrahim, Om Alhosein Ali Garbo, Idris Ibrahim Adam, Abdalla Mohammed Suleiman, and Adam Abdalla Amin for their extraordinary efforts towards the success of this work.

Authors gratefully acknowledge financial support from multiple sources for this work. The primary funding for this work comes from a DIV Phase-1 grant by the United States Agency for International Development (USAID) to Potential Energy. Daniel Wilson and Angeli Kirk are grateful for support from the National Science Foundation (NSF) Graduate Research Fellowship. Additional funding for personnel and materials for this project has been generously provided by funding agencies including the Development Impact Lab (USAID Cooperative Agreement AID-OAA-A-13-00002) which is part of the USAID Higher Education Solutions Network, the Blum Center for Developing Economies, a Behavioral Sensing Grant from The Center for Effective Global Action (CEGA), and Department of Energy Contract DE-AC02-05CH11231 to Lawrence Berkeley National Laboratory (LBNL), operated by the University of California.

A preliminary analysis of this data set originally appeared in a conference paper for the 2014 Tech4Dev International Conference put on by the UNESCO Chair in Technologies for Development with the theme "What is Essential?" The conference paper, among others, was published as a book chapter by Springer (Wilson et al. 2015). This paper in its current form represents significant progress in analysis and insight into the original data set. For readability and flow, this paper summarizes the Background and Methods sections as well as some analysis presented in the conference paper and book chapter, but no copywritten text or figures are duplicated herein. Additionally, the photos for the TOC art in the paper were taken by field staff in Darfur during the completion of this study.

References

Benjamin Arnold, D. M. A. H., Byron Arana; John M Colford, J. Evaluation of a pre-existing, 3-year household water treatment and handwashing intervention in rural Guatemala. International Journal of Epidemiology, 2009, 38, 1651–1661.

Aronson, E, 1980. In Retrospection on socialpsychology; Festinger, L., Ed.; Oxford University Press: Oxford, pp 3–21.

Bandura, A., 1977. Self-efficacy: Toward a Unifying Theory of Behavioral Change. Psychological Review. 84, 191–215.

Bank, T. W., 2011. Household Cookstoves, Environment, Health, and Climate Change.

Berkeley Air Monitoring Group, 2012. Monitoring and Evaluation of the Jiko Poa Cookstove in Kenya Report.

Burwen, J. From Technology to Impact: Understanding and Measuring Behavior Change with Improved Biomass Cookstoves. 2011.

Burwen, J.; Levine, D. I. A rapid assessment randomized-controlled trial of improved cookstoves in rural Ghana. Energy for Sustainable Development 2012, 16, 328–338.

Dickerson, Chris Ann, Ruth Thibodeau, 1992. "Using Cognitive Dissonance to Encourage Water Conservation," Journal of Applied Social Psychology, 22, 841–854.

Das, J.; Hammer, J.; Sanchez-Paramo, C. The impact of recall periods on reported morbidity and health seeking behavior. Journal of Development Economics 2012, 98, 76–88.

Edwards, A. L. The social desirability variable in personality assessment and research.; Dryden Press, 1957.

Lewis, J. J.; Pattanayak, S. K. "Who Adopts Improved Fuels and Cookstoves? A Systematic Review." *Environmental Health Perspectives* 2012, 120, 637–645.

Lim, S. S.; Vos, T.; Flaxman, A. D.; Danei, G.; Shibuya, K. "A Comparative Risk Assessment of Burden of Disease and Injury Attributable to 67 Risk Factors and Risk Factor Clusters in 21 Regions, 1990-2010: a Systematic Analysis for the Global Burden of Disease Study 2010." The Lancet 2013,

Kremer, M.; Leino, J.; Miguel, E.; Zwane, A. P. Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions. Quarterly Journal Economics 2011, 126, 145–205.

Mukhopadhyay, R.; Sambandam, S.; Pillarisetti, A.; Jack, D.; Mukhopadhyay, K.; Balakrishnan, K.; Vaswani, M.; Bates, M. N.; Kinney, P. L.; Arora, N.; Smith, K. R. Cooking practices, air quality, and the acceptability of advanced cookstoves in Haryana, India: an exploratory study to inform large-scale interventions. Global Health Action 2012, 5, 1347.

Nunnally, J. C. Psychometric Theory, 2nd ed.; McGraw-Hill: New York, 1978.Pillarisetti, A.; Vaswani, M.; Jack, D.; Balakrishnan, K.; Bates, M. N.; Arora, N. K.; Smith, K. R. Patterns of Stove Usage after Introduction of an Advanced Cookstove: The Long-Term Application of Household Sensors. Environmental Science & Technology 2014, 48, 14525–14533.

Parsons, H. M. What Happened at Hawthorne?: New evidence suggests the Hawthorne effect resulted from operant reinforcement contingencies. Science 1974, 183, 922–932.

Preble, C. V.; Hadley, O. L.; Gadgil, A. J.; Kirchstetter, T. W. Emissions and Climate Relevant Optical Properties of Pollutants Emitted from a Three-Stone Fire and the Berkeley-Darfur Stove Tested under Laboratory Conditions. Environmental Science & Technology 2014, 48, 6484–6491.

Ram, P. K.; Halder, A. K.; Granger, S. P.; Jones, T.; Hall, P.; Hitchcock, D.; Wright, R.; Nygren, B.; Islam, M. S.; Molyneaux, J. W.; Luby, S. P. Is Structured Observation a Valid Technique to Measure Handwashing Behavior? Use of Acceleration Sensors Embedded in Soap to Assess Reactivity to Structured Observation. American Journal of Tropical Medicine and Hygiene 2010, 83, 1070–1076.

Ruiz-Mercado, I.; Masera, O.; Zamora, H.; Smith, K. R. Adoption and sustained use of improved cookstoves. Energy Policy 2011, 39, 7557–7566.

Ruiz-Mercado, I.; Canuz, E.; Walker, J. L.; Smith, K. R. Quantitative metrics of stove adoption using Stove Use Monitors (SUMs). Biomass and Bioenergy 2013, 57, 136–148.

Darhl M. Pedersen, K. B., Sheila Keithly Effects of an observer on conformity to handwashing norm. Perceptual and Motor Skills 1986, 62, 169–70.

Alan Russell, D. M., Graeme Russell Observer Influences on Mothers and Fathers: Self-Reported Influence During a Home Observation. Merrill-Palmer Quarterly 1992, 38, 263–283.

Smith, K. R.; Bruce, N.; Balakrishnan, K.; Adair-Rohani, H.; Balmes, J.; Chafe, Z.; Dherani, M.; Hosgood, H. D.; Mehta, S.; Pope, D.; Rehfuess, E. Millions Dead: How Do We Know and What Does It Mean? Methods Used in the Comparative Risk Assessment of Household Air Pollution. Annual Review of Public Health 2014, 35, 185–206.

Smith, K. R. Changing Paradigms in Clean Cooking. EcoHealth 2015, April.

Thomas, E. A.; Barstow, C. K.; Rosa, G.; Majorin, F.; Clasen, T. Use of remotely reporting electronic sensors for assessing use of water filters and cookstoves in Rwanda. Environmental Science & Technology 2013, 47, 13602–13610.

UNOCHA, Sudan: Humanitarian Bulletin Issue 19, 4 10 May; 2015.

Wilson, D. L.; Adam, M. I.; Abbas, O.; Coyle, J. R.; Kirk, A. E.; Rosa, J.; Gadgil, A. J. In Technologies for Development: What is Essential?, 1st ed.; Hostettler, S., Hazbourn, E., Bolay, J.-C., Eds.; Springer: New York, New York, 2015; pp 211–221.

World Bank, The, 2011. "Household Cookstoves, Environment, Health, and Climate Change." Technical report.

Chapter 3

Labor and Time Use Response to El Salvador's Conditional Cash Transfer Program

Abstract

I use a panel dataset from El Salvador to examine household short-term responses in time use to the Comunidades Solidarias Rurales conditional cash transfer program during 2007/2008 using difference-in-differences and regression discontinuity design. The introduction of a cash transfer program in El Salvador with the standard education conditions for children ages 6-12 and healthcare conditions for infants ages 0-3 was done in stages based nonrandomly on a set of observable municipality traits that precluded household-level influence over eligibility. This program design allows for a selection-on-observables approach to estimation. Because baseline analysis shows significant differences in a few household and individual characteristics between earlier and later phases, I use fixed effects specifications to control for any time-invariant differences between treatment and control. Because of the availability of only one baseline period, however, I cannot provide evidence against differences in time-variant trends. For each specification (DD and RDD, along with interactions with initial asset level and gender), I present results using two different bandwidths from the treatment cutoff. To accommodate a small number of municipalities in the sample, I apply wild cluster bootstrapping and present the resulting p-values along those obtained from clustered standard errors as typically applied for larger samples, and show that standard methods would lead to over-rejection of the null hypothesis in multiple instances. I use clustering at the municipality level in both cases.

Overall, many of my results are small and somewhat variable across alternative specifications, potentially due to measurement error, a small number of clusters, or simply a small response in the short run to a program offering a relatively small sum of \$15-20 a month. Still, a few consistent patterns emerge. My findings suggest that for children 6-12, the program appears to have increased school attendance for girls by a small amount relative to boys. There were no gains in enrollment in most specifications, though this may not be surprising in a context where primary school enrollment is already around 90 percent. At the household level, the program may result in a slight reduction of household labor (defined to exclude housework or time allocated to program compliance) for wealthier households relative to poorer households, but a more important change seems to be the shift of productive labor from adult females toward men.

Introduction

Since the well-documented success of Mexico's conditional cash transfer program in inducing greater schooling outcomes and healthcare participation among children, many other countries, especially in Latin America, have opted to implement CCTs in an effort to boost human capital accumulation and fight poverty (Rawlings and Rubio, 2005; Fiszbein and Schady, 2009). A sizable literature has grown around subject of CCTS, including the direct impacts on schooling (Skoufias and Parker, 2006), child labor (Edmonds and Schady, 2008; Skoufias and Parker, 2001; Filmer and Schady 2009c; Maluccio and Flores 2005; Yap, Sedlacek, and Orazem, 2008; Attanasio et al. (2006), Glewwe and Olinto (2004), child health (Gertler, 2004), and consumption (Hoddinott and Skoufias, 2004); role as a safety net for exposure to shocks (de Janvry et al., 2006); and the role of the imposed conditions (Schady and Araujo, 2008; de Brauw and Hoddinott, 2008).

The literature has been relatively more thin, however, on the subject of adult labor effects, especially compared to the number of studies on child labor and CCTs. Faced with the option of a new revenue stream as well as new costs (time, and in some cases financial), households may choose to alter their previous allocations of "productive" labor (here meant to exclude time spend in education or helping others participated in education or health services) to maximize utility over these new conditions, by changing total labor time or shifting labor from some members to other members, depending on program constraints, benefits, and the household's utility function. From a policy perspective, there are a number of reasons one may care about these changes. If programs give households cash transfers, there may be concern over the possibility that households will respond with lower adult labor, potentially at odds with programs' poverty reduction goals (Fiszbein and Schady, 2009; Moffit, 2002; Skoufias and Di Maro, 2006). Such a reduction could result from a preference for leisure or program responsibilities displacing available work hours, potentially exacerbated by any perceived incentives for households reduce work to ensure eligibility for a program. Another concern is that program compliance requirements may fall disproportionately on certain members of the household (particularly mothers and school age children). If compliance meaningfully detracts from their wellbeing through the reduction of leisure, one may worry that household benefit may come at the expense of more vulnerable members (Escobar and Gonzales de la Rocha 2008, Gammage 2010).[27] If new responsibilities lead to other members sharing additional burdens to compensate, however, this potential distortion may be less of a concern, and income directed to the control of the mother may shift intrahousehold dynamics in ways favorable to those members responsible for program compliance (Escobar and Gonzales de la Rocha 2008, Gammage 2010). From an policymaking perspective, measurement and prediction of household labor responses may desirable for being able to predict ex ante or estimate ex post impacts on household income and poverty shifts in response to transfer programs.

Of studies thus far, only Nicaragua's relatively-generous transfer program RPS has been found to generate any meaningful reductions in adult labor, concentrated primary among men (Maluccio and Flores, 2005). In Mexico, Parker and Skoufias (2001) find small short-run reductions in adult labor from Mexico's Progresa, but these effects disappear over time. They are unable to answer the question as to whether the disappearing effect might be attributed to long-run shifts in household preferences, adaptation of productive inputs, spillover effects, or something else. Skoufias and Di Maro (2006) also find few labor impacts using the Mexican

[27]This is not a new concern or one limited to program compliance. For example, Hochschild (1989) complained of the "second shift" among working mothers in the U.S.

Progresa data, while qualitative work by Escobar and Gonzales de la Rocha (2008) shows that women find program compliance onerous and generally incompatible with wage work. Addressing a similar question for children, Edmonds and Schady (2009) find reductions in child labor resulting from participation in Ecuador's cash transfer, even without the educational conditions, but no impact on adult labor.

In this chapter, however, I discuss conditions that would predict a fall in women's productive labor relative to men, with ambiguous effects on labor overall. I then find suggestive but somewhat weak evidence for this pattern, using a panel dataset from El Salvador to examine household responses in time use - specifically educational participation and labor - to the government's Comunidades Solidarias Rurales (CSR, formerly Red Solidaria) conditional cash transfer program during 2008.

El Salvador's Conditional Cash Transfer: Comunidades Solidarias Rurales

In 2005, the government of El Salvador initiated its conditional cash transfer program Red Solidaridad, now Comunidades Solidarias Rurales (CSR) with the goals of increasing human capital among children and also to increase consumption among the poor, who represented approximately 35 percent of the population at that time (CIA World Fact Book, 2011). Described in more detail below, the program was rolled out in several blocks between 2005 and 2009. Eligibility was determined by municipality-level poverty level and marginality index scores (referred to as IIMM hereafter), and all households with kids ages 0-3 and/or 6-12 still in primary school in eligible municipalities were eligible regardless of individual household wealth.[28] The IIMM is a composite score (with an undisclosed formula) generated by the government using poverty levels, education rates, and housing characteristics (de Brauw and Gilligan 2011). Voluntary participation in CSR's CCT gave participant households $15-$20 of additional income per month ($15 for either kids under 4 or primary school students, $20 for both) and imposed various time requirements as well as costs for compliance (for example, transportation costs and any school-related expenses). In particular, all primary school-aged children (6-12, though the upper bound has changed some over time) were required to attend at least 80 percent of possible school days, and mothers were required to take children ages 0-3 for specific sets of check-ups and vaccinations and attend health lectures. [29]

The years during which the CCT program was being rolled out were marked by a few strong macroeconomic trends that the program could not have anticipated. First, the "food price crisis" saw food prices rise around the world over the course of 2007 and peaking in 2008. Then in 2008 and 2009, the financial crisis hit, leading to high levels of unemployment in El Salvador.[30] At the same time, crop production in El Salvador was particularly favorable in

[28]In urban areas, there was also a household proxy means component to eligibility, but in the rural areas for which I have data, all households were eligible.

[29]Attending lectures was not actually a condition to receive the transfer, but a survey found that nearly all participants believed it was a requirement.

[30]http://www.worldbank.org/en/country/elsalvador/overview, accessed 17 April, 2016.

2008, approximately 15 percent higher than in the two previous years. [31]

Theoretical motivations and predictions

In order to look at how the program might affect time allocations in the household, I start with utility-maximizing households, whose members engage in some combination wage labor and own production. When a household becomes eligible for a CCT, it can choose whether or not to participate. Again, participation in CSR's CCT gave participant households $15 or $20 of additional income per month and imposed various time requirements as well as financial costs for compliance (for example, transportation costs and any school-related expenses). [32] Thus, participation gives the household additional income that eases the budget constraint, to be used directly for consumption, educational expenses, investment, or the "purchase" of leisure. Households will choose to participate if net utility change is positive. The "cost" of compliance includes any additional action a household must take over and above what it would have taken otherwise and the distortion this allocation generates over their preferred allocation. Households will choose to participate if net utility gain (additional benefits minus additional costs) is positive. For households whose children already attend school and attend at least 80 percent of possible days a month, the educational condition does not impose any additional cost, and the transfer would be the same as a pure income transfer.

How then might labor be impacted among participating households? As discussed in Parker and Skoufias (2000) and drawing from Killingsworth's (1983) model of household labor supply, insofar as consumption and leisure are both normal goods for all members and the household is at an interior solution with nonzero labor and leisure, transfer income would allow the household to enjoy more of both through an income effect, reducing time allocated to labor and increasing leisure. At the same time, the transfer is conditioned on time allocations to program compliance, which increases the opportunity cost of market labor for school-aged children and mothers and home labor for children, if program compliance is different from their normal allocation of activities. If these members reduce their market labor for program compliance, household income (excluding the transfer) falls, mitigating the income effect on market labor for the remaining members. These effects together lead to a prediction of an increased differential in market labor participation of individuals not responsible for program compliance over those who are responsible for compliance. Gertler et al. (2006) show further that in the presence of liquidity or credit constraints, the transfer could increase market labor through the purchase of complementary inputs such as inventory, fertilizer, or transportation to a work site.

Under this framework, CSR is expected to lead to an increase in time allocation to schooling among primary school age children, which has already been shown by de Brauw and Gilligan (2011) (using a different subset of earlier treatment groups). This could take the form of higher enrollment, but given that primary school enrollment is already around 90 percent,

[31] Source: FAO, via Index Mundi, available at: http://www.indexmundi.com/facts/el-salvador/crop-production-index

[32] As a reference point in 2005, approximately 11 percent of El Salvador's population fell below the $1.25 PPP extreme poverty line (World Bank, World Development Indicators).

gains may more likely to come from higher attendance and time spent on homework among those already enrolled but attending less than the program threshold for benefits, since the marginal cost of compliance is lower relative to unenrolled children. At the point that 90 percent of primary-age children are in school at least some of the time, the marginal cost of enrollment for the remaining households who have not yet chosen to enroll may be in the high and may not be primarily monetary.

Since Gilligan and Peterman (2011) find that the program increased antenatal visits, and in my data, many program participants report attending multiple CSR workshops in addition to checkups and vaccinations for infants, so even if many households were already in compliance for schooling, it is likely that program mothers faced extra time costs for health conditionalities.[33] In this case, women's labor is expected to fall relative to men's. [34] We would not expect to see this if a) program benefits were too small to make an impact, b) households were already compliant with program requirements, or c) intrahousehold bargaining conditions prevent women from reducing labor despite taking on new household responsibilities.

The start of the financial crisis followed by favorable agricultural conditions may have increased the likelihood or severity of liquidity constraints. The transfer would then alleviate these constraints, either for consumption, which would not increase labor, or productive investment, which might increase returns to labor and observed amounts of labor. Wealthier households are less likely to be liquidity constrained, so these households are expected to have a more negative labor response than poorer households. Of course, any labor effects are dependent on program participation. For those households who would have "complied" even without any requirements, we would expect to see no enrollment effects but potentially some attendance effects (if education is a normal good) and labor responses coming from the transfer.

Overall, if the program has any impact, I would predict:

- Women's productive labor time decreases relative to men's.
- Total household "productive" labor time could decrease, increase, or stay the same.
- More negative or less positive labor change for wealthier households.

Two other necessary predictions feed into these labor predictions, since program participation is necessary for the transfer to be received and have any impact.

- Increases in time allocation to schooling/homework among primary school age children, or
- Increased child healthcare activities among mothers (not testable here)

[33] I am unable to estimate from my dataset the total time spent by women specifically for program-required activities.

[34] As an alternative, we could see a shift in other home tasks to men, but cultural gender roles and the fact that program requirements may reinforce women's specialization in home tasks, make this unlikely.

Dataset

I use the first two rounds of a panel dataset collected in El Salvador twice in 2008 for the purpose of monitoring and impact evaluation of the Comunidades Solidarias Rurales (CSR, formally Red Solidaria). The data contain detailed household and individual-level information on demographics, education, income, use of financial services, household production, as well as a thorough time-use module that covers paid, unpaid, and self-employment activities; school work; and household tasks including childcare and homework help. Modules on household production and asset ownership also allow me consider heterogeneity according to baseline wealth proxied by a principle components asset score, following Filmer and Pritchett (2001).

The first "baseline" collection of data occurred very early in 2008, just as the 2007 group ("Group 2") was entering (since "2007" rollout actually only started at the very end of 2007/beginning of 2008) and well before the entrance of the "early 2008" group ("Group 3") and the "late 2008" group ("Group 4"). The 2006 group ("Group 1") was already enrolled at that time, precluding a true baseline with the exception of some basic information collected for administrative purposes prior to roll-out and reducing the number outcomes for which 2006 and 2007 can be compared, and Group 1 is excluded from this analysis.[35] Because the first data collection occurred after the phase-in of CSR was complete for Group 1 and had already begun in a few municipalities in Group 2, retrospective questions about activities from the past year (and even further back for education) were included to recapture some information about households prior to the program.

Data were collected for households and for individuals. The individual outcomes I estimate as dependent variables are:

Education (ages 6-12 and 6-17):

- enrollment
- number of days attended in past four weeks

Labor:

- hours worked, past 12 months
- hours in housework, average week

At the households level, I will use:

- total days of school attendance in the past week, all members
- household per capita hours of labor
- weekly hours in housework

[35] There was also a 2005 group, who were enrolled long before there were any plans for evaluation, and they were excluded from the data collection entirely.

I will also consider household expenditures in agriculture and per capita income, toward the story of liquidity constraints in a favorable agricultural year.

I estimate outcomes at the household and individual level because decisions are being made and budgets aggregated at this level, but because eligibility was assigned at the municipality, all standard errors must be clustered at the municipal level. Once a municipality was determined to be eligible for CSR, all households in the municipality with children who satisfied either of the age criteria were eligible to participate. This will allow me to examine heterogeneous treatment effects across pre-CSR poverty levels proxied by asset ownership and gender. The relevant sample cover 40 municipalities, with 15 households drawn from each of two randomly-selected cantons (subunits) from each municipality.[36]

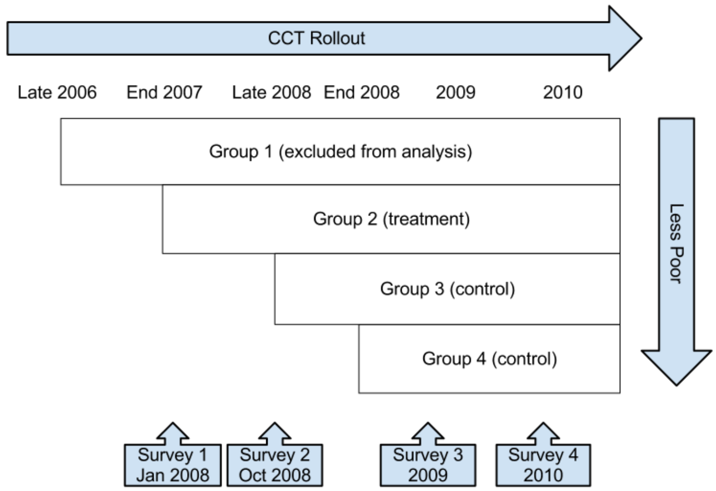
Plan for Estimation and Testing

Identification Strategy

My main identification strategy to examine the impacts of the cash transfer lies in a program rollout that lends itself well to a regression discontinuity design. To accommodate capacity constraints, CSR was phased in over the course of several years (2005-2009), as can be seen in Figure 1. There was no randomized component to program assignment; with poverty-reduction objectives, the government wanted to reach all of the poorest municipalities, starting with the very poorest first. Fortunately for the purposes of identification, however, eligibility and the order of rollout were completely determined by observable characteristics. As described in de Brauw and Gilligan (2011), first a poverty score that combined two municipality-level characteristics (poverty rate and stunting rates for children under three years old) was used to assign municipalities to broad poverty categories, and the two poorest categories (Extreme Poverty and Severe Poverty) would all eventually be enrolled. From there, each of the broader poverty categories was further subdivided to allow for five years of roll-out from 2005-2009.

Figure 1 Program Rollout:

[36] 29 municipalities have IIMM marginality scores within five points of the cutoff between Group 2 and Group 3.

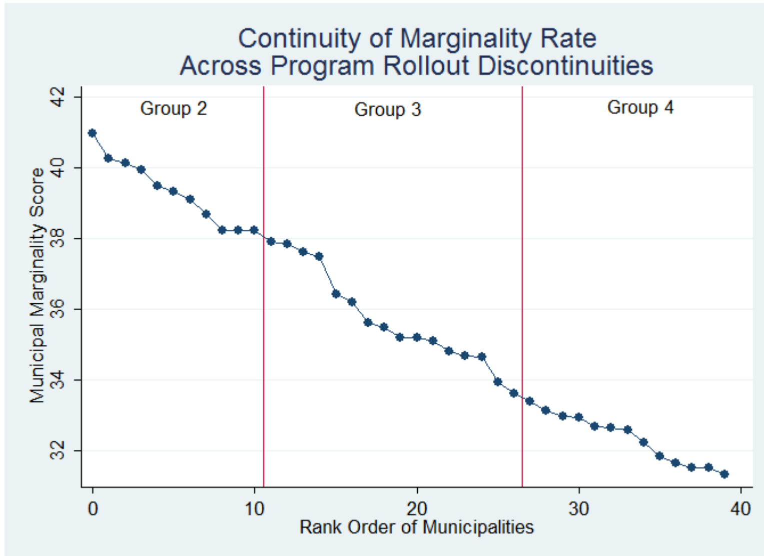


The order of allocation within each broader category was assigned by another previously-determined poverty score, the IIMM ("Marginalization Index," with high scores indicating more poverty along a number of dimensions that included school enrollment rates), with the municipalities with the highest IIMM scores (the poorest) being selected first. Because the number of municipalities receiving CSR in each phase was widely agreed to have been determined by administrative constraints rather than by natural discontinuities in the poverty rate, the cutoff points can be considered arbitrary, and phase assignment generates a natural forcing variable. The Extreme Poverty group received the rollout in 2005 (2005 entrants were completely excluded from data collection and ignored in this analysis) and 2006, and the Severe Poverty group became eligible in three phases in 2007 (IIMM score 38.16 and higher), early 2008 (IIMM score 33.5 and higher), and late 2008, though the program was actually rolled out behind schedule in each case (still I will keep the names 2006, 2007, etc., in keeping with earlier documentation and studies of the program). The underlying characteristics of the IIMM and similarity of municipalities around the threshold are address below in the Summary Statistics section.

Again, since eligibility was completely determined at the municipality level for all household with children in the relevant age ranges, no individual or household level characteristics were factored into eligibility other than how they contributed to municipal averages.[37] In order to alter its own eligibility, a household would have had to known well ahead of time which municipalities would have been selected, and relocated there, which seems implausible, especially for such a small transfer.

Figure 2: Rollout by Marginality Score (IIMM)

[37]Because eligibility is determined at the municipal level, all estimation use municipality-clustered standard errors.



Exploiting the arbitrary nature of the cutoffs between groups in the absence of random assignment, the natural comparisons come from households in municipalities just above each cutoff to households in municipalities below each cutoff. The differences in the expected values of the characteristics of the municipalities included and excluded (and the households who live in them) become smaller and smaller as we consider a more and more narrow window along the running variable (IIMM score) around the cutoff point. This means that in stages, the 2007 group (Group 2 in the figures) can provide comparison households for 2006 (Group 1) prior to the 2007 roll-out (but without a baseline) and the poorer municipalities in the two 2008 groups (Groups 3 and 4) will provide potential comparison households for 2007 prior to the 2008 roll-outs. 29 municipalities have IIMM marginality scores within five points of the cutoff between Group 2 and Group 3 – 11 from Group 2, 16 from Group 3, and two from Group 4. 16 have scores within 2.5 points of the cutoff – 10 from Group 2 and 6 from Group 3. While a smaller cutoff is more preferable for comparable samples, the sample size becomes much smaller in this case, so I will present results for both.

The first collection of data occurred very early in 2008, well before the entrance of the first 2008 group but after the entrance of a "2006" group and just as rollout was beginning to ramp up to the "2007" group, since implementation ran behind schedule. This precludes a baseline for the 2006 group. Because of this, for now I drop Group1 and use only the 2007/2008 comparison (Group 2 vs Group 3 and part of Group 4), to take advantage of the availability of the (imperfect) baseline. Using the panel, however, cannot eliminate the risk that a small part of the 2007 group maybe be partially "contaminated" by a the very beginning of treatment in the first survey, which would tend to attenuate any estimated effects for the 2007/2008 comparison, but the fraction who were enrolled in the program by the baseline is reported to be "very small." Using only the 2007/2008 comparison means the relevant cutoff point corresponds to a relatively lower degree of poverty than the 2006/2007 cutoff,[38] and thus prevents us from saying anything about outcomes for the poorest part of the population. Still, the 2007/2008 comparison allows us to look at impacts closer to

[38] At least along certain dimensions. Because the first division was based on a difference set of character-

the margin of enrollment, which may be useful for policymakers who may want to consider whether to continue expanding coverage to additional municipalities.

For comparison, and because the sample comes from relatively few municipalities with similar poverty scores, I will also provide estimates using a difference in differences approach, which controls for time-invariant differences between may be that by choosing a matched control sample for which the distribution of (observed) characteristics is like that of the treatment sample, it may reduce the risk of time-variant differences unrelated to the actual receipt of treatment.[39]

Outcomes of Interest and Estimating Equations

Difference in differences

I will estimate the following reduced form equations for household and individual outcomes:

The standard difference-in-differences (DD) comparison (used as a reference point for the regression discontinuity) is

$$y_{ihm} = \alpha + \beta_{DD} T_m t_t + \gamma_1 T_m + \gamma_2 t_t + \lambda Z_{ihm} + \epsilon_{ihm},$$

where for individual i in household h in municipality m, the interaction $T_m t_t$ of treatment status T_m and time t_t gives the program effect β_{DD}. T_m controls for any time-invariant differences inherent to those selected to treatment, and t_t captures any time trends between periods for all individuals. Z_{ihm} is a vector of individual level predetermined characteristics.

I add individual child fixed effects to the DD, estimating:

$$y_{ihm} = \beta_{DD} T_m t_t + \gamma_2 t_t + I_{ihm} + \epsilon_{ihm},$$

where I_{ihm} is an individual-level fixed effect. The vector Z_{ihm} is captured by the fixed effect and so is dropped.

Regression discontinuity

Regression discontinuity design (RDD) can be used to treatment effects when inclusion into treatment has been decided by some sort of eligibility threshold. RDD requires three conditions: 1) discontinuity of the probability of treatment at the cutoff point along the eligibility criteria (in our case, early treatment if the municipality has an IIMM score of greater than 38.16, and later treatment otherwise), 2) similarity of characteristics (observable and unobservable) between units on either side of the cutoff, and 3) continuity of the outcome of interest across the threshold in the absence of the intervention. Simply put, RDD compares a just-barely-eligible population to the just-barely-ineligble population, generally controlling

istics, municipalities in the Severe Extreme poverty group do not necessarily have lower IIMM scores than municipalities that received the program later, though the two scores are correlated, as might be expected.

[39]Ideally, I would also be able to test for differences in trends prior to treatment, but with only one baseline survey, I can only look for and control for differences in pre-treatment levels, not changes.

for the slope of the outcome of interest across the eligbility threshold in the absence of treatment (Edmonds, Mammen, and Miller 2005). For regression discontinuity using local linear regressions, I estimate:

$$y_{ihm} = \alpha + \beta_{RDD} T_m t_t + \gamma_1 T_m + \gamma_2 t_t + \gamma_3 D_m + \gamma_4 T_m D_m + \gamma_5 D_m t_t + \gamma_6 T_m D_m t_t + \lambda Z_{ihm} + \epsilon_{ihm}, ,$$

where, combining the RDD and DD, the interaction $T_m t_t$ of treatment status T_m and time t_t gives the program effect β_{RDD}. Here D_m is the running or forcing variable, which gives the distance to the cutoff for inclusion in treatment ($D_m = 0$). Its inclusion captures the slope in the dependent variable that is related to the progressive changes across individuals and households as they are further from the cutoff, assumed to be continuous and linear in this specification.

I again add individual fixed effects to the RDD, estimating:

$$y_{ihm} = \beta_{RDD} T_m t_t + \gamma_2 t_t + \gamma_5 D_m t_t + \gamma_6 T_m D_m t_t + I_{ihm} + \epsilon_{ihm}.$$

I include both the DD and the RDD for comparison, to see how similar their results are. For robustness checks, I also vary the maximum distance from the cutoff point. Additionally, while I cannot establish a causal relationship, for interest, I will explore heterogeneity by including interaction terms in the RDD, child fixed effect specification - asset score[40] at baseline (t=0) and gender (for the individual regressions). This enters as:

$$y_{ihm} = \beta_{RDD} T_m t_t + \gamma_2 t_t + \gamma_5 D_m t_t + \gamma_6 T_m D_m t_t + \beta_{interact} T_m t_t * X_{ihm} + \gamma_7 t_t * X_{ihm} + I_{ihm} + \epsilon_{ihm},$$

where X_{ihm} is assets, farmings, or gender, and $\beta_{RDD} + \beta_{interact} * X_{ihm}$ gives the average effect for each subgroup of interest. Because of eligibility being assigned on the municipal level, standard errors in all specifications are clustered at the municipal level as well.

Small sample considerations and wild bootstrap

Because there are few municipalities, even with municipality-level clustering, standard errors are prone to underestimation bias (leading to over-rejection of the null hypothesis), increasing in the size of the cluster and in the within-cluster correlations of regressors and of errors (Bertrand, Duflo, and Mullainathan 2004, Cameron et al. 2008,). To correct for this, I use the wild cluster bootstrap with asymptotic refinement as proposed by Cameron et al. (2008) to generate p-values alongside the standard clustered standard errors. The wild bootstrap has the advantage of accommodating errors that are not iid across clusters or where clusters are not perfectly the same size. [41]

[40] My asset index is a Principal Components score based on a series of indicator variables for owning various household assets. Assets = 0 falls between the median and mean values for the sample.

[41] To implement this procedure, I draw from Stata code used by Eduardo Montoya in his dissertation chapter "Violence and economic disruption: firm-level evidence from Mexico."

Summary Statistics

The panel data were collected as municipality-representative, unbalanced panel (to ensure inclusion of households with the very youngest children). Because there is no information for the added households for the period $t = 0$ – limiting the number of controls and precluding the use of child or household fixed effects or PSM – my panel analysis will only use the portion of the sample for whom there are observations for the beginning and end of 2008 ($t = 0$ and $t = 1$, respectively). Table 1 gives baseline summary statistics for households in $t = 0$ for the entire unbalanced sample (within bandwidth 5 of cutoff point 38.16) of 1539 households in the first column and for the balanced sample of 1236 over a selected demographic and productive characteristics in the second column. The third column gives the differences between the full unbalanced and restricted balanced samples, and we see that there are no significant differences, so I will only used the balanced panel portion going forward. [42]

Household Characteristics

Note that the use of the balanced subsample reduces representation of households whose only eligible children were on the older end in the first year or the younger end in the second year.

[42] I do not have data on attrition, though it was reported to be low (in a conversation with Alan de Brauw, affiliated with the IFPRI team that managed data collection for the program evaluation). The excluded sample comprises attriters as well as those rotated out of the sample, but given the lack of significant differences between the whole sample and the balanced subsample, concerns about bias or unrepresentativeness introduced from the use of the subsample may be minimal.

Table 1: Baseline Household Comparisons: Bandwidth 5
Unbalanced and Balanced Panel, Treatment and Control

	Unbalanced Total	Balanced Total	Diff U-B	Balanced Subset Treatment	Balanced Subset Control	Balanced Subset Diff
Household size	5.745	5.791	-0.0460	5.817	5.771	0.0465
	(2.246)	(2.196)	(0.0847)	(2.285)	(2.124)	(0.126)
Female head	0.242	0.235	0.00774	0.229	0.239	-0.0110
	(0.429)	(0.424)	(0.0163)	(0.420)	(0.427)	(0.0243)
Highest schooling	6.605	6.649	-0.0439	6.546	6.730	-0.184
	(2.874)	(2.824)	(0.109)	(2.750)	(2.881)	(0.162)
Female schooling	3.006	3.012	-0.00564	2.746	3.223	-0.477**
	(3.148)	(3.129)	(0.120)	(3.081)	(3.154)	(0.179)
Time to school, min	15.93	15.78	0.145	16.02	15.60	0.416
	(15.12)	(15.18)	(0.579)	(14.99)	(15.33)	(0.869)
Housing index	2.361	2.357	0.00383	2.311	2.393	-0.0825*
	(0.603)	(0.607)	(0.0231)	(0.659)	(0.561)	(0.0347)
Household grew crops	0.657	0.667	-0.00975	0.689	0.649	0.0404
	(0.475)	(0.472)	(0.0181)	(0.463)	(0.478)	(0.0270)
Household had animals	0.606	0.612	-0.00541	0.600	0.621	-0.0216
	(0.489)	(0.488)	(0.0186)	(0.490)	(0.485)	(0.0279)
Land cultivated, mn	0.966	0.993	-0.0266	1.045	0.952	0.0931
	(1.496)	(1.565)	(0.0586)	(1.652)	(1.492)	(0.0896)
Per capita inc,no ag	356.5	364.9	-8.404	325.7	396.1	-70.41
	(959.6)	(1030.3)	(38.17)	(1041.8)	(1020.7)	(58.99)
Per capita income	423.1	422.3	0.831	361.9	470.2	-108.3+
	(1202.1)	(1148.7)	(44.79)	(1041.5)	(1225.7)	(65.74)
Per capita ag profit	66.65	57.41	9.236	36.29	74.18	-37.89
	(742.5)	(526.7)	(24.14)	(109.2)	(698.5)	(30.16)
Per capita days labor	81.92	81.87	0.0515	83.92	80.24	3.672
	(57.77)	(56.76)	(2.185)	(57.61)	(56.06)	(3.250)
N	1539	1236		547	689	

Standard deviation/error in parentheses, + $p<0.10$, * $p<0.05$, ** $p<0.01$
Other variables not found to be different: # 0-3, # 6-12, # 13-17, # 18-64, age of head, p/c labor hours, harvest value.

Looking at the differences between treatment and control within the balanced subset, the statistically significant differences we see are that in treatment households in $t = 0$, average schooling attainment among primary adult females is lower by almost half a grade and the housing quality index scores (0-3, one point for "improved" materials for each of floor, walls, and roof) are slightly lower. Per capita income is also 23 percent lower (significant at 10 percent). Notice that approximately two-thirds of households grew crops and nearly as many raised animals (these do not overlap for some households). Using a bandwidth of 5, which includes all of the 2007 rollout and much of the 2008 groups, the average for treatment households is higher for crops, but the difference is not significantly different.

Dropping down to a bandwidth of 2.5 as shown in Table 2, we see that treatment households were significantly less likely to have crops or animals than the comparison group, though average amount of land cultivated is similar between the two. Treatment households are also significantly closer to the nearest school, though only by an average of four minutes. Female education and housing are no longer significantly different. Given that the number of municipalities drops to only 16 with the 2.5 bandwidth, we may consider the sample with bandwidth 5 as the primary sample.

Seeing significant differences at "baseline" reduces confidence that those excluded by the treatment cutoff are an ideal comparison for those included and increases the importance of including controls and individual fixed effects. Moreover, we may worry when using the smaller bandwidth that farming outcomes may be particularly subject to time variant effects because of, say, weather or prices, which would violate the assumptions needed for identification using individual fixed effects.

Table 2: Baseline Household Comparisons: Bandwidth 2.5
Unbalanced and Balanced Panel, Treatment and Control

	Unbalanced Total	Balanced Total	Diff U-B	Balanced Subset Treatment	Control	Diff
Household size	5.626	5.669	-0.0431	5.691	5.622	0.0686
	(2.112)	(2.123)	(0.106)	(2.184)	(1.992)	(0.169)
Female head	0.236	0.225	0.0104	0.232	0.210	0.0220
	(0.425)	(0.418)	(0.0211)	(0.423)	(0.408)	(0.0332)
Highest schooling	6.511	6.546	-0.0353	6.638	6.351	0.287
	(2.840)	(2.785)	(0.141)	(2.769)	(2.815)	(0.221)
Female schooling	2.833	2.845	-0.0120	2.803	2.935	-0.132
	(3.114)	(3.081)	(0.155)	(3.148)	(2.939)	(0.245)
Time to school, min	16.79	16.71	0.0754	15.42	19.45	-4.033**
	(15.96)	(16.42)	(0.814)	(14.72)	(19.30)	(1.297)
Housing index	2.341	2.341	0.000302	2.343	2.335	0.00867
	(0.644)	(0.648)	(0.0324)	(0.670)	(0.601)	(0.0515)
Household grew crops	0.681	0.694	-0.0129	0.671	0.742	-0.0718+
	(0.466)	(0.461)	(0.0233)	(0.470)	(0.438)	(0.0366)
Household had animals	0.598	0.599	-0.000503	0.570	0.661	-0.0912*
	(0.491)	(0.490)	(0.0246)	(0.496)	(0.474)	(0.0388)
Land cultivated, mn	0.937	0.944	-0.00697	0.921	0.992	-0.0711
	(1.402)	(1.417)	(0.0708)	(1.505)	(1.209)	(0.113)
Per capita inc,no ag	342.7	348.6	-5.862	343.8	358.7	-14.86
	(1023.0)	(1081.4)	(52.95)	(1090.7)	(1063.6)	(85.97)
Per capita income	401.3	382.6	18.68	377.2	394.2	-16.98
	(1267.1)	(1080.9)	(58.67)	(1090.8)	(1061.6)	(85.93)
Per capita ag profit	58.58	34.04	24.54	33.36	35.48	-2.121
	(766.2)	(102.7)	(26.20)	(109.5)	(86.84)	(8.166)
Per capita days labor	84.51	83.59	0.925	83.82	83.09	0.734
	(59.10)	(57.33)	(2.917)	(58.96)	(53.81)	(4.558)
N	874	728		495	233	

Standard deviation/error in parentheses, + $p<0.10$, * $p<0.05$, ** $p<0.01$
Other variables not found to be different: # 0-3, # 6-12, # 13-17, # 18-64, age of head, p/c labor hours, harvest value.

Additionally, to compare the treatment and control groups, I estimate the logit regression on baseline X's for bandwidth 5 and 2.5, presented in Table 1b. Individually, only distance to school is marginally statistically significant, and only for bandwidth 2.5, but in both specifications the X's were jointly statistically significant (p-value < 0.01). I include the household-level comparison here, with the p-value for the joint F test in the last row below the number of observations:

Table 1b: Logistic Regression Predicting Treatment from Baseline Statistics
Dependent Variable: =1 if Early Treatment Municipality
Bandwidths 5 and 2.5

	band = 5	band = 2.5
Household size	-0.0334	0.00195
	(-0.75)	(0.03)
Female head	-0.0469	-0.0156
	(-0.22)	(-0.07)
Highest schooling	0.0281	0.0544
	(0.62)	(0.94)
Female schooling	-0.0480	-0.0514
	(-1.53)	(-1.01)
Time to school, min	-0.00128	-0.0108+
	(-0.31)	(-1.95)
Housing index	-0.110	-0.0516
	(-0.44)	(-0.18)
Household grew crops	0.211	-0.398
	(0.71)	(-1.32)
Household had animals	-0.0298	-0.0537
	(-0.12)	(-0.17)
Qty land cultivated (tareas)	0.0250	0.00677
	(0.72)	(0.34)
Per capita household income	0.000114	-0.000594
	(0.83)	(-0.96)
Per capita agric. income	-0.000114	0.000612
	(-1.32)	(1.05)
Asset index score	-0.0981	-0.0190
	(-1.36)	(-0.14)
Per capita days labor, annual	0.00131	0.00171
	(1.42)	(1.23)
Constant	-0.0519	1.017
	(-0.06)	(0.95)
N	2496	1464
p-value for joint F test:	0.00784	0.00000

Notes: t-statistics in parentheses; ste errors clustered at the municipal level; + p<0.10, * p<0.05, ** p<0.01; housing index ranges 0-3, with 1 point each for improved roof, walls, and floor. Asset index is a principal components score (Filmer and Pritchett 2001) based on indicator variables for owning any of each of the following: radio, stereo, tv, vcr/dvd, video games, fan, computer, typewriter, sewing maching, vehicle, boat, bike, cart, motorcycle, fridge, stove, mill, blender, generator, solar panels, bed, furniture.

Marginalization Index Scores - IIMM

The IIMM marginalization index was a measure of 1) poverty levels, 2) education rates, and 3) housing quality, combined according to an undisclosed government algorithm at the municipal level. We are interested in knowing how similar households are within our chosen bandwidths. For the purposes of my analysis, the similarity of the sampled households that are being compared to each other is more relevant than the similarity of their underlying municipalities, though these are the same in expectation. From the household and individual descriptive statistics, using the 5 point bandwidth, there are differences between treatment and comparison groups for both income and schooling for children ages 6 to 12, though these are not statistically significant. These go in opposite directions: average per capita household income at baseline was quite low for both groups but 30 percent higher for the comparison group (470 USD vs 362 USD), and reported enrollment rates were higher for the treatment group (92.1% vs. 88.7%). The composite nature of the index and the use of sampling allow for higher schooling rates among just-included municipalities relative to just-excluded municipalities. For the 2.5 bandwidth, these differences become even smaller betwen treatment and control. There were no significant differences in the housing quality index (0 to 3, with 1 point each for "improved" material for floor, walls, and roof) at either bandwidth. Using a logistic regression to use the three elements – income, housing, and schooling – to predict treatment, we see that only schooling remains marginally statistically significant using a bandwidth of 5, and the elements are not jointly significant using either bandwidth (p-values 0.158 and 0.821 for 5 and 2.5, respectively).

Table 1c: Comparison of IIMM Index Components by Treatment
Logistic Regression, Dependent Variable: Treatment

	bandwidth 5	bandwidth 2.5
Per capita income	-0.0000886	-0.0000278
	(0.489)	(0.719)
Housing quality index	-0.253	0.00397
	(0.334)	(0.989)
% enrollment ages 6-12	0.569+	0.430
	(0.069)	(0.457)
Constant	-0.109	0.362
	(0.894)	(0.751)
Observations	978	570
p-value for joint F test	0.158	0.821

p-values in parentheses; + p<0.10, * p<0.05, ** p<0.01; housing index ranges 0-3, with 1 point each for improved roof, walls, and floor; std errors clustered at the municipality level.

Individual Characteristics

Tables 3 and 4 compare treatment and control characteristics at the individual level within the balanced panel at baseline, for adults ages 18-64, youth ages 13-17, and children ages

6-12 using bandwidths 5 and 2.5, respectively. In the actual estimation, however, I only the 13-17 age group in a sample combined with the younger children, because of the small sample of youths, reflecting the fact the sample was chosen in a way to prioritize the availability of the younger children.

Looking at adults, under bandwidth 5 (975 treatment, 467 controls), treatment individuals, are more likely to have done any work, worked 15.6 more days on average, are more likely to have engaged in any business and in agriculture. Using bandwidth 2.5, they are more likely to have engaged in non-agricultural work, and this is the only significant difference.

I have included 13-17 year-olds, but note that the sample is quite small, especially with the balanced panel (since households more more likely to be dropped from the sample if they were no longer in the age-eligibility range). With bandwidth 5 (352 treatment, 401 controls), treatment are more likely to work and for more days, more likely to engage in family work, and less likely to engage in non-agricultural work (though the overall percentage is small). No differences are significant with bandwidth 2.5, but the sample has fallen to 311 treatment and 134 controls.

For 6-12 year-olds, the treatment group has a significantly lower proportion of males (47.6 versus 52.3) Perhaps surprisingly, treatment kids are more likely to be in school, at 92 versus 89 percent[43], and also more likely to have engaged in any work, family work, and non-agricultural work. At bandwidth 2.5 these differences cease to be significant, but again the sample has fallen to 699 treatment and only 333 controls.

The pre-treatment differences I find between treatment and control provide strong support for the use of the panel component of my dataset to control for household and individual characteristics or fixed effects.

[43]This is perhaps a puzzling finding. One potential explanation may be that since most households in the 2007 knew by the time of the survey that they would be part of the program, there may have been a greater perception that they should have had their children enrolled in school, or that their future benefits might be affected. Another possibility is an anticipation effect, but this should be unlikely, since the news about the program arriving to the community would have come toward the end of the year in question, and school decisions would have been made at the beginning. At the same time, school enrollment rates were only one part of IIMM score that was used as the selection criterion for program assignment, so differences are possible.

Table 3: Baseline Individual Comparisons: Treatment and Control, Bandwidth 5

	Age 18-64			Age 13-17			Age 6-12		
	Treatment	Control	Diff	Treatment	Control	Diff	Treatment	Control	Diff
Age (years)	34.11	34.04	0.0635	14.72	14.63	0.0914	9.020	8.993	0.0271
	(11.52)	(11.79)	(0.473)	(1.386)	(1.349)	(0.0995)	(1.948)	(1.930)	(0.0920)
Gender = male	0.442	0.423	0.0186	0.501	0.517	-0.0160	0.476	0.523	-0.0473*
	(0.497)	(0.494)	(0.0201)	(0.501)	(0.500)	(0.0365)	(0.500)	(0.500)	(0.0237)
Has own children in house	0.768	0.736	0.0318+	0.0479	0.0373	0.0106	-	-	-
	(0.422)	(0.441)	(0.0175)	(0.214)	(0.190)	(0.0147)	(-)	(-)	(-)
In school in recent year	-	-	-	0.718	0.766	-0.0479	0.921	0.887	0.0343*
	(-)	(-)	(-)	(0.450)	(0.424)	(0.0318)	(0.270)	(0.317)	(0.0141)
Any homework, 12 mo	0.0331	0.0266	0.00655	0.532	0.577	-0.0447	0.702	0.729	-0.0271
	(0.179)	(0.161)	(0.00685)	(0.500)	(0.495)	(0.0362)	(0.458)	(0.445)	(0.0214)
Weekly hours in hmwk	0.244	0.233	0.0110	4.363	4.957	-0.594	4.959	5.183	-0.224
	(1.582)	(2.160)	(0.0781)	(6.175)	(6.412)	(0.459)	(5.581)	(5.387)	(0.260)
Engaged in any work	0.689	0.632	0.0568**	0.420	0.311	0.109**	0.0854	0.0587	0.0267*
	(0.463)	(0.482)	(0.0192)	(0.494)	(0.463)	(0.0348)	(0.280)	(0.235)	(0.0121)
Engaged in own business	0.462	0.395	0.0676**	0.0451	0.0299	0.0152	0.00126	0	0.00126
	(0.499)	(0.489)	(0.0200)	(0.208)	(0.170)	(0.0138)	(0.0354)	(0)	(0.00112)
Engaged in family work	0.117	0.0942	0.0228+	0.313	0.239	0.0739*	0.0854	0.0557	0.0297*
	(0.322)	(0.292)	(0.0124)	(0.464)	(0.427)	(0.0324)	(0.280)	(0.229)	(0.0120)
Engaged in paid work	0.236	0.260	-0.0238	0.0817	0.0821	-0.000399	0.00377	0.00498	-0.00121
	(0.425)	(0.439)	(0.0175)	(0.274)	(0.275)	(0.0200)	(0.0613)	(0.0704)	(0.00316)
Engaged in agriculture	0.541	0.494	0.0475*	0.366	0.313	0.0528	0.0766	0.0557	0.0209+
	(0.499)	(0.500)	(0.0202)	(0.482)	(0.464)	(0.0344)	(0.266)	(0.229)	(0.0117)
Engaged in nonag work	0.202	0.191	0.0104	0.0563	0.0174	0.0389**	0.0138	0.00498	0.00884*
	(0.401)	(0.393)	(0.0161)	(0.231)	(0.131)	(0.0135)	(0.117)	(0.0704)	(0.00445)
Individual income	664.1	803.8	-139.8	108.8	108.9	-0.0604	0.206	1.027	-0.821
	(2901.4)	(3349.1)	(128.0)	(483.3)	(497.7)	(35.76)	(4.109)	(20.48)	(0.737)
Annual hours of work	1062.1	984.4	77.68	435.0	377.5	57.51	52.23	39.05	13.17
	(1178.9)	(1191.5)	(48.02)	(698.8)	(824.9)	(55.96)	(246.5)	(215.9)	(10.91)
Annual days of work	173.0	157.3	15.72*	81.30	59.31	21.99**	11.37	8.348	3.017
	(162.0)	(154.6)	(6.395)	(120.4)	(110.1)	(8.380)	(48.50)	(43.48)	(2.172)
N	1086	1391		355	402		796	1005	

Standard deviation/error in parentheses, + $p<0.10$, * $p<0.05$, ** $p<0.01$

Balanced panel observations only.

Table 4: Baseline Individual Comparisons: Treatment and Control, Bandwidth 2.5

	Age 18-64			Age 13-17			Age 6-12		
	Treatment	Control	Diff	Treatment	Control	Diff	Treatment	Control	Diff
Age (years)	34.12	33.59	0.537	14.71	14.58	0.129	9.041	9.141	-0.0997
	(11.62)	(11.45)	(0.651)	(1.379)	(1.395)	(0.142)	(1.945)	(1.911)	(0.129)
Gender = male	0.439	0.433	0.00643	0.494	0.533	-0.0397	0.466	0.474	-0.00809
	(0.497)	(0.496)	(0.0279)	(0.501)	(0.501)	(0.0515)	(0.499)	(0.500)	(0.0333)
Has own children in house	0.762	0.775	-0.0131	0.0350	0.0519	-0.0168	-	-	-
	(0.426)	(0.418)	(0.0238)	(0.184)	(0.223)	(0.0202)	(-)	(-)	(-)
In school in recent year	-	-	-	0.742	0.726	0.0161	0.923	0.907	0.0158
	(-)	(-)	(-)	(0.438)	(0.448)	(0.0454)	(0.267)	(0.291)	(0.0183)
Any homework, 12 mo	0.0318	0.0193	0.0125	0.557	0.526	0.0314	0.694	0.736	-0.0419
	(0.176)	(0.138)	(0.00924)	(0.497)	(0.501)	(0.0513)	(0.461)	(0.442)	(0.0303)
Weekly hours in hmwk	0.244	0.116	0.128	4.714	4.050	0.664	4.966	4.968	-0.00212
	(1.592)	(0.998)	(0.0803)	(6.410)	(5.381)	(0.630)	(5.662)	(4.712)	(0.358)
Engaged in any work	0.681	0.647	0.0343	0.395	0.356	0.0393	0.0801	0.0721	0.00804
	(0.466)	(0.479)	(0.0265)	(0.490)	(0.480)	(0.0501)	(0.272)	(0.259)	(0.0178)
Engaged in own business	0.449	0.437	0.0124	0.0382	0.0444	-0.00623	0.00143	0	0.00143
	(0.498)	(0.497)	(0.0280)	(0.192)	(0.207)	(0.0202)	(0.0378)	(0)	(0.00207)
Engaged in family work	0.109	0.101	0.00808	0.296	0.311	-0.0149	0.0787	0.0721	0.00661
	(0.311)	(0.301)	(0.0173)	(0.457)	(0.465)	(0.0473)	(0.269)	(0.259)	(0.0177)
Engaged in paid work	0.246	0.255	-0.00866	0.0764	0.0741	0.00236	0.00286	0.00300	-0.000142
	(0.431)	(0.436)	(0.0244)	(0.266)	(0.263)	(0.0273)	(0.0535)	(0.0548)	(0.00359)
Engaged in agriculture	0.528	0.537	-0.00927	0.354	0.378	-0.0243	0.0687	0.0661	0.00260
	(0.499)	(0.499)	(0.0281)	(0.479)	(0.487)	(0.0495)	(0.253)	(0.249)	(0.0168)
Engaged in nonag work	0.209	0.165	0.0443*	0.0446	0.0148	0.0298	0.0143	0.00901	0.00530
	(0.407)	(0.371)	(0.0223)	(0.207)	(0.121)	(0.0191)	(0.119)	(0.0946)	(0.00743)
Individual income	711.3	657.7	53.53	94.46	64.77	29.69	0.114	0.432	-0.318
	(3049.9)	(2416.7)	(161.0)	(428.7)	(338.1)	(41.54)	(3.026)	(7.891)	(0.341)
Annual hours of work	1053.8	1000.2	53.61	402.7	431.2	-28.44	46.68	51.55	-4.869
	(1185.5)	(1168.8)	(66.41)	(679.1)	(940.4)	(78.92)	(229.8)	(244.3)	(15.62)
Annual days of work	171.1	162.2	8.941	76.14	69.60	6.543	10.72	11.41	-0.690
	(162.9)	(154.4)	(9.017)	(119.6)	(119.9)	(12.32)	(47.95)	(50.56)	(3.250)
N	975	467		314	135		699	333	

Standard deviation/error in parentheses, + $p<0.10$, * $p<0.05$, ** $p<0.01$

Balanced panel observations only.

Empirical Findings

I present the results in a set of three tables (one for the household, one for each age group). In order to include the relatively large number of outcome variables and specification, I present only the coefficient on the interaction of treatment and time (β_{DD} and β_{RDD}) for the first two specifications and then β_{DD} and β_{RDD} as will as the treatment and time interacted with wealth and then gender.

Household Results

Table 5 presents results for the household regressions. Looking at RDD with bandwidth 5, there is a positive coefficient of 4.6 days of school attendance (summed among all members) in the last four weeks that school was in session. With the regular clustered standard errors, this result is significant at 5 percent, but falls to 10 percent with the bootstrapped errors – a reminder that standard methods may lead to over-rejection of the null for small samples. At the same time, the DD coefficient is nearly zero, suggesting the RDD result is not very robust. With the smaller bandwidth, the coefficients are both halfway between the bandwidth 5 results, pointing to variability across municipalities in our sample range.

For per capita labor, DD and RDD results show opposite signs, but both values are small over 12 months. When treatment is also interacted with asset scores,[44] the base treatment effects keep their signs, but both DD and RDD show negative coefficients of -24 days for the interaction term (p-values 0.06 and 0.076). This would mean that wealthier households are more likely to decrease their labor hours in response to the program. Since the baseline asset index is mean zero, with 25th, 50th, and 75th percentiles of -1.6,-0.5, and 0.94, respectively. Thus the coefficient of -24 would translate into a 61 hour per capita difference in impact between a household at the 25th percentile versus the 75th percentile. At a little more than an hour a week, a fairly subtle difference. At the mean wealth level, there is no significant change in work hours, giving no support to the possibility that households as a whole are increasing labor (as might be expected if households were using the transfers to ease a credit constraint to purchase more labor-complementing inputs). There are no significant results for household tasks, and coefficients are small, suggesting the program did not induce meaningful increases in tasks overall, though infrequent but time-consuming required tasks may not have been reported for "an average week." All overall results for per capita income (excluding the transfer) have positive coefficients, but none is statistically significant. Overall, there are no large or robust impacts at the household level.

[44] Again, the asset index is a Principal Components score based on a series of indicator variables for owning various household assets. Assets = 0 falls between the median and mean values for the sample.

Table 5 -- Household Results

	Overall		Asset Score Interaction			
	DD	RDD	DD		RDD	
Dep vars:	b_T	b_T	b_T	b_T*asset	b_T	b_T*asset
Bandwidth 5						
Days in school, last month	0.489	4.589*	0.220	0.364	4.872*	0.390
clustered p-value	(0.743)	(0.042)	(0.879)	(0.475)	(0.040)	(0.455)
wild cluster p-value	(0.695)	(0.088)	(0.902)	(0.447)	(0.076)	(0.439)
Hours work p/c, 12 mo	28.10	-36.51	31.43	-24.11+	-44.69	-23.60+
clustered p-value	(0.532)	(0.527)	(0.481)	(0.067)	(0.423)	(0.065)
wild cluster p-value	(0.551)	(0.591)	(0.467)	(0.060)	(0.487)	(0.076)
Weekly housework p/c	0.351	-0.0122	0.249	0.229	0.107	0.233
clustered p-value	(0.612)	(0.991)	(0.707)	(0.536)	(0.924)	(0.533)
wild cluster p-value	(0.655)	(0.998)	(0.667)	(0.579)	(0.986)	(0.551)
Income per capita, 12 mo	91.08	134.5	67.82	-7.861	148.0	-4.876
clustered p-value	(0.131)	(0.364)	(0.315)	(0.897)	(0.348)	(0.937)
wild cluster p-value	(0.132)	(0.387)	(0.311)	(0.934)	(0.427)	(0.926)
Bandwidth 2.5						
Days in school, last month	2.919+	2.702	2.899+	-0.139	2.826	-0.208
clustered p-value	(0.081)	(0.270)	(0.074)	(0.830)	(0.253)	(0.758)
wild cluster p-value	(0.084)	(0.351)	(0.092)	(0.695)	(0.355)	(0.790)
Hours work p/c, 12 mo	-11.50	-14.48	-19.59	-40.17*	-37.68	-35.31*
clustered p-value	(0.840)	(0.819)	(0.705)	(0.023)	(0.518)	(0.030)
wild cluster p-value	(0.826)	(0.826)	(0.623)	(0.032)	(0.647)	(0.052)
Weekly housework p/c	0.444	-0.556	0.438	-0.0490	-0.491	-0.0700
clustered p-value	(0.578)	(0.688)	(0.585)	(0.931)	(0.742)	(0.907)
wild cluster p-value	(0.587)	(0.703)	(0.615)	(0.958)	(0.750)	(0.950)
Income per capita, 12 mo	110.9	210.0	102.5	-45.78	214.6	-34.70
clustered p-value	(0.285)	(0.299)	(0.362)	(0.413)	(0.346)	(0.586)
wild cluster p-value	(0.387)	(0.407)	(0.439)	(0.495)	(0.571)	(0.659)

Only treatment effects and interactions presented. Full results available.
Errors clustered at municipal level. + 0.1, *0.05, ** 0.01

Individual Results

Adults 18-64

The results for adult individuals are included in Table 6. There are no statistically significant impacts overall for productive labor or housework, echoing the household findings. There is a negative coefficient of 124.5 hours for RDD of bandwidth 5, but the point estimates move around a lot in the different specifications. Once we allow the interaction by asset score, however, all specifications again show a negative coefficient for higher asset scores. This pattern suggests that wealthier households are more likely to use the transfer to afford more leisure relative to poorer households. The p-values all increase with the wild cluster bootstrap but remain below 0.05 for RDD and DD at both bandwidths. The base coefficients are less consistent across specifications and are not individually significant, but are generally negative. For gender, too, both specifications with bandwidth 5 show that in households in treatment communities, there is an increased disparity between the hours of productive work for men and women, with a negative coefficient for the base treatment effect (-276, marginally significant at 10%) but a large positive coefficient for men. At nearly 350 hours for 12 months this is a difference of nearly an hour a day. The total productive work with treatment may be falling for women and appears to be increasing for men. This support the hypothesis that with increased program demands for women's time, they reduce work activities. The increase for men suggests that men are replacing more of those hours. Under bandwidth 2.5, all of the coefficients in the gender specification are qualitatively similar but none is significant.

No specification for housework is significant, but the gender coefficients are the opposite of those for market labor: positive base coefficients and negative coefficients for treated men. Unfortunately, phrasing of the questions in the survey did not make it clear whether or which tasks related to program compliance may have been included in home tasks, and some tasks may have been time consuming but happened on an irregular basis, potentially not factoring into a "typical" week.

Table 6 -- Individual Results: Adults, Age 18-64

	Overall		Asset Score Interaction				Gender Interaction			
	DD	RDD	DD		RDD		DD		RDD	
	b_T	b_T	b_T	b_T*asset	b_T	b_T*asset	b_T	b_T*male	b_T	b_T*male
Bandwidth: 5										
Hours work, 12 mo	23.50	-124.5	30.57	-74.39*	-151.4	-71.09*	-111.0	344.0**	-276.4+	347.2**
clustered p-value	(0.834)	(0.416)	(0.782)	(0.011)	(0.301)	(0.012)	(0.273)	(0.002)	(0.053)	(0.002)
wild cluster p-value	(0.874)	(0.451)	(0.774)	(0.020)	(0.355)	(0.024)	(0.295)	(0.012)	(0.080)	(0.008)
Hours housework, weekly	1.591	0.442	1.461	0.802	0.766	0.782	2.468	-1.743	1.239	-1.711
clustered p-value	(0.261)	(0.810)	(0.296)	(0.281)	(0.696)	(0.291)	(0.244)	(0.475)	(0.525)	(0.481)
wild cluster p-value	(0.339)	(0.874)	(0.351)	(0.319)	(0.727)	(0.275)	(0.327)	(0.527)	(0.579)	(0.507)
Bandwidth: 2.5										
Hours work, 12 mo	-58.30	7.182	-88.15	-113.9**	-55.98	-94.23**	-150.3	222.1	-85.82	219.7
clustered p-value	(0.714)	(0.966)	(0.539)	(0.002)	(0.714)	(0.004)	(0.263)	(0.167)	(0.594)	(0.168)
wild cluster p-value	(0.754)	(0.962)	(0.583)	(0.004)	(0.774)	(0.028)	(0.283)	(0.251)	(0.723)	(0.244)
Hours housework, weekly	1.321	-1.050	1.448	0.802	-0.676	0.782	3.290	-1.743	0.966	-1.711
clustered p-value	(0.388)	(0.605)	(0.366)	(0.752)	(0.780)	(0.777)	(0.148)	(0.217)	(0.643)	(0.226)
wild cluster p-value	(0.423)	(0.683)	(0.379)	(0.854)	(0.782)	(0.758)	(0.160)	(0.255)	(0.671)	(0.248)

Only treatment effects and interactions presented. Full results available. Std. errors clustered at municipal level. + 0.1, *0.05, ** 0.01

Children 6-12

For 6 to 12 year olds, none of the coefficients for being enrolled in school at any point in the year is significant, with the exception of a small negative for the interaction of treatment with asset score marginally significant only under the difference in differences specification – there appears to be little impact. This was another case in which the small sample bootstrapping for standard errors pushed the results from the multiple asset specifications from marginally significant under standard methods to insignificant at the 10 percent level. Though insignificant, the signs are in a direction that makes sense, since wealthier children already tend to have higher enrollment in the absence of the program. These finding differ from the general findings of de Brauw and Gilligan (2011) using larger census data on enrollment, but they were studying a comparison at an earlier phase of the rollout, between groups at higher poverty rates. Again, the lack of impact could be partially attributed to the fact that once enrollment rates are quite high, $15 a month may not be enough to draw non-compliers into school.

For attendance in the last four weeks when school was in session, which is set to zero for children who never enrolled or who dropped out, again there are no overall impacts and coefficients for the wealth specifications are in the same direction as enrollment but with the small sample bootstrapped errors fall out of marginal significance. There is a small result that is significant in all gender specifications, however. A positive base coefficient and negative male interactions that suggests that there may be a small increase in attendance for girls, either because they go more often when enrolled or are slightly less likely to drop out. The coefficients for RDD at bandwidth 5, though only bordering on 10 percent significance, would represent about 10 percent more time in school. Given that attendance was fairly high at baseline, this might be a reasonable gain but would constitute only a minimal additional constraint on the amount of time children have available for labor or leisure.

There are no significant differences in productive labor, and point estimates are small for 12 months, and there are few noteworthy results for time spent in household tasks, except with the gender interaction. The negative base coefficient and positive male interaction suggests that girls may have decreased housework slightly, with little impact for boys, and more precisely, that boys experienced an increase relative to girls. This result would be compatible with a slight attendance increase for girls, or if mothers were taking over more home tasks while engaged in less work outside the home. With bandwidth 2.5, there are no significant labor outcomes, though magnitudes are qualitatively comparable.

Table 7 -- Individual Results: Children, Age 6-12

	Overall		Asset Score Interaction				Gender Interaction			
	DD	RDD	DD		RDD		DD		RDD	
Dep vars:	b_T	b_T	b_T	b_T*asset	b_T	b_T*asset	b_T	b_T*male	b_T	b_T*male
Bandwidth 5										
Enrolled in school	0.0205	0.00550	0.0115	-0.019*	0.00058	-0.0179+	0.0210	-0.0012	0.0059	-0.0008
clustered p-value	(0.259)	(0.859)	(0.499)	(0.048)	(0.985)	(0.056)	(0.346)	(0.969)	(0.858)	(0.979)
wild cluster p-value	(0.303)	(0.946)	(0.471)	(0.080)	(0.950)	(0.124)	(0.343)	(0.990)	(0.770)	(0.962)
Days in school, last month	-0.00811	1.032	-0.205	-0.322+	0.965	-0.290	0.725	-1.485*	1.711+	-1.416*
clustered p-value	(0.987)	(0.229)	(0.678)	(0.053)	(0.264)	(0.100)	(0.252)	(0.040)	(0.064)	(0.049)
wild cluster p-value	(0.954)	(0.263)	(0.631)	(0.060)	(0.303)	(0.136)	(0.232)	(0.040)	(0.108)	(0.068)
Hours work, 12 mo	-9.626	27.54	-6.554	4.259	28.52	3.914	-18.19	13.78	20.78	14.76
clustered p-value	(0.554)	(0.372)	(0.678)	(0.577)	(0.349)	(0.583)	(0.195)	(0.503)	(0.480)	(0.465)
wild cluster p-value	(0.570)	(0.411)	(0.611)	(0.639)	(0.459)	(0.635)	(0.232)	(0.551)	(0.579)	(0.475)
Hours housework, weekly	-0.986	-0.363	-1.272	-0.333	-0.418	-0.268	-2.272+	2.519+	-1.598	2.603*
clustered p-value	(0.369)	(0.848)	(0.233)	(0.368)	(0.822)	(0.454)	(0.083)	(0.053)	(0.461)	(0.041)
wild cluster p-value	(0.331)	(0.818)	(0.271)	(0.423)	(0.802)	(0.419)	(0.084)	(0.092)	(0.543)	(0.048)
Bandwidth 2.5										
Enrolled in school	0.0182	-0.0023	0.0121	-0.0165+	-0.0086	-0.0165	0.0358	-0.0369	0.0147	-0.0350
clustered p-value	(0.484)	(0.949)	(0.622)	(0.099)	(0.810)	(0.106)	(0.283)	(0.477)	(0.719)	(0.501)
wild cluster p-value	(0.587)	(0.946)	(0.643)	(0.112)	(0.782)	(0.140)	(0.323)	(0.946)	(0.727)	(0.555)
Days in school, last month	0.908	0.911	0.747	-0.491+	0.677	-0.494+	2.002*	-2.300*	2.050+	-2.302*
clustered p-value	(0.163)	(0.387)	(0.232)	(0.074)	(0.516)	(0.086)	(0.010)	(0.031)	(0.065)	(0.031)
wild cluster p-value	(0.152)	(0.439)	(0.271)	(0.236)	(0.535)	(0.236)	(0.024)	(0.092)	(0.144)	(0.092)
Hours work, 12 mo	7.524	27.24	5.946	-7.955	21.18	-6.873	-13.25	43.30	6.992	42.33
clustered p-value	(0.736)	(0.519)	(0.789)	(0.429)	(0.606)	(0.427)	(0.444)	(0.123)	(0.834)	(0.124)
wild cluster p-value	(0.810)	(0.587)	(0.822)	(0.615)	(0.619)	(0.535)	(0.455)	(0.128)	(0.870)	(0.128)
Hours housework, weekly	-0.206	0.0849	-0.420	-0.674+	-0.275	-0.668+	-1.322	2.354	-1.059	2.339
clustered p-value	(0.882)	(0.972)	(0.760)	(0.087)	(0.910)	(0.096)	(0.490)	(0.170)	(0.712)	(0.176)
wild cluster p-value	(0.918)	(0.962)	(0.707)	(0.096)	(0.950)	(0.104)	(0.539)	(0.279)	(0.731)	(0.240)

Only treatment effects and interactions presented. Full results available. Errors clustered at municipal level. + 0.1, *0.05, ** 0.01

Multiple Inference Adjustments

Because of the large number of estimations conducted for this study, we may worry that there is a higher likelihood of finding some statistically significant results by random chance even when the null hypothesis of no impact is true. To address this, I apply a multiple inference adjustment that approximates the multiple inference penalty of the free step down method according to how correlated the study outcomes are (McKenzie 2012, Sankoh et al. 1997). In essence, I use an exponential form of the mean correlation among the outcome variables:

$$\begin{aligned} p_{adj} &= 1-(1-p(k))^{g(k)}, \\ g(k) &= M^{1-r(.k)} \end{aligned}$$

with $p(k)$ as the p-value of outcome k, M as the number of outcomes, and $r(.k)$ as the correlation of all outcomes except for k. Applying this approach to the smallest p-value of the my set of RDD regressions for bandwidth 5 (p-value = 0.008), and using $M = 10$ to cover only the RDD specification so for bandwidth 5, and calculated mean correlation of 0.109, I get an adjusted p-value of 0.0606. We can thus infer that under the large number of estimations included in this study, the small number of statistically significant results are within the percentage of results we would expect to find at 95% confidence intervals even if the null hypothesis were always true. Thus, on the conservative side, we can say that we cannot detect any true impacts in this study. However, as we already acknowledge that if there are any impacts, they appear to be sparse and small, the additional conservative correction may not add much additional information.

Discussion

For all of the above estimations, I have provided DD and local linear RDD regression estimates. Difference in differences (DD) depends on the assumption that the treatment and a comparison group would have experienced the same *change* over time in the absence of the intervention, even if the comparison group does not match the treatment group in *levels* prior to the intervention. The change experienced by the control group is then understood to be the counterfactual change for the treatment group. This approach is most convincing when treatment and comparison groups are similar at baseline, and when a history of "parallel trends" can be shown. Regression discontinuity design (RDD) can be used to treatment effects when inclusion into treatment has been decided by some sort of eligibility threshold. RDD requires three conditions: 1) discontinuity of the probability of treatment at the cutoff point along the eligibility criteria (in our case, early treatment if the municipality has an IIMM score of greater than 38.16, and later treatment otherwise), 2) similarity of characteristics (observable and unobservable) between units on either side of the cutoff, and 3) continuity of the outcome of interest across the threshold in the absence of the intervention. Simply put, RDD compares a just-barely-eligible population to the just-barely-ineligble population, generally controlling for the slope of the outcome of interest across the eligbility threshold in the absence of treatment (Edmonds, Mammen, and Miller 2005).

Under my specification, if the parallel trends assumption needed for DD is satisfied for the entire sample within the selected bandwidth and if treatment effects are uniform throughout the sample, DD and RDD would be expected to give the same treatment coefficients, but this is not generally what I find. Thus there are a couple of reasons that DD and RDD might have generated different results in some cases. First, parallel trends may not have been satisfied (we do not have multiple baseline periods available to be able to test this). The RDD does not rely on parallel trends but rather assumes that the counterfactual changes over time would be functions (linear, in my specification) of the running variable (IIMM score), so it allows different slopes along the running variable in the pre- and post- periods, while the DD cannot have this flexibility. This distinction would be important if, for example, areas with lower schooling were already "catching up" or were falling further behind. In this case, differences between DD and RDD would likely reflect bias in the DD. Second, the RDD design ignores any differential treatment effects along the running variable, since the estimate comes only from the intercept value at the cutoff value of the running value. The DD estimates the average treatment effect across all treatment individuals in the sample, which could be lesser or greater depending on whether treated individuals further from the cutoff are more or less responsive to the program than those who were just barely eligible. (This estimate is still unbiased as long as the counterfactual trend without the program was still the same along the running variable). In this case, neither estimate would be biased necessarily, but the two approaches would be measuring effects for two different populations. A third possibility is misspecified functions along the running variable in the RDD estimates. I have used local linear forms, which might not be correct, but my small sample and the nature of the running variable may also introduce problematic noise. Specifically, of the outcome variables, only schooling outcomes factored into the marginalization index, but they were combined in some way with two other measures (poverty levels and housing quality). As a result, there is more variability in the outcomes along the running variable compared to a case in which the study outcomes at baseline were themselves the running variable.[45] The higher variance may reduce the robustness of the RDD estimates for our small sample.

Other potential limitations for this study may be the risk of measurement error over variables with a long recall period and a potential for contamination. There is a small number of clusters in the sample, and in applying a wild cluster bootstrap for standard errors, I find multiple instances where the null would be wrongly rejected at 5 or 10 percent using standard methods. These are common challenges for program evaluations given the costs and foresight required for implementation.

There appear to be some interesting trends in the results, despite the fact that the program offered small transfers compared to other similar programs, with the major caveat that the number of statistically significant findings falls within the range that we might expect by random fluctation even if there were no impacts at all. There is some suggestive evidence that the program may generate a small decrease in productive labor for wealthier households relative to poorer ones, which would be consistent with either decreasing returns to consumption or greater liquidity constraints among poorer households. There may also be reallocation of productive labor away from women toward men, which we would predict if the program induced women to spend more time on program compliance. While the results are far from conclusive, given that the increase for men may be more than the decrease for women, two possibilities that are worth exploring are either that the transfer has increased women's bargaining power in a way that induces men to work harder, or perhaps the cash has been put toward inputs in ways that made labor more productive. For children, I estimate small gains to attendance for girls on the intensive margin and potentially a small reduction of housework relative to boys, but in general, it appears that the program had little effect on children ages 6 through 12, which might not be unexpected where school participation rates were already high before the program initiated. Longer term studies of conditional programs might benefit from attention to whether and how program benefits and requirements alter long-term labor patterns, focusing on intrahousehold bargaining as well as the implications of the ways household members may specialize to comply with conditions.

[45] Averages among sampled households may also deviate from the municipal averages with limited observations per municipality.

Acknowledgements

I would like to thank Alan de Brauw at IFPRI as well as my professors and classmates at University of California, Berkeley's department of Agricultural and Resource Economics for many helpful comments. I am grateful for financial support of the National Science Foundation Graduate Fellowship. All errors remain my own.

References

Attanasio, Orazio, and Alice Mesnard, 2006. "The Impact of a Conditional Cash Transfer Programme on Consumption in Colombia." *Fiscal Studies* 27 (4): 421-42.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*, MIT Press, 119(1): 249-275.

Cameron, Colin, Jonah Gelbach and Douglas Miller, 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 90, August: 414-427.

de Brauw, Alan and Gilligan, Daniel, 2011. "Using the regression discontinuity design with implicit partitions: The impacts of comunidades solidarias rurales on schooling in El Salvador," IFPRI discussion papers 1116.

de Brauw, Alan, and John Hoddinott, 2008. "Must Conditional Cash Transfer Programs Be Conditioned to Be Effective? The Impact of Conditioning Transfers on School Enrollment in Mexico." Discussion Paper.

de Janvry, Alain, Federico Finan, Elisabeth Sadoulet, and Renos Vakis, 2006. "Can conditional cash transfer programs serve as safety nets in keeping children at school and from working when exposed to shocks?" *Journal of Development Economics*, 79(2), April: 349-373.

Edmonds, Eric, and Norbert Schady, 2008. "Poverty Alleviation and Child Labor." Policy Research Working Paper 4702, World Bank, Washington, DC.

Escobar Latapí, Agustín and Mercedes González de la Rocha, 2008. "Girls, Mothers, and Poverty Reduction in Mexico: Evaluating Progresa-Oportunidades," Chapter 10 in S. Razavi (ed.) The Gendered Impacts of Liberalization: Towards Embedded Liberalism? Chapter 10: 435-468. Routledge/UNRISD, New York.

Filmer and Pritchett, 2001."Estimating Wealth Effects without Expenditure Data - or Tears: An Application to Educational Enrollment in States of India." *Demography* 38(1), February: 115-132.

Fiszbein and Schady, 2009. "Conditional cash transfers: reducing present and future poverty," World Bank Policy Research Report, World Bank, Washington DC.

Gammage, Sarah, 2010. "Time Pressed and Time Poor: Unpaid Household Work in Guatemala" *Feminist Economics*, 16 (3): 79-112, available at: https://ideas.repec.org/a/taf/femeco/v16y2010i3p79-112.html

Gertler, Paul, 2004. “Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA”s Control Randomized Experiment.” *American Economic Review* 94 (2): 336-41.

Gertler, Paul, Sebastian Martinez, and Marta Rubio-Codina, 2006. “Investing Cash Transfers to Raise Long-Term Living Standards,” World Bank Policy Research Working Paper 3994 (August). Available at: http://www1.worldbank.org/prem/poverty/ie/dime_papers/1082.pdf

Glewwe, Paul, and Pedro Olinto, 2004. “Evaluating the Impact of Conditional Cash Transfers on Schooling: An Experimental Analysis of Honduras.” Unpublished manuscript, University of Minnesota, Minneapolis.

Hochschild, Arlie, 1989/2003. The Second Shift. Avon Books.

Hoddinott, John, and Emmanuel Skoufias, 2004. “The Impact of PROGRESA on Food Consumption.” *Economic Development and Cultural Change* 53 (1): 37-61.

Maluccio, John, and Rafael Flores, 2005. “Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan Red de Proteccion Social.” Research Report 141, International Food Policy Research Institute, Washington, DC.

McKenzie, David, 2012. “Tools of the Trade: A quick adjustment for multiple hypothesis testing,” Development Impact Blog post. The World Bank, October 21, 2012.

Moffitt, Robert, 2002. “Welfare Programs and Labor Supply.” In Handbook of Public Economics, ed. Alan J. Auerbach and Martin Feldstein, 2393-2430. Amsterdam, The Netherlands: Elsevier.

Parker and Skoufias, 2000. “Impact of Progresa on Work, Leisure, and Time Allocation,” IFPRI Report, October.

Rawlings and Rubio, 2005. ”Evaluating the impact of conditional cash transfer programs,” World Bank Research Observer, 20: 29-55.

Sankoh, Huque, and Dubey, 1997. “Some comments on frequently used multiple endpoint adjustment methods in clinical trials,” Journal: Statistics in Medicine, 16(22): 2529-2542.

Sahn and Alderman, 1996. ”The Effect of Food Subsidies on Labor Supply in Sri Lanka,” *Economic Development and Cultural Change*, 45(1) (Oct., 1996), pp. 125-145.

Schady, Norbert, and Maria Caridad Araujo, 2008. “Cash Transfers, Conditions, and School Enrollment in Ecuador.” *Economia* 8 (2): 43-70.

Skoufias and di Maro, 2008. ”Conditional Cash Transfers, Adult Work Incentives, and Poverty.” *Journal of Development Studies*, 44:7, 935-960.

Skoufias and Parker, 2002. ”Labor market shocks and their impacts on work and schooling,” FCND discussion papers 129, IFPRI.

Yap, Yoon-Tien, Guilherme Sedlacek, and Peter Orazem, 2008. “Limiting Child Labor through Behavior-Based Income Transfers: An Experimental Evaluation of the PETI Program in Rural Brazil.” In Child Labor and Education in Latin America: An Economic Perspective, ed. Orazem, Sedlacek, and Tzannatos. Palgrave.