IDB WORKING PAPER SERIES Nº IDB-WP-01109

All They're Cracked Up to Be?
The Impact of Chicken Transfers in Guatemala

Conner Mullally
Mayra Rivas
Travis McArthur

Inter-American Development Bank
Multilateral Investment Fund - IDB Lab

October 2019

All They're Cracked Up to Be?
The Impact of Chicken Transfers in Guatemala

Conner Mullally*
Mayra Rivas
Travis McArthur

* University of Florida

Inter-American Development Bank
Multilateral Investment Fund - IDB Lab

October 2019

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
Mullally, Conner.
All they're cracked up to be?: the impact of chicken transfers in Guatemala / Conner Mullally, Mayra Rivas, Travis McArthur.
p. cm. — (IDB Working Paper Series ; 1109)
Includes bibliographic references.
1. Chickens-Guatemala. 2. Livestock-Guatemala. 3. Nutrition-Guatemala. 3. Poor-Nutrition-Guatemala. I. Rivas, Mayra. II. McArthur, Travis. III. IDB Lab. IV. Title. V. Series.
IDB-WP-1109

http://www.iadb.org

All They're Cracked Up to Be? The Impact of Chicken Transfers in Guatemala*

Conner Mullally, Mayra Rivas, and Travis McArthur

University of Florida

October 21, 2019

Abstract

We evaluate a program in Guatemala offering training and transfers of a local chicken variety using a randomized phase-in design with imperfect compliance. We do not find strong evidence for or against positive intent-to-treat effects on household-level outcomes, including indicators of expenditure, calorie and protein intake, diet quality, egg consumption and production, as well as chicken ownership and management. Among girls between the ages of six and 60 months, we find that the program reduced stunting by 23.5 ($\pm$ 19.4) percentage points, while also improving other height and weight outcomes. Boys are more likely to suffer from intestinal illness, which could explain differences in program impacts by sex. Children in the poorest households experienced the largest impacts on dietary diversity and the probability of consuming animal-source foods, but these impacts did not translate into larger effects on height or weight.

Keywords: Livestock, stunting, nutrition, impact evaluation, Latin America, Guatemala, machine learning, randomized trial

JEL codes: O22, O12, Q12

*Corresponding author: connerm@ufl.edu. We would like to thank Yuri Soares, Lorena Mejicanos, Patricia Yanez Pagans, discussants at the 2018 Annual Meeting of the Agricultural and Applied Economics Association, and seminar participants at the University of Minnesota Department of Applied Economics for helpful comments on an earlier draft. We would also like to thank Khanti Consulting for data collection and the Mancomunidad Copan Ch'orti' for their support. This paper was made possible through a consultancy for the Multilateral Investment Fund of the Inter-American Development Bank.

1 Introduction

Transferring chickens to poor households has attracted attention in the media (CNN, 2016) and in development policy circles (Gates, 2016) as a promising antipoverty tool. The enthusiasm stems partly from the fact that a modest chicken flock can generate income while requiring minimal investments of time, money, or land, making small-scale poultry farming an appropriate enterprise for households with few resources (Sonaiya and Swan, 2004). In addition, chicken transfers could improve human capital outcomes if they increase consumption of animal-source foods among young children. Multi-country analysis from Headey, Hirvonen, and Hoddinott (2018) reveals a strong correlation between animal-source food consumption and child health, where the latter is measured by the incidence of stunting. Poor child health can decrease earnings in adulthood, primarily through impaired cognitive development (Attanasio et al., 2018; Attanasio, Meghir, and Nix, 2019; Figlio et al., 2014; Hoddinott et al., 2008). Egg consumption may be particularly effective in improving child health, as eggs have a better digestability-corrected amino acid profile than meat, fish, or soy while also being rich in micronutrients and fatty acids that are essential for brain development (Jin and Iannotti, 2014). The potential of eggs to improve child health is supported by evidence from a randomized feeding trial in Ecuador, where giving infants one egg per day for six months reduced stunting and underweight by 47% and 74%, respectively (Iannotti et al., 2017a). But caution is warranted, as increased poultry and livestock ownership can undermine child health gains by increasing exposure to pathogens contained in animal waste (Headey and Hirvonen, 2016; Headey et al., 2017). Furthermore, chicken transfer programs could be quickly undone by unexpected animal health shocks, as chickens are highly susceptible to disease.

In this article, we add to the evidence on chicken transfers by evaluating a portion of the Recovery of Natural Capital of the Dry Corridor Region program (hereafter "the program") in Guatemala. The program component studied here offered participating households a "chicken set" in exchange for completing a poultry extension program and meeting other program requirements. The chicken set included ten females and two males of the local "naked-neck" variety, a fifty pound bag of commercial chicken feed, forage plants for use once the feed supply was exhausted, and animal health services.

Using baseline and follow-up data collected from 791 households, we estimate the effects of the program by comparing 14 clusters of two or more communities randomly assigned to be early recipients of the program (i.e. the treatment group) to 14 clusters of two or more communities randomly assigned to be late recipients of the program (the control group). We estimate intent-to-treat effects on two sets of outcomes: household-level indicators of expenditure, diet quality, nutrient intake, egg production, chicken ownership, and poultry management, and individual-level indicators of bodyweight, height, diet, exposure to animal waste, and intestinal illness among boys and girls between the ages of six and 60 months. Baseline data were collected about three months before the start of chicken set distribution in the treatment group while follow-up data were collected shortly after the beginning of chicken set distribution in the control group. Participation was not randomized or compulsory within communities. At follow-up, 36% of households in the treatment group sample had received a chicken set at least 80 days prior, which is the amount of time meeded for program chickens to begin producing eggs, versus 9% of households in the control group sample.

We find no statistically significant results at the household level, but the 95% confidence intervals around estimated intent-to-treat impacts are too wide to be characterized as precise null effects. In addition, the slow pace of program implementation in communities assigned to the treatment group suggests that impacts on some indicators may have dissipated before follow-up data were collected. We conclude that we do not have enough evidence to state whether or not the program affected household-level indicators. In contrast, the program had large positive impacts on anthropometric indicators for girls. Assignment to the treatment group raised the average weight-for-age and height-for-age Z-scores by 0.349 standard deviations ($\pm$ 0.281 standard deviations according to the 95% confidence interval) and 0.539 standard deviations ($\pm$ 0.433), respectively. Stunting among girls fell by 23.5 percentage points ($\pm$ 19.4), an improvement of 57% relative to the control group. Severe stunting among girls fell by 14.3 percentage points ($\pm$ 13.1), a decrease of 18% relative to the control group. Impacts on girls are robust to multiple hypothesis testing adjustments, mode of inference, regression specification, and sample construction. Average impacts on anthropometric indicators for boys are positive but small and imprecisely estimated. We speculate that the higher incidence of intestinal illness among boys could explain differences in impacts on height and weight by sex. Relative to girls between the ages of six and 60 months, boys are 23% more likely to have had intestinal illness in the 30 days prior to follow-up interviews and have had intestinal illness for almost a full extra day over the same time horizon. Importantly, the program does not appear to have increased intestinal illness or exposure to animal waste. Estimated average impacts on dietary indicators are small and imprecise for girls and boys.

To explore impacts that average intent-to-treat effects may miss, we estimate heterogeneous intent-to-treat effects using an algorithm developed by Chernozhukov et al. (2018b). The method of Chernozhukov et al. (2018b) uses machine learning methods to estimate intent-to-treat effects on individual observations, allowing us to compare the households and children who were most and least affected by the program both in terms of their respective program impacts and their characteristics. We find no evidence of impact heterogeneity for household-level outcomes. Among children, we find that children from the poorest households enjoyed the largest impacts on the probability of consuming animal-source foods in the past day. We explore why the poorest children did not also enjoy the biggest positive effects on height and weight and present descriptive evidence for differences in hygiene and the incidence of intestinal illness by wealth as offering plausible explanations.

Our study adds to a growing literature on assessing livestock transfer programs. One large strand of this literature studies the effects of "graduation" programs offering some combination of entrepreneurial training, life skills coaching, regular home visits, a stipend, health services, and assets (usually livestock) to impoverished women. Examples include Misha et al. (2019); Raza, de Poel, and Ourti (2018); Banerjee et al. (2015); Roy et al. (2015); Emran, Robano, and Smith (2014); Bandiera et al. (2013); and Krishna, Poghosyan, and Das (2012), among others. The vast majority of graduation evaluations focus on variants of a single intervention—BRAC's "Targeting the Ultra Poor" program in Bangladesh—and tend to find large positive effects on income, assets, and food security. Among existing randomized control trial evaluations of graduation programs, only Raza, de Poel, and Ourti (2018) evaluate impacts on child height and weight, finding that Targeting the Ultra Poor improves weight-for-height but not height-for-age. But given the bundle of interventions and variety of asset types included in Targeting the Ultra Poor, it is difficult to say what portion of its child health impacts can be attributed to livestock.

A smaller strand of the livestock transfer literature evaluates programs offering training and livestock transfers with a "pass-on-the-gift" component, i.e. program participants must agree to pass offspring from transferred livestock on to other households in order to receive program benefits. Nearly all studies of pass-on-the-gift evaluate programs implemented by Heifer International. The lone exception to this rule is Glass et al.'s (2017) evaluation of a program that distributed pigs in Democratic Republic of Congo, where they find positive effects on financial inclusion as well as self-reported mental and physical health. A series of studies on Heifer International's goat program in Nepal find positive effects on financial inclusion and women's empowerment (Janzen et al., 2018) and suggestive evidence for improved dietary diversity among children (Darrouzet-Nardi et al., 2016) as well as child height and weight (Miller et al., 2016, 2014). A pair of evaluations of Heifer's program offering dairy cows, draft cattle, and goats in Zambia find that all three livestock types increase household expenditures while cows and goats increase dietary diversity (Jodlowski et al., 2016; Kafle, Winter-Nelson, and Goldsmith, 2016). Using propensity score matching, Rawlins et al. (2014) find positive effects of dairy cows on child height and weight in Rwanda but no impacts of meat goats on these same outcomes.

We make three main contributions to the literature. First, the context of the program studied here differs substantially from that of existing livestock transfer evaluations, both in terms of the implementing agency and location. Virtually all previous evaluations of livestock transfers focus on programs administered by non-governmental organizations with decades of implementation experience. We evaluate a program managed by a government agency that had never previously attempted a large-scale livestock transfer program, and the implementer's lack of experience was severely tested by a disease outbreak at a program breeding facility. In addition, the existing evidence on livestock transfers comes almost exclusively from South Asia and Sub-Saharan Africa, with the exceptions of Honduras and Peru in Banerjee et al. (2015). Guatemala has a substantially higher GDP per capita and is more urbanized than Bangladesh, Nepal, Zambia, or Rwanda, for example (World Bank, 2019a,c). But Guatemala is characterized by extreme income inequality, with poverty heavily concentrated in rural areas (Guatemala National Institute of Statistics, 2015). Among the countries studied in the literature reviewed above, only Zambia and Honduras have larger Gini coefficients (World Bank, 2019b).

Second, we use random variation in program implementation to identify impacts on child height and weight in an environment where chronic malnutrition is rampant in rural areas. At 46.5%, Guatemala has one of the highest rates of stunting in the world among children under five years of age (World Food Program, 2018). By comparison, stunting rates in Bangladesh, Nepal, and Rwanda (the three countries for which we have evidence on livestock transfers and child height) for children under five are 36%, 38%, and 36%, respectively (National Institute of Statistics of Rwanda, 2015; National Institute of Population Research and Training, 2016; Nepal Ministry of Health, 2016). Third, our approach to exploring heterogeneous program effects allows for a richer analysis than what is typical in the livestock transfer literature. In particular, our analysis demonstrates that children in the poorest households fail to translate relatively large impacts on animal-source food consumption into bigger gains in height and weight than their wealthier counterparts, suggesting that complementary interventions may be needed to improve program impacts for those who need them most.

In what follows, section 2 describes the context of the study and the details of the program itself. Section 3 describes the evaluation design, data collection, and presents summary statistics and balance tests. Section 4 describes our estimation and inference methods, while section 5 reports the results. We discuss our results in section 6 and conclude in section 7.

2 Background

The Recovery of Natural Capital of the Dry Corridor Region program was supported by the Multilateral Investment Fund of the Inter-American Development Bank and implemented by the Mancomunidad Copan Ch'orti'. The latter is an association formed by four municipalities in the Guatemalan department of Chiquimula to facilitate cooperation in regional development projects. Chiquimula is located in the "Dry Corridor", a region that stretches from the southern tip of Mexico through Central America to Panama. Severe droughts affected the Dry Corridor during the agricultural years ending in 2013, 2014, and 2015, capped by a drought in 2015-2016 caused by an El Niño event characterized as the worst the region had seen in 30 years (El Nuevo Diario, 2016; FAO, 2017). The population in Chiquimula is largely indigenous, reliant on agriculture, and has a poverty rate of 71% (INE, 2015). Chiquimula has some of the highest rates of malnutrition among children and overall food insecurity in Guatemala (Government of Guatemala, 2013). In the 2014-2015 National Survey of Maternal and Child Health, 55.6% of children in Chiquimula below five years of age were stunted and 19.2% were underweight according to standards set by the World Health Organization (2011), while 40.2% of children in Chiquimula between the ages of six and 60 months were anemic (Ministry of Public Health and Social Assistance, 2017).

The main objective of the program studied here was to increase the resilience of households located in the Guatemalan Dry Corridor. A major feature of the program was the creation of farmer field schools in communities to train households in basic grain production, agroforestry, adaptation to climate change, and poultry management. Each topic was covered through weekly three-hour classes over the course of eight weeks. As mentioned in the introduction, a second major program component was the distribution of naked-neck chickens. By breeding and distributing naked-neck chickens, the Mancomunidad sought to build resilience for program participants while conserving a productive asset well-suited to the hot and drought-prone conditions of the Dry Corridor.

Evidence for the productivity of naked-neck chickens in adverse conditions is found in the animal science literature. In a laboratory experiment, Chen et al. (2004) found that normally-feathered chickens produce about 7% more eggs over a year (173 versus 161) than naked-neck chickens when the ambient temperature is 22°C. But at 32°C the additional ventilation provided by the naked-neck's absence of neck feathers results in 9.7% additional eggs produced (124 versus 113), higher average weight per egg (42.8 grams versus 40 grams), and better maintenance of body weight (90% versus 85%). As discussed by Wong et al. (2017), indigenous chicken varieties have co-evolved alongside their environments, and we therefore also expect naked-neck chickens to have advantages over specialized egg layers with respect to disease resistance, the ability to scavenge for food, and avoidance of predators. Laboratory results like those of Chen et al. (2004) will not fully translate to conditions in Chiquimula. But the Mancomunidad estimated that the ten hens given to each program participant receiving chickens could produce around 500 eggs per year, suggesting a potentially large impact on access to animal protein.

Program implementation worked as follows. Program personnel would arrive at a community and conduct a household census, followed by a meeting to publicize the program and describe the farmer field schools and other program details. Communities too small to support their own farmer field schools were combined into clusters with other neighboring communities, and larger communities sometimes had multiple farmer field schools. Households in the cohort studied in this article were allowed to pick which modules they attended and any resident of the community served by a farmer field school (according to the program's census) was eligible to enroll.

Each participating household received a chicken set if they satisfied the following criteria: near-perfect attendance at the local farmer field school's poultry sessions, constructing a hen house using local materials, dedicating a small area to forage plants for their chickens, and signing a pledge to transfer ten female and two male chickens back to the program once their birds began to reproduce and their offspring were old enough to be moved (i.e. pass-on-the-gift). The cost per household of the chicken set, including training and all materials, was approximately $500 by 2018 and paid for by the program.[1] To put this figure in context, estimated average annual total consumption expenditure per capita was $1,113 in 2017 for the households in our data set. While the program's various components are directed at the entire household, the individuals attending the poultry module and receiving chickens are all women. As of January 2018, the poultry component of the program had produced 147,340 birds and distributed chickens to 4,239 households (Multilateral Investment Fund, 2018).

At each farmer field school, the chickens received by the first group of households to complete the poultry module were raised by program technicians at a dedicated facility. Subsequent graduates of the poultry module received chickens that had been raised by earlier program participants from the same cluster and later passed back to the program to satisfy the above-mentioned pledge. In the present study, fewer than 1% of households assigned to the treatment group had received pass-on-the-gift birds when follow-up data were collected, versus 37% of treatment group households overall. Waiting for further distribution of pass-on-the-gift chickens would have led to contamination of the control group. In addition, only 6% of households in the treatment group had participated in a non-poultry farmer field school module. The Mancomunidad made attendance of all modules compulsory for future program cohorts, but this change did not affect the treatment group used in our study. We therefore interpret our results as reflecting the impact of poultry training and chicken

[1] Cost figures are taken from program documentation provided by the Mancomunidad Copan Ch'orti'.

transfers delivered directly to beneficiary households through the program.

3 Evaluation design and data

All 133 communities located in the most drought-prone areas served by the Mancomunidad were included in the program. Capacity constraints dictated that the program would be rolled out over approximately five years. To that end, communities were divided into five cohorts. The Mancomunidad's initial plan was to phase the program into one cohort per year, but implementation challenges described below resulted in slow rollout for some cohorts and much quicker implementation in others.

3.1 Evaluation design

An initial cohort of 15 communities received the program in 2013-2014. Communities assigned to the first cohort were selected by the Mancomunidad and were prioritized because of the severe effects of a drought. An additional 50 communities in relatively affluent coffee-producing areas were assigned by the Mancomunidad to the program's last cohort. The remaining 68 communities were grouped into clusters large enough to support a farmer field school and randomly assigned to cohort two, three, or four by the research team. Under the evaluation plan agreed to with the Mancomunidad, program technicians would finish distributing chickens raised by the program in one cohort before moving on to another. But the evaluation plan did not set rules on the distribution of chickens between participating households (i.e. through pass-on-the-gift) or on the order and pace of distribution within a given cohort.

The randomization was carried out at the cluster level as follows. At the request of the Mancomunidad, three clusters (totaling five communities) were randomly assigned to cohort three to make the pace of program rollout as close to the Mancomunidad's ideal as possible.

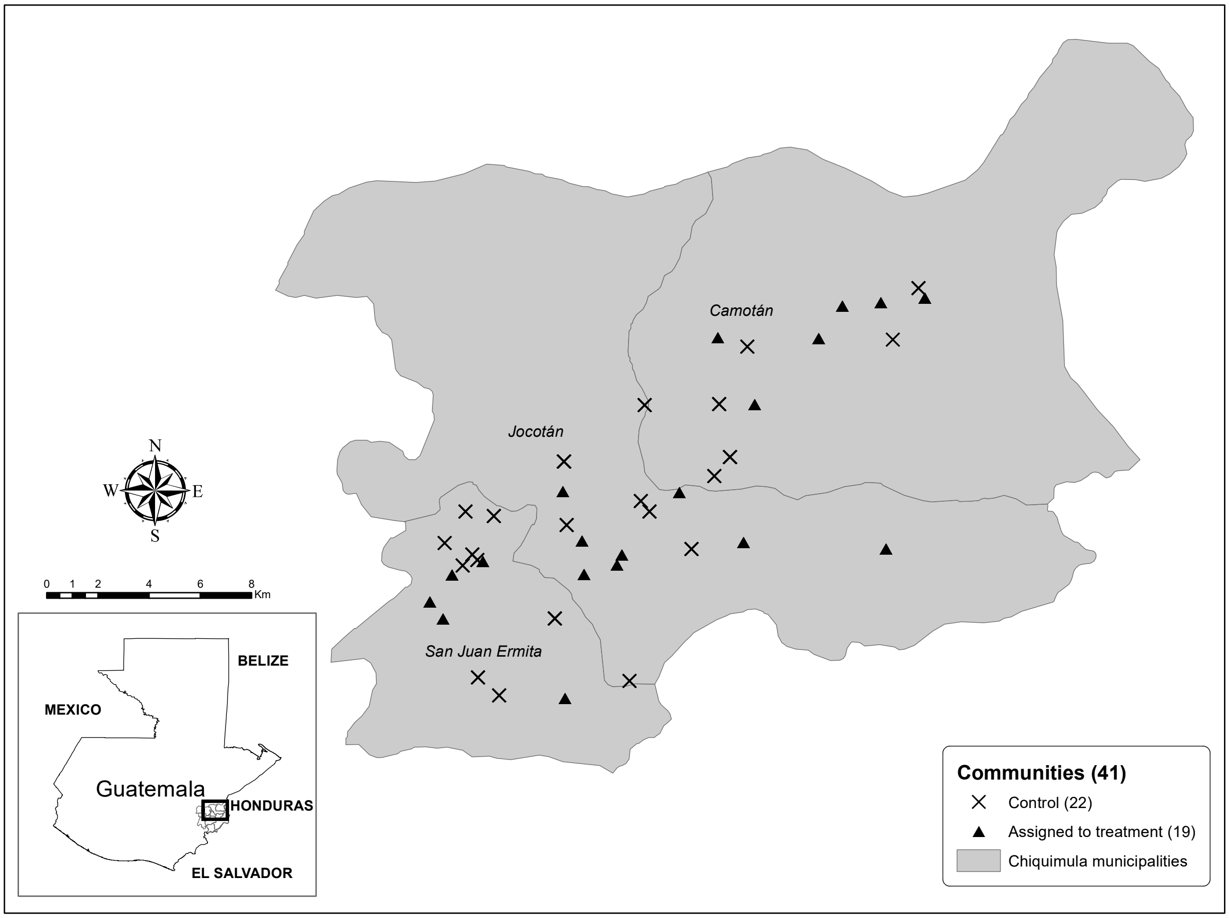
The remaining 48 clusters (with 63 communities) were matched into strata of three clusters each using Euclidean distance as a function of population size, number of communities, altitude (indicators for low, medium, and high), and indicators for being located in one of three watersheds. Community-level data were provided by the Mancomunidad. Each member of a stratum was randomly assigned to program cohort two, three, or four. Cohort four would serve as the control group for the evaluation, with follow-up data being collected before cohort four could plausibly be affected by the program. Communities were not told in advance when exactly they could expect the program to arrive, which should limit bias from anticipatory effects.

There was a budget constraint on the evaluation that limited the number of surveys to be collected, ruling out a midline survey and forcing the research team to decide how to distribute the survey budget between cohorts two, three, and four. The research team assumed that if data were collected before cohort four was affected by the program, then cohort three would only be weakly affected at that point, particularly as compared to cohort two. Therefore survey data were only collected from the 16 strata of clusters assigned to cohorts two and four, including 16 clusters/19 communities assigned to cohort two (i.e. the treatment group) and 16 clusters/22 communities assigned to cohort four (the control group). Approximate locations of communities included in the data set are shown in figure 1.

3.2 Data collection and sample construction

The sample of interviewed households was constructed in three stages. First, since a census was not available or feasible given budget constraints when baseline data were collected, supervisors from the data collection team met with local leaders and created rosters of individuals living in each of the 41 communities assigned to program cohorts two and four. The field team made it as clear as possible to community leaders that no benefit would be gained from appearing on the list, and that the purpose of the list was to gather information that

Figure 1: Program area

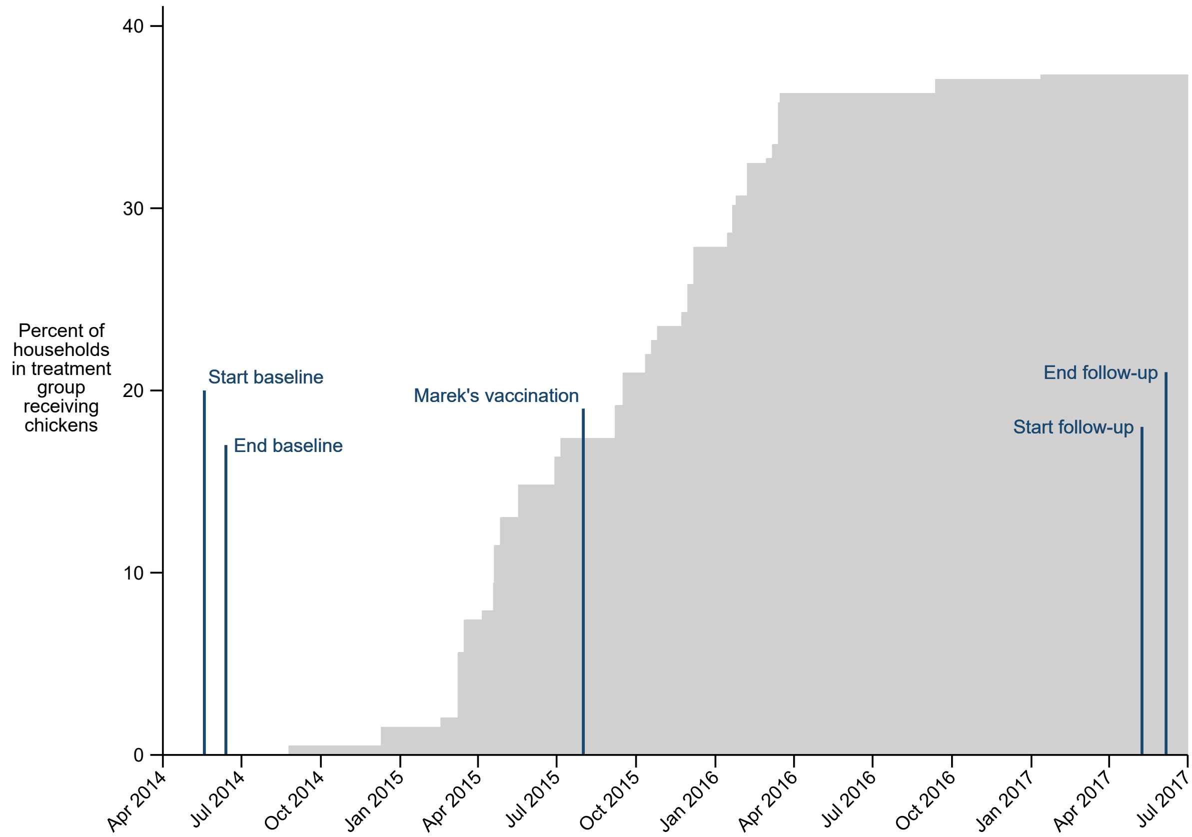


would help the Mancomunidad serve the community. Next, a random sample of households was selected from each cluster. Finally, households were retained in the sample if they could be located and if they indicated participating in at least one productive agricultural activity (raising animals or planting crops). Households not satisfying these critera were replaced in the sample. The purpose of filtering households by participation in agriculture was to identify households with a high probability of future participation in a farmer field school. The baseline survey was carried out in May 2014, or about three months before the start of chicken distribution in the treatment group. Follow-up data were collected in May 2017, or about five months after the completion of chicken distribution in the treatment group and one month into the distribution of chickens in the control group.

Figure 2 shows the timing of baseline and endline data collection as well as the cumulative number of chicken set deliveries made to treatment group households in our data set. Chicken set distribution began in August 2014, accelerated in August 2015, and finished in January 2015. As mentioned in the introduction and noted in figure 2, the program chicken-breeding facility suffered an outbreak of Marek's disease, which is contagious and deadly to chickens. We discuss the causes and consequences of the outbreak as well as other program challenges in appendix A.1. A vaccine against Marek's disease was introduced into the program in August 2015. Although we do not have data on chicken deaths, the outbreak could have affected the 47% of program participants in the treatment group who received their chickens before the vaccine was introduced. When discussing our results later in the paper, we consider the role of the disease outbreak in limiting program impacts.

The decision of when to collect follow-up data was complicated by lack of control on the part of the research team over the timing of program implementation and the fact that rollout accelerated once management issues (e.g. the Marek's outbreak) were resolved. The change in the pace of implementation made the program's start in the control group a moving target. But the follow-up survey seems to have been timed as appropriately as possible given the circumstances. At follow-up, 38% of households in the treatment group had received a chicken set versus 20% of households in the control group. Virtually all participating households in the treatment group had been exposed to the program long enough to experience impacts on most indicators, i.e. at least 80 days, which is how long it would have taken for chickens received through the program to begin laying eggs. In contrast, 9% of households in the control group had received a chicken set at least 80 days before their follow-up interviews. Waiting longer to collect follow-up data would have increased the participation rate in treatment group communities (through pass-on-the-gift) but worsened contamination of the control group. Collecting follow-up data early enough to cut into the 9% of control households with prolonged exposure to the program would have reduced the number of treatment group households with sufficient program exposure. We discuss the

Figure 2: Timing of data collection and chicken transfers



Notes: Households would have needed around 80 days for chickens to begin producing eggs.

details of follow-up data collection and the duration of exposure needed for the program to have an effect on different indicators in appendix A.2.

Enumerators were able to complete follow-up interviews with 92.3% of the baseline sample. We restrict our analysis to panel households and show that our results are robust to attrition in appendix A.10. Prior to analysis, we dropped two contaminated strata where the Mancomunidad intervened in the control clusters but not in the clusters assigned to treatment (this appears to have been done purely by mistake). Assignment to the treatment group was randomized within the remaining strata, so our identification strategy is unaffected by dropping the contaminated observations.[2] Our final sample includes 14 strata, 28 clusters,

[2]In appendix A.7 we compare baseline characteristics for our main estimating sample and the contami-

and 791 households observed over two years, among which 14 clusters with 391 households were assigned to treatment and 14 clusters with 400 households were assigned to control. Power calculations for the experimental design are shown in section A.6 of the appendix.

3.3 Baseline summary statistics and balance

Table 1 shows averages and standard deviations by treatment status as well as estimated differences in means for key demographic indicators and household-level outcomes, all estimated using baseline data. Although baseline chicken ownership is fairly high, egg production is quite low and heavily skewed, as the median level of egg production among chicken owners is zero over the six months preceding baseline interviews. Egg consumption also looks somewhat high, but the average of roughly 1.35 egg consumed daily per adult male equivalent is misleading, as the median is 0.65 and 31% of households reported consuming zero eggs.

In general, differences in means are small relative to the spread of the data, with calorie consumption and naked-neck chickens showing statistically significant differences. It is unsurprising that we would find at least some statistically significant differences given the large number of hypothesis tests shown in table 1 and the fact that we did not directly stratify on baseline outcomes in our randomization. We control for lagged outcomes when estimating program impacts and show robustness to controlling for imbalanced baseline covariates in appendix A.15.

Table 2 shows baseline summary statistics and balance for age and anthropometric indicators among girls and boys, respectively, ages six to 60 months. We focus on this age range for two reason. First, it is critical for child growth and development that children complement breastfeeding with nutrient-dense solid foods (e.g. eggs and chicken meat) between six and 23 months of age (Choudhury, Headey, and Masters, 2019). Second, children

nated strata that were dropped, and in general they look very similar. In addition, we present results using the full sample in appendix A.17. None of our conclusions is sensitive to whether we include all strata in our analysis.

Table 1: Baseline summary statistics and balance: household-level variables

	Treatment	Control	Difference
Household size (adult male equivalents)	4.069	3.976	0.055
	{1.799}	{1.802}	[0.136]
Dependency ratio	0.390	0.395	-0.003
	{0.225}	{0.223}	[0.038]
Woman-headed household (0/1)	0.161	0.188	-0.029
	{0.368}	{0.391}	[0.040]
Wealth (log)	9.619	9.412	0.220
	{1.396}	{1.353}	[0.294]
Annual food expenditure per adult male equivalent (log)	7.250	7.432	-0.190
	{1.269}	{1.131}	[0.193]
Daily calories per adult male equivalent (log)	8.340	8.463	-0.137
	{0.730}	{0.734}	[0.058]**
Daily grams of animal protein per adult male equivalent	7.702	7.883	-0.445
	{12.590}	{11.304}	[2.076]
Daily servings of eggs	0.393	0.440	-0.069
	{0.570}	{0.678}	[0.094]
Eggs consumed per day per adult male equivalent	1.307	1.408	-0.137
	{2.063}	{2.354}	[0.291]
Food consumption score	20.376	21.345	-1.110
	{8.419}	{8.239}	[1.601]
Chickens owned	11.263	10.355	1.363
	{11.867}	{10.670}	[1.212]
Naked-neck chickens owned	1.402	1.210	0.346
	{3.305}	{2.863}	[0.143]**
Uses poultry registry (0/1)	0.107	0.095	0.012
	{0.310}	{0.294}	[0.024]
Eggs produced in last six months (log)	1.421	1.527	-0.084
	{2.299}	{2.504}	[0.219]
Sold at least one egg in last six months (0/1)	0.038	0.037	0.009
	{0.192}	{0.190}	[0.019]
Observations	391	400	791

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard deviations in curly braces, standard errors in brackets. All regressions for differences in means include the treatment indicator, an intercept, and indicators for thirteen strata. Standard errors and degrees of freedom were estimated as in Young (2016). Continuous outcomes were top coded their 1$^{\text{st}}$ and 99$^{\text{th}}$ percentiles. Wealth includes the value of land, livestock, agricultural implements, housing, consumer durables, and savings. Food quantity data were collected using a consumption module similar to that of the Guatemalan National Survey of Living Standards, modified for the study context. Calorie and protein data were obtained using a food composition table for Central America (INCAP, 2012). The food consumption score is a quality-weighted measure of dietary diversity (World Food Program, 2008).

who were 23 months of age when baseline data were collected would be no more than sixty months of age at follow-up. In other words, children sixty months of age and younger at follow-up would have received their chickens sets during the most critical period for child development. Our indicators include Z-scores for weight-for-age and height-for-age as well as indicators for not being underweight, severely underweight, stunted, or severealy stunted, all calculated according to guidelines from the World Health Organization (2011). Impacts on Z-scores would indicate that the program shifted the distribution of height or weight, while changes in indicators for stunting and underweight would imply that children with the poorest health status were affected by the program.

Table 2: Baseline summary statistics and balance: children ages six to 60 months

	Treatment	Control	Difference
Girls			
Age in months	32.604	31.655	1.298
	{17.057}	{16.194}	[2.458]
Height-for-age (Z-score)	-1.474	-1.724	0.233
	{1.977}	{1.739}	[0.248]
Not stunted (0/1)	0.516	0.462	0.054
	{0.502}	{0.501}	[0.091]
Not severely stunted (0/1)	0.802	0.790	0.007
	{0.401}	{0.409}	[0.060]
Weight-for-age (Z-score)	-0.940	-0.944	-0.035
	{1.305}	{1.391}	[0.149]
Not underweight (0/1)	0.813	0.773	0.021
	{0.392}	{0.421}	[0.054]
Not severely underweight (0/1)	0.956	0.941	0.024
	{0.206}	{0.236}	[0.036]
Observations	91	119	210
Boys			
Age in months	32.210	30.694	1.082
	{15.016}	{16.049}	[1.806]
Height-for-age (Z-score)	-1.766	-1.762	-0.055
	{1.794}	{1.891}	[0.439]
Not stunted (0/1)	0.572	0.478	0.103
	{0.497}	{0.501}	[0.094]
Not severely stunted (0/1)	0.826	0.776	0.045
	{0.380}	{0.418}	[0.079]
Weight-for-age (Z-score)	-0.911	-0.995	0.059
	{1.513}	{1.395}	[0.306]
Not underweight (0/1)	0.790	0.791	-0.016
	{0.409}	{0.408}	[0.045]
Not severely underweight (0/1)	0.949	0.948	0.006
	{0.220}	{0.223}	[0.036]
Observations	138	134	272

Notes: * p <0.10, ** p <0.05, *** p <0.01. Standard deviations in curly braces, standard errors in brackets. All regressions for differences in means include the treatment indicator, an intercept, and indicators for thirteen strata. Z-scores as well as indicators for stunting and underweight calculated according to World Health Organization standards (World Health Organization, 2011). Weight-for-age and height-for-age were top and bottom coded at 6 and -6, respectively, following guidelines set by the World Health Organization (2011). Standard errors and degrees of freedom estimated as in Young (2016).

4 Empirical approach

In our empirical analysis, we focus on identification and estimation of intent-to-treat effects, i.e. the average effect of assignment to the treatment group rather than the control group. We opt not to estimate effects of the treatment itself in the main text because with imperfect treatment compliance the randomization only identifies program treatment effects in the absence of spillovers (as well as other assumptions laid out in Imbens and Angrist (1994)). For households in our data set, average distance to the nearest neighbor is 74 meters. Ruling out any exchange of eggs (for example) wihin communities seems like an excessively strong assumption in the present case.

4.1 Estimation and inference for average intent-to-treat effects

Household-level intent-to-treat effects are estimated using the following regression:

$$y_{hcst} = \gamma_s + \rho y_{hcst-1} + \delta Treat_c + \varepsilon_{hcst} \tag{1}$$

where h, c, s, and t index household, cluster, stratum, and time period, respectively. The γ_s parameter is a stratum fixed effect, y_{hcst-1} is the lagged outcome, $Treat_c$ is a dummy variable equal to one for households in clusters assigned to the treatment group, and δ is the average intent-to-treat effect.[3]

We use a slightly different specification when estimating impacts on height and weight indicators among children because lagged outcomes are only observed for individuals who were already born at baseline. We compensate by using lagged outcomes among siblings, when available:

$$y_{ihcst} = \gamma_s + \gamma_{y_i} + \gamma_{\bar{y}} + \rho_1 y_{ihcst-1} + \rho_2 \bar{y}_{hcst-1} + \beta Age_{ihcst} + \delta Treat_c + \varepsilon_{ihcst} \tag{2}$$

[3]We used lagged food expenditure for y_{icst-1} when estimating impacts on total expenditure because the latter was only measured at follow-up.

where i indexes child, $y_{ihcst-1}$ is the lagged outcome (set to zero for individuals not yet born at baseline), $\bar{y}_{hcst-1}$ is the average lagged outcome among siblings of the same sex as child i at baseline (e.g. the lagged average height-for-weight Z-score among girl siblings if child i is a girl, set to zero if no such siblings exist), γ_{y_i} is the coefficient on a dummy variable equal to one if $y_{ihcst-1}$ is observed, $\gamma_{\bar{y}}$ is the coefficient on a dummy variable equal to one if $\bar{y}_{hcst-1}$ is observed in the data, and Age_{ihcst} is age in months as measured at follow-up.

Since individual-level diet indicators were not collected at baseline, we remove γ_{y_i}, $\gamma_{\bar{y}}$, $y_{ihcst-1}$, and $\bar{y}_{hcst-1}$ from the model when estimating impacts on dietary diversity and consumption of animal-source foods among children and replace them with the lagged outcome measured at the household level using seven-day recall. For inference, we estimate cluster-robust standard errors and effective degrees of freedom as in Young (2016), which should result in valid inference despite having just 28 clusters in the data. We adjust for multiple comparisons by reporting q-values and 95% confidence intervals adjusted for the false discovery rate (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2005). We describe our rules for multiple hypothesis testing in detail in appendix A.5.

4.2 Estimation and inference for heterogeneous intent-to-treat effects

We use a method proposed by Chernozhukov et al. (2018b) to explore program impact heterogeneity. The method of Chernozhukov et al. (2018b) allows us to model heterogeneous treatment effects as a function of a large number of observed characteristics while avoiding "overfitting", i.e. without obtaining an excellent in-sample fit at the expense of highly-variable out-of-sample performance. Overfitting could lead to "discovering" treatment effect heterogeneity that is a quirk of a given sample rather than reflective of the population of interest. In addition, when applying the method of Chernozhukov et al. (2018b) we can predict heterogeneous treatment effects using machine learning methods without placing any distributional assumptions on the estimates generated by a given algorithm. In general, the theory needed to justify hypothesis testing using machine learning estimates is not well developed, although there are exceptions (Athey, Tibshirani, and Wager, 2019; Belloni, Chernozhukov, and Hansen, 2014). In contrast, hypothesis testing is carried out just as in any application of linear regression when applying the method of Chernozhukov et al. (2018b). We present key details of our approach to estimating impact heterogeneity below and reserve a detailed discussion for appendix A.3.

To apply the method of Chernozhukov et al. (2018b), we begin by estimating the following equation:

$$y_{ics} = \boldsymbol{\alpha}'\boldsymbol{X}_{ics} + \beta_1(Treat_c - p(\boldsymbol{Z}_{ics})) + \beta_2(Treat_c - p(\boldsymbol{Z}_{ics}))(S(\boldsymbol{Z}_{ics}) - \bar{S}) + \varepsilon_{ics} \tag{3}$$

The vector $\boldsymbol{Z}_{ics}$ includes covariates that could explain impact heterogeneity and are observed at baseline or otherwise unaffected by assignment to treatment. In appendix A.4 we describe and justify each variable included in $\boldsymbol{Z}_{ics}$. The propensity score is given by $p(\boldsymbol{Z}_{ics})$ and is set to 0.5 for all observations by virtue of randomization,[4] $S(\boldsymbol{Z}_{ics})$ is a "proxy predictor" for the conditional intent-to-treat effect obtained through machine learning methods, and $\bar{S}$ is the mean of the proxy predictor. The average intent-to-treat effect is given by β_1, while $\beta_1 + \beta_2(S(\boldsymbol{Z}_{ics}) - \bar{S})$ is a linear approximation to the intent-to-treat effect conditional on $\boldsymbol{Z}_{ics}$. The properties of ordinary least squares ensure that $\beta_1 + \beta_2(S(\boldsymbol{Z}_{ics}) - \bar{S})$ is the "best linear predictor" of the conditional intent-to-treat effect given $S(\boldsymbol{Z}_{ics})$, i.e. the linear predictor that yields the smallest mean squared error (Chernozhukov et al., 2018b). The vector $\boldsymbol{X}_{ics}$ contains a column of ones and additional terms meant to improve precision, including stratum indicators, $S(\boldsymbol{Z}_{ics})$, and a proxy predictor for the conditional mean of y_{ics} when assigned to the control group.

[4]In our application, subtracting the propensity score from the treatment indicator has no effect on our results since the propensity score does not vary. However, for research designs where the probability of treatment depends on observed characteristics, failing to difference out the propensity score can result in substantial bias (Chernozhukov et al., 2018a).

We also estimate equation 4:

$$\begin{aligned} y_{ics} = & \boldsymbol{\alpha}'\boldsymbol{X}_{ics} + \gamma_1 G_1(Treat_c - p(\boldsymbol{Z}_{ics})) + \gamma_2 G_2(Treat_c - p(\boldsymbol{Z}_{ics})) \\ & + \gamma_3 G_3(Treat_c - p(\boldsymbol{Z}_{ics})) + u_{ics} \end{aligned} \tag{4}$$

where G_1 is an indicator variable equal to one for observations in the lowest tercile of $S(\boldsymbol{Z}_{ics})$, and G_2 and G_3 are indicators for the middle and top terciles of $S(\boldsymbol{Z}_{ics})$, respectively. If $S(\boldsymbol{Z}_{ics})$ closely approximates the true conditional intent-to-treat effect, then γ_1, γ_2, and γ_3 will closely correspond to the average intent-to-treat effects among households or children least affected, moderately affected, and most affected by the program. We refer to γ_1, γ_2, and γ_3 as group average intent-to-treat effects.

Finally, we have:

$$Z_{ics}^k = \delta_1 G_1 + \delta_2 G_2 + \delta_3 G_3 + e_{ics} \tag{5}$$

where Z_{ics}^k represents a single covariate from $\boldsymbol{Z}_{ics}$. The parameters δ_1, δ_2, and δ_3 are the means of Z_{ics}^k for the least, middle, and most affected observations. By testing whether $\delta_1 = \delta_3$ we can check whether the most and least affected observations differ in their characteristics. We refer to the comparison of δ_1 and δ_3 as classification analysis, following Chernozhukov et al. (2018b). We limit the classification analysis presented in the main text to outcomes that exhibit evidence of treatment effect heterogeneity, i.e. where we can reject $\beta_2 = 0$ or $\gamma_1 = \gamma_3$ after adjusting for multiple hypothesis testing. Other classification analysis results are given in appendices A.13 for children and A.14 for household-level outcomes.

The method of Chernozhukov et al. (2018b) avoids overfitting by using repeated sample splitting, i.e. randomly dividing the sample into an auxiliary part used for model selection and a main part used for estimation and hypothesis testing. We repeat the sample splitting procedure 199 times and report the median of each point estimate, p-value, and upper and lower confidence interval bound in our final results. Using medians generated by many sample splits increases robustness, e.g. by avoiding "cherry picking" a favorable sample split. The cost of sample splitting is that a nominal significance level of α translates to a true significance level of 2α (Chernozhukov et al., 2018b).

We use two machine learning methods to create the proxy predictors (i.e. the $S(\boldsymbol{Z}_{ics})$): the elastic net (Zou and Hastie, 2005) and the random forest algorithm (Breiman, 2001). The elastic net is a penalized least squares method that arguably combines the best features of two other well-known penalized least squares methods: the LASSO and ridge regression. When predictors are not highly correlated with one another and the correct regression model has a relatively small number of non-zero parameters, then the elastic net will select a small number of predictors, just as in the LASSO. When a subset of predictors are highly correlated with one another, the elastic net will include them all in the final model, just as in ridge regression (Zou and Hastie, 2005). The random forest algorithm is an "ensemble" method that averages predictions generated by many "regression trees". Regression trees are grown by splitting the sample into "leaves" where observations in a leaf have similar values of covariates that are important for predicting the outcome. Each observation is assigned the average of the outcome within its leaf as its predicted value. The elastic net should provide a better fit for outcomes where the conditional mean is linear in parameters whereas random forest may have more success for nonlinear problems. Both methods strong predictive accuracy while avoiding overfitting.

Since we are using two machine learning methods, we could potentially have two tests for every hypothesis of interest that should be accounted for when adjusting for multiple comparisons. We opt to present results from only one machine learning method per outcome. When deciding which set of results to select for a given outcome, we first check whether one of the machine learning methods strictly dominates based on two separate goodness-of-fit measures taken from Chernozhukov et al. (2018b) and described in appendix A.3. If the two goodness-of-fit measures disagree, then we select the set of results to report at random.

Note that we do not conduct separate heterogeneity analyses for boys and girls. Dividing the subsample of children by sex would in our view cut the data too finely, limiting our ability to detect meaningful effect heterogeneity. Instead, we include sex as one of the characteristics that could shape treatment effect heterogeneity. Results obtained when splitting the sample of children by sex are similar but less precise than what we obtain by pooling the sample of children.

5 Results

5.1 Average intent-to-treat effects on households

Table 3 presents average intent-to-treat effects for household-level outcomes. Virtually all intent-to-treat estimates are small and imprecisely estimated. Although the negative point estimates on number of chickens owned and naked-neck chickens owned are somewhat alarming, they are not significant and contradicted by a positive effect on egg production that is estimated with comparable precision. Overall, impacts at the household level are not estimated with enough precision to make a strong statement about program effects on the outcomes shown in table 3.

5.2 Average intent-to-treat effects on children

Table 4 shows estimated intent-to-treat effects among children ages six to 60 months. Among girls, impacts on weight-for-age, height-for-age, stunting, and severe stunting are large, significant, and robust to adjustment for multiple comparisons. Estimated effects for boys are generally small and none are statistically significant. Impacts on consumption of animal-source foods and the dietary diversity score are imprecisely estimated for both sexes. As shown in the rightmost column of table 4, there are significant differences in intent-to-treat effects by sex for weight-for-age and stunting, although none of the differences in intent-to-treat effects by sex remain statistically significant after adjusting for multiple comparisons.

Table 3: Intent-to-treat effects, household-level outcomes

	Intent-to-treat effect	Control mean
Annual expenditure per adult male equivalent (log)	-0.008	9.087
	[-0.434, 0.418]	
	(-0.511, 0.494)	
Annual food expenditure per adult male equivalent (log)	0.004	8.364
	[-0.290, 0.298]	
	(-0.290, 0.298)	
Daily calories per adult male equivalent (log)	-0.001	8.075
	[-0.250, 0.248]	
	(-0.250, 0.248)	
Daily grams of animal protein per adult male equivalent	2.420	26.793
	[-14.317, 19.157]	
	(-18.309, 23.148)	
Daily servings of eggs	-0.024	0.439
	[-0.243, 0.195]	
	(-0.272, 0.224)	
Eggs consumed per day per adult male equivalent	-0.007	1.867
	[-1.009, 0.996]	
	(-1.068, 1.054)	
Food consumption score	-0.897	7.214
	[-2.839, 1.045]	
	(-3.652, 1.858)	
Chickens owned	-1.289	8.377
	[-3.318, 0.740]	
	(-3.802, 1.224)	
Naked-neck chickens owned	-0.354	2.425
	[-1.882, 1.175]	
	(-1.971, 1.264)	
Uses poultry registry (0/1)	0.026	0.015
	[-0.008, 0.060]	
	(-0.022, 0.075)	
Eggs produced in last six months (log)	0.203	2.600
	[-0.503, 0.910]	
	(-0.598, 1.005)	
Sold at least one egg in last six months (0/1)	0.004	0.075
	[-0.050, 0.059]	
	(-0.050, 0.059)	
Observations	791	

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $^{+}$ $q < 0.10$, $^{++}$ $q < 0.05$, $^{+++}$ $q < 0.01$, where q is the false discovery rate. 95% confidence intervals in brackets, 95% false discovery rate-adjusted confidence intervals in parentheses. Standard errors and degrees of freedom estimated as in Young (2016). See equation 1 for the regression specification.

Table 4: Intent-to-treat effects for children, ages 6 to 60 months

	Girls		Boys		
	Intent-to-treat effect	Control mean	Intent-to-treat effect	Control mean	Difference
Weight-for-age (Z-score)	0.349	-1.305	0.027	-1.163	0.322
	[0.111, 0.587]***		[-0.397, 0.450]		[-0.050, 0.694]*
	(0.068, 0.629)$^{++}$		(-0.397, 0.450)		(-0.154, 0.798)
Not underweight (0/1)	0.090	0.773	0.023	0.802	0.068
	[-0.021, 0.202]		[-0.098, 0.143]		[-0.113, 0.249]
	(-0.027, 0.207)		(-0.120, 0.165)		(-0.122, 0.257)
Not severely underweight (0/1)	0.054	0.938	0.032	0.954	0.021
	[-0.013, 0.120]		[-0.022, 0.087]		[-0.081, 0.124]
	(-0.013, 0.120)		(-0.048, 0.112)		(-0.081, 0.124)
Height-for-age (Z-score)	0.539	-2.209	0.065	-1.997	0.474
	[0.203, 0.875]***		[-0.631, 0.762]		[-0.178, 1.125]
	(0.106, 0.972)$^{++}$		(-0.706, 0.836)		(-0.295, 1.242)
Not stunted (0/1)	0.235	0.412	0.038	0.420	0.197
	[0.103, 0.366]***		[-0.121, 0.196]		[-0.029, 0.423]*
	(0.041, 0.428)$^{++}$		(-0.166, 0.242)		(-0.132, 0.526)
Not severely stunted (0/1)	0.143	0.784	0.012	0.794	0.131
	[0.025, 0.261]**		[-0.115, 0.139]		[-0.064, 0.326]
	(0.012, 0.274)$^{++}$		(-0.121, 0.145)		(-0.085, 0.347)
Consumed animal-source foods in past day (0/1)	-0.039	0.598	0.061	0.547	-0.100
	[-0.339, 0.260]		[-0.210, 0.331]		[-0.294, 0.094]
	(-0.394, 0.315)		(-0.259, 0.380)		(-0.329, 0.129)
One-day dietary diversity score	-0.014	5.649	-0.041	5.484	0.027
	[-1.049, 1.021]		[-1.272, 1.190]		[-0.626, 0.680]
	(-1.049, 1.021)		(-1.272, 1.190)		(-0.626, 0.680)
Observations	195		241		436

Notes: * p <0.10, ** p <0.05, *** p <0.01; $^{+}$ q <0.10, $^{++}$ q <0.05, $^{+++}$ q <0.01, where q is the false discovery rate. 95% confidence intervals in brackets, 95% false discovery rate-adjusted confidence intervals in parentheses. Standard errors and degrees of freedom estimated as in Young (2016). Z-scores and indicators for stunting and underweight based on standards from World Health Organization (2011). See equation 2 for the regression specification.

5.3 Robustness checks on average intent-to-treat effects

We subject our results to a large number of robustness checks, including testing hypotheses by randomization inference (appendix A.8), adding imbalanced baseline outcomes to the covariate set (A.15), and checking for bias from spillovers across clusters (A.9) and attrition (A.10). We also run a series of robustness checks specifically for impacts on children, including dropping age and lagged outcomes from the covariate set when estimating average intent-to-treat effects (A.11), including children below six months of age in the sample (A.19), limiting the sample to children alive at baseline (A.21), estimating average intent-to-treat efects using machine learning methods (A.12), and limiting the sample to households that had children at baseline (A.20). None of our robustness checks negates our conclusions.

5.4 Heterogeneous intent-to-treat effects

Our analysis of heterogeneous intent-to-treat effects uncovered no statistically significant heterogeneity at the household level, and we present household-level results in appendix A.14. Table 5 reports the results of estimating equation 3 for our child-level outcomes. The effect on stunting remains statistically significant when pooling boys and girls, although its q-value is above standard cutoff values. The "Heterogeneity" parameter (i.e. β_2 from equation 3) is statistically significant for the one-day dietary diversity score and the indicator for having consumed animal-source foods in the past day.[5] For the remaining outcomes, estimates in the "Heterogeneity" column are quite noisy. Table 6 presents the estimated group average intent-to-treat effects for child-level outcomes. The differences in intent-to-treat effects for the most and least affected children are estimated somewhat noisily, which is to be expected given that each parameter is estimated using one-third of the available data. Given that the

[5]The 90% confidence interval for the dietary diversity score heterogeneity parameter excludes zero despite the fact that the q-value is greater than 0.10 (the q-value is 0.1018). This is possible because the reported p-values (upon which the q-values are based) and the reported confidence intervals are medians generated by sample splitting. The median p-value might come from a different sample split than the median upper or lower bound of the corresponding confidence interval.

relevant "Heterogeneity" parameters are significant after adjusting for multiple comparisons, we conclude that there is evidence of impact heterogeneity for both diet outcomes.

Table 5: Best linear predictor of the average intent-to-treat effect and impact heterogeneity: child outcomes

	Intent-to-treat	Heterogeneity	ML method
Weight-for-age (Z-score)	0.157	0.570	Elastic net
	[-0.063, 0.390]	[-0.501, 1.649]	
	(-0.090, 0.432)	(-0.830, 1.981)	
Not underweight (0/1)	0.022	0.075	Elastic net
	[-0.053, 0.097]	[-2.008, 2.357]	
	(-0.053, 0.097)	(-2.322, 2.562)	
Height-for-age (Z-score)	0.236	0.362	Elastic net
	[-0.085, 0.551]	[-0.914, 1.610]	
	(-0.160, 0.645)	(-0.914, 1.610)	
Not stunted (0/1)	0.117	-0.598	Random forest
	[0.021, 0.209]**	[-1.563, 0.426]	
	(-0.008, 0.243)	(-2.144, 0.805)	
Consumed animal-source foods in past day (0/1)	0.055	1.134	Random forest
	[-0.048, 0.155]	[0.295, 1.979]**	
	(-0.060, 0.165)	(0.116, 2.262)	
One-day dietary diversity score	0.125	0.911	Random forest
	[-0.254, 0.510]	[0.188, 1.630]**	
	(-0.268, 0.524)	$(0.035, 1.717)^{+}$	

Notes: * p <0.10, ** p <0.05, *** p <0.01; $^{+}$ q <0.10, $^{++}$ q <0.05, $^{+++}$ q <0.01, where q is the false discovery rate. 90% confidence intervals in brackets, 90% false discovery rate-adjusted confidence intervals in parentheses. Standard errors and degrees of freedom estimated as in Young (2016). See equation 3 for the regression specification.

Tables 7, 8, and 9 each shows estimated differences in means obtained for ten separate characteristics when comparing the children most and least affected by the program. Children enjoying the largest impacts on having consumed animal-source foods in the past day are from households that are relatively poor and unlikely to be connected to an electricity grid. Other baseline characteristics showing significant differences include the dependency ratio, having a dirt floor, giving government-provided micronutrient supplements to their children, and having a loan. But these additional differences are sensitive to adjusting for multiple comparisons. For the one-day dietary diversity score, children experiencing the largest im-

Table 6: Intent-to-treat effects for most and least affected children

	Most	Least	(Most - Least)
Weight-for-age (Z-score)	0.371	-0.024	0.382
	[-0.095, 0.810]	[-0.533, 0.496]	[-0.370, 1.143]
	(-0.172, 0.870)	(-0.533, 0.496)	(-0.459, 1.248)
Not underweight (0/1)	0.053	0.005	0.033
	[-0.135, 0.246]	[-0.188, 0.192]	[-0.254, 0.356]
	(-0.142, 0.255)	(-0.199, 0.203)	(-0.273, 0.371)
Height-for-age (Z-score)	0.347	0.272	0.052
	[-0.338, 1.060]	[-0.382, 0.876]	[-0.955, 1.107]
	(-0.386, 1.128)	(-0.482, 0.930)	(-0.998, 1.146)
Not stunted (0/1)	0.002	0.199	-0.194
	[-0.225, 0.225]	[-0.022, 0.429]	[-0.548, 0.175]
	(-0.225, 0.225)	(-0.110, 0.516)	(-0.623, 0.255)
Consumed animal-source foods in past day (0/1)	0.281	-0.124	0.404
	[0.052, 0.509]**	[-0.373, 0.130]	[0.005, 0.798]*
	(-0.022, 0.640)	(-0.414, 0.177)	(-0.022, 0.821)
One-day dietary diversity score	0.824	-0.421	1.244
	[0.003, 1.635]*	[-1.321, 0.402]	[-0.029, 2.595]
	(-0.173, 1.902)	(-1.606, 0.611)	(-0.029, 2.595)

Notes: * p <0.10, ** p <0.05, *** p <0.01; $^{+}$ q <0.10, $^{++}$ q <0.05, $^{+++}$ q <0.01, where q is the false discovery rate. 90% confidence intervals in brackets, 90% false discovery rate-adjusted confidence intervals in parentheses. Standard errors and degrees of freedom estimated as in Young (2016). See equation 4 for the regression specification.

pacts are from households that have relatively low average household height-for-age, are more likely to use micronutrient supplements, and own larger chicken flocks at baseline. But all of the differences in characteristics between the children most and least affected with respect to dietary diversity lose statistical significance when adjusting for multiple comparisons.

Table 7: Classification analysis for child outcomes

		(Most - Least)	
	Sample average	Animal-source foods (0/1)	Dietary diversity (count)
Woman (0/1)	0.445	-0.014	-0.000
		[-0.180, 0.147]	[-0.166, 0.163]
		(-0.188, 0.155)	(-0.172, 0.169)
Age in months	31.743	1.208	2.528
		[-3.946, 6.519]	[-2.656, 7.712]
		(-4.023, 6.601)	(-3.561, 8.616)
Household average weight-for-age	-0.616	-0.155	-0.182
		[-0.590, 0.281]	[-0.628, 0.260]
		(-0.667, 0.349)	(-0.691, 0.326)
Household average height-for-age	-1.056	-0.427	-0.747
		[-1.010, 0.161]	[-1.318, -0.160]**
		(-1.173, 0.317)	(-1.658, 0.170)
Baseline weight-for-age	-0.155	-0.059	-0.032
		[-0.408, 0.289]	[-0.371, 0.310]
		(-0.434, 0.318)	(-0.405, 0.343)
Baseline height-for-age	-0.285	-0.059	-0.355
		[-0.540, 0.401]	[-0.811, 0.109]
		(-0.550, 0.412)	(-0.939, 0.237)
Meters above sea level	765.182	47.808	12.225
		[-52.605, 146.907]	[-85.503, 111.672]
		(-72.885, 167.034)	(-86.245, 112.418)
Rainfall (millimeters)	467.225	30.933	87.053
		[-78.066, 139.856]	[-20.176, 195.244]
		(-78.066, 139.856)	(-48.889, 224.365)
Days with extreme temperatures	173.961	-5.653	-7.194
		[-26.275, 14.736]	[-26.922, 13.317]
		(-28.147, 16.807)	(-28.941, 15.377)
Household size	4.102	-0.229	-0.293
		[-0.787, 0.351]	[-0.862, 0.292]
		(-0.892, 0.441)	(-0.967, 0.404)

Notes: * p <0.10, ** p <0.05, *** p <0.01; $^{+}$ q <0.10, $^{++}$ q <0.05, $^{+++}$ q <0.01, where q is the false discovery rate. 90% confidence intervals in brackets, 90% false discovery rate-adjusted confidence intervals in parentheses. Standard errors are of the heteroskedasticity-robust "HC1" variety (MacKinnon and White, 1985). All characteristics displayed in the rows are measured at baseline except for sex, age, weather variables, and altitude of the dwelling. Weather variables are measured for the 2015 crop season, approximately one year after baseline interviews. "Days with extreme temperatures" are days with maximum temperature over 30°C. "Baseline weight-for-age" and "Baseline height-for-age" are lagged Z-scores set to zero in the case of missing values, while "Household average weight-for-age" and "Household average height-for-age" are the lagged average Z-scores among siblings of the same sex as a given observation. Household size is in adult male equivalents.

Table 8: Classification analysis for child outcomes, continued

		(Most - Least)	
	Sample average	Animal-source foods (0/1)	Dietary diversity (count)
Dependency ratio	0.488	0.070	0.014
		[0.012, 0.129]**	[-0.048, 0.073]
		(-0.009, 0.150)	(-0.050, 0.074)
Average education (years)	2.639	-0.569	0.403
		[-1.639, 0.514]	[-0.658, 1.452]
		(-1.857, 0.722)	(-0.796, 1.574)
Wealth (log)	9.185	-0.809	-0.366
		[-1.204, -0.415]***	[-0.786, 0.051]
		(-1.477, -0.172)$^{+++}$	(-0.915, 0.169)
Dwelling has dirt floor (0/1)	0.748	0.153	0.042
		[0.016, 0.290]*	[-0.103, 0.186]
		(-0.028, 0.334)	(-0.103, 0.186)
Micronutrient supplement (0/1)	0.511	0.222	0.181
		[0.061, 0.383]**	[0.018, 0.344]*
		(-0.011, 0.454)	(-0.065, 0.427)
Receives CCT (0/1)	0.409	-0.014	-0.139
		[-0.176, 0.151]	[-0.303, 0.019]
		(-0.200, 0.175)	(-0.361, 0.075)
Women's share of wealth	0.306	0.020	0.010
		[-0.081, 0.121]	[-0.091, 0.112]
		(-0.084, 0.125)	(-0.098, 0.119)
Had credit at baseline (0/1)	0.116	-0.139	-0.097
		[-0.241, -0.032]**	[-0.202, 0.014]
		(-0.281, 0.011)	(-0.240, 0.050)
Distance to market (minutes)	64.264	1.847	-3.361
		[-14.908, 18.441]	[-20.110, 12.890]
		(-16.185, 19.712)	(-20.377, 13.137)
Dwelling has water filter (0/1)	0.132	0.014	0.042
		[-0.098, 0.121]	[-0.075, 0.143]
		(-0.111, 0.135)	(-0.088, 0.154)

Notes: * p <0.10, ** p <0.05, *** p <0.01; $^{+}$ q <0.10, $^{++}$ q <0.05, $^{+++}$ q <0.01, where q is the false discovery rate. 90% confidence intervals in brackets, 90% false discovery rate-adjusted confidence intervals in parentheses. Standard errors are of the heteroskedasticity-robust "HC1" variety (MacKinnon and White, 1985). All characteristics displayed in the rows are measured at baseline.

Table 9: Classification analysis for child outcomes, continued

		(Most - Least)	
	Sample average	Animal-source foods (0/1)	Dietary diversity (count)
Connected to water network (0/1)	0.577	0.042	0.153
		[-0.121, 0.205]	[-0.005, 0.315]
		(-0.128, 0.212)	(-0.067, 0.379)
Connected to electricity network (0/1)	0.461	-0.250	-0.069
		[-0.410, -0.090]***	[-0.234, 0.096]
		(-0.491, -0.008)$^{+}$	(-0.256, 0.118)
Social capital index, household	1.466	0.097	0.167
		[-0.495, 0.723]	[-0.490, 0.782]
		(-0.500, 0.727)	(-0.540, 0.827)
Social capital, women	0.807	0.125	0.097
		[-0.267, 0.495]	[-0.298, 0.490]
		(-0.314, 0.543)	(-0.312, 0.502)
Calories (log)	8.341	-0.046	0.162
		[-0.303, 0.201]	[-0.094, 0.421]
		(-0.317, 0.215)	(-0.150, 0.479)
Animal protein (log grams)	5.576	-1.748	0.603
		[-4.414, 0.986]	[-2.133, 3.532]
		(-5.015, 1.690)	(-2.288, 3.689)
Food consumption score	19.389	-1.785	0.403
		[-4.272, 0.706]	[-2.110, 2.726]
		(-4.891, 1.301)	(-2.232, 2.854)
Egg unit value (quetzales/egg)	1.186	-0.034	-0.020
		[-0.070, 0.003]	[-0.057, 0.015]
		(-0.078, 0.012)	(-0.064, 0.023)
Eggs produced (units in last six months)	41.922	-13.299	19.404
		[-55.510, 27.832]	[-22.479, 57.994]
		(-60.992, 31.701)	(-29.760, 67.178)
Chickens owned	10.107	0.069	3.708
		[-3.445, 3.614]	[0.127, 7.150]*
		(-3.780, 3.868)	(-1.464, 8.833)

Notes: * p <0.10, ** p <0.05, *** p <0.01; $^{+}$ q <0.10, $^{++}$ q <0.05, $^{+++}$ q <0.01, where q is the false discovery rate. 90% confidence intervals in brackets, 90% false discovery rate-adjusted confidence intervals in parentheses. Standard errors are of the heteroskedasticity-robust "HC1" variety (MacKinnon and White, 1985). All characteristics displayed in the rows are measured at baseline. "Social capital" and "Social capital, women" are counts of memberships in organizations and groups of which the household and women in the household, respectively, are members. "Egg unit value" is the price paid per purchased egg.

6 Discussion

6.1 Are our household-level results consistent with a well-functioning program?

Our household-level analysis did not uncover clear evidence of program impacts. But there are several mitigating factors that could explain this result. First, the Marek's outbreak likely made impacts on chicken ownership short lived for some households. Second, the power analysis in appendix section A.6 shows that minimum detectable effects for most household-level outcomes are somewhat large, and low power is reflected by the wide 95% confidence intervals around estimated household-level impacts. Third, the program was rolled out over the course of a year and a half in the treatment group, and follow-up data were collected about two and a half years after the start of implementation. Naked-neck chickens have a productive life of two to three years, while program chickens produced offspring for 56% of surveyed households who received birds that were vaccinated against Marek's and 13% of households receiving chickens who did not get the vaccination (we do not know whether offspring produced eggs). We would not expect the household-level indicators used in our analysis to be permanently affected by a temporary increase in egg production and chicken ownership. In other words, program impacts on household-level indicators may have come and gone for some households by the time follow-up data were collected. The potentially transitory nature of household-level program effects stands in contrast to impacts on stunting (and to a lesser degree underweight), which can be permanently improved through proper nutrition in early childhood (Choudhury, Headey, and Masters, 2019).

6.2 Explaining larger average impacts on girls than boys

Our results from section 5.2 show a clear pattern of larger impacts on girls. A first possible explanation for this pattern is survivor effects, e.g. the program is allowing boys with relatively low Z-scores to survive until the follow-up survey. But only two children who were young enough at baseline to later be included in our impact analysis did not survive to the follow-up survey, and the program had no impact on the number of boys or girls under 60 months of age at the household level (see appendix A.21). A second explanation is that our results reflect sex-specific differences in anthropometric indicators in the absence of treatment, or in the treatment group prior to intervention. We examine this possibility by comparing Z-scores for boys and girls in the control group at follow-up and in the treatment group at baseline. Kolmogorov-Smirnov tests and t-tests fail to reject the null of no difference by sex in all cases, regardless of whether we condition on being stunted (when testing height-for-age) or underweight (when testing weight-for-age).

Alternatively, it could be that boys are more likely than girls to suffer from intestinal illness, undermining gains in height and weight through poor absorption of nutrients (Guerrant et al., 2012). As mentioned earlier, the program itself could increase intestinal illness by worsening exposure to pathogens in animal waste. We explore whether the program affected intestinal illness among children as well as differences in intestinal illness by sex in table 10. The first and second columns of table 10 show no evidence that the program had deleterious effects on child intestinal health. But in the three rightmost columns we see that the incidence and average duration of intestinal illness are substantially higher among boys, where the latter is calculated using all children, not just those who were sick.[6]

If we could identify the mechanism that is driving higher rates of intestinal illness among boys, then we might be able to propose a complementary intervention that could raise the effectiveness of future livestock transfer programs. One possible explanation for what we observe in table 10 is greater exposure to pathogens among infant boys through consumption of drinks mixed with unclean water, e.g. infant formula (Anttila-Hughes et al.,

[6]Note that the q-values and confidence intervals shown in table 10 do not incorporate the hypothesis tests for our main outcomes of interest shown in table 4. We consider the tests in each table to represent different families of hypotheses. The tests in table 4 each give us a separate chance to determine whether the program had a positive effect on children, whereas the tests in table 10 all relate to potential negative effects of the program.

2018). In our data, boys between the ages of six and 60 months are more than twice as likely to consume infant formula than girls ($p < 0.005$) at follow-up, but consumption is small overall (5.8% of boys versus 2.6% of girls). There are no significant differences by sex in the probability of consuming any other drinks that might deliver pathogens.

Another candidate explanation is that girls are favored by households in the distribution of eggs or meat. Regressing a series of binary indicators for having consumed different animal-source foods in the past 24 hours on sex and age in months reveals no pattern of favoritism towards girls or boys. Differences could be with respect to quantity, however. But any existing bias in food distribution in the Guatemalan context seems more likely to favor boys than girls (Frongillo and Begín, 1993). Our results could also be explained by differences in breastfeeding patterns. For example, girls might stop exclusively breastfeeding earlier than boys, and move on to complementary foods. Our data do not have detailed information on breastfeeding, so we turn to the the 2014-2015 National Survey of Maternal and Child Health (NSMCH), which includes a random sample of 566 children between the ages of zero and 60 months from Chiquimula. The NSMCH data show that conditional on age in months, sex has no detectable correlation with whether a child is currently breastfeeding or duration of breastfeeding. And as already noted, we find no differences in diet by sex among children under 60 months of age in our own data. Overall, differences in breastfeeding and consumption of complementary foods do not seem to provide an explanation for higher rates of intestinal illness among boys. We do not have the data needed to consider other channels of pathogen exposure, such as differences in the play habits of boys and girls.

Finally, several previous studies have evaluated policy changes that increased asset ownership or control over income among women while affecting girls but not boys, including Duflo (2003), Qian (2008), Matz and Narciso (2010), and de Carvalho Filho (2012), while suggesting improved women's bargaining strength as an explanation. Given our imprecise impact estimates on chicken ownership, we cannot rule out improved bargaining strength through greater asset ownership as a mechanism.

Table 10: Intestinal illness, children ages 6 to 60 months

	Intent-to-treat effects			Sample averages		
	Girls	Boys	Difference	Girls	Boys	Difference
Animal feces in or around the dwelling (0/1)	0.111	0.034	0.077	0.438	0.412	0.018
	[-0.125, 0.346]	[-0.100, 0.167]	[-0.151, 0.305]			[-0.074, 0.110]
	(-0.150, 0.371)	(-0.114, 0.181)	(-0.216, 0.370)			(-0.074, 0.110)
Intestinal illness in past 30 days (0/1)	0.067	0.082	-0.015	0.304	0.395	-0.098
	[-0.056, 0.189]	[-0.144, 0.308]	[-0.204, 0.174]			[-0.183, -0.013]**
	(-0.091, 0.225)	(-0.209, 0.373)	(-0.204, 0.174)			(-0.190, -0.006)$^{++}$
Days with intestinal illness in past 30 days	0.356	0.202	0.154	1.861	2.761	-0.997
	[-0.706, 1.418]	[-1.662, 2.065]	[-1.454, 1.761]			[-1.753, -0.241]***
	(-0.706, 1.418)	(-1.662, 2.065)	(-1.624, 1.931)			(-1.922, -0.072)$^{++}$
Observations	196	244	440	196	244	440

Notes: * p <0.10, ** p <0.05, *** p <0.01; $^{+}$ q <0.10, $^{++}$ q <0.05, $^{+++}$ q <0.01, where q is the false discovery rate. 95% confidence intervals in brackets, 95% false discovery rate-adjusted confidence intervals in parentheses. Standard errors and degrees of freedom estimated as in Young (2016). All regressions include an intercept, indicators for thirteen strata, and child age in months. The standard errors for the rightmost column are adjusted for clustering at the household level.

6.3 Heterogeneous impacts on diet among children

Children in the poorest households enjoy the largest impacts on animal-source foods consumption at follow-up, over two years after the start of the program. While we might expect higher current consumption of animal-source foods to be accompanied by gains in height or weight, we do not have strong evidence for heterogeneous impacts on Z-scores, stunting, or underweight. One possible explanation for the disconnect between impacts on animal-source foods and height and weight is that poorer households have had their chickens long enough to enjoy increased access to animal source foods, but not long enough for chickens to reach the end of their productive lives or for child height and weight to be affected. This situation might occur if the Mancomunidad opted to treat wealthier clusters first. We check this possibility by regressing the chicken set delivery date (specified as the number of days from an arbitrary point in time to the delivery date) on baseline household wealth and find no relationship (p=0.566).

Another explanation is that the poorest children face additional constraints that prevent them from transforming higher animal-source foods consumption into anthropometric gains. Oral-fecal contamination, for example, has been blamed for the failure of nutritional interventions and oral vaccines in other contexts (Ngure et al., 2014), motivating the incorporation of water, sanitation, and hygiene components into nutrition interventions (e.g. Tofail et al. (2018)). As shown in table 10, 35% of children in our data set between the ages of six and 60 months had diarrhea in the 30 days prior to follow-up interviews, suggesting that oral-fecal contamination is common in the program's intervention area.

Access to clean water is scarce in the Mancomunidad, and is likely a problem for households of all wealth levels. But hygiene appears to be positively correlated with wealth in our data. Enumerators confirmed the presence of a hand-washing station with soap in about 50% of household at follow-up.[7] Baseline wealth is 41% higher among households with a

[7]Note that we did not ask this question at baseline so it was not included in our machine learning analysis.

hand-washing station than in households without. Children who reported having diarrhea in the 30 days prior to follow-up data collection are from households with 10% less wealth on average than their counterparts who did not report having diarrhea. Although household wealth could be correlated with a variety of factors that mediate program impacts, it is at least plausible that hygiene and sanitation considerations are partly responsible for the patterns we see in impacts on anthropometrics and consumption of animal-source foods among children.

7 Conclusion

We evaluated the chicken transfer and training component of a program seeking to build resilience in rural Guatemala using a cluster-randomized phase-in design. We estimate program intent-to-treat effects by comparing 14 clusters of communities assigned to receive the program in its second year to 14 clusters assigned to receive the program in year four. At follow-up, 36% of households assigned to treatment had participated in the program long enough to plausibly have experienced impacts, whereas the same was true of 9% of households assigned to the control group.

We find no statistically significant average intent-to-treat effects on household-level measures of expenditure, nutrient intake, diet quality, chicken ownership, egg consumption, egg production and sales, and chicken management. In contrast, we find clear evidence for large impacts on weight and height among girls between the ages of six and 60 months. Most notably, stunting fell by 23.5 percentage points ($\pm$ 19.4) among girls, an improvement of 57% relative to the control group. We find no statistically significant average intent-to-treat effects on height or weight among boys. Differences in height and weight impacts by sex could be a product of higher rates of intestinal illness among boys. Importantly, the program does not appear to have increased the rate or severity of intestinal illness among children. Although we cannot reject the null of a zero average intent-to-treat effect on the one-day dietary diversity score or an indicator for having consumed animal-source foods in the past day for children of either sex, we do find evidence of heterogeneous effects for these two indicators, with children from the poorest households experiencing the largest impacts on consumption of animal source foods. But we find no corresponding heterogeneity in height or weight impacts by wealth. Descriptive evidence suggests that differences in hygiene by wealth could be to blame, as poor hygiene can lead to increased intestinal illness and reduced absorption of nutrients.

We acknowledge that power hampers our ability to detect household-level effects. Furthermore, the slow pace of program rollout and the typical productive lifespan of a naked-neck chicken may have resulted in impacts dissipating for some households before follow-up data were collected. As for impacts on health among children, the lack of effects on mechanisms like chicken ownership and the fact that some households participated in other program components complicates interpretation of our results. But we see the bulk of the evidence presented here as pointing to chicken transfers driving improvements in health among girls.

The main policy implication of our results is that livestock transfer programs can have dramatic impacts on child health even when undermined by shocks like disease outbreaks. In addition, there is room to increase the effectiveness of livestock transfers through complementary interventions, such as training on nutrition and intrahousehold food distribution, animal waste management, and improved hygiene. Future research can help improve livestock transfer effectiveness by building complementary interventions into randomized trials. Researchers would also be well-advised to collect detailed intra-household data on the drivers of treatment effect heterogeneity within the household, such as women's bargaining strength and how households divide food among their members.

References

Abadie, A., S. Athey, G. Imbens, and J. Wooldridge. 2017. "When Should You Adjust Standard Errors for Clustering?" Working paper, National Bureau of Economic Research.

Anttila-Hughes, J., L. Fernald, P. Gertler, P. Krause, and B. Wydick. 2018. "Mortality from Nestl's Marketing of Infant Formula in Low and Middle-Income Countries." Working paper No. 24452, National Bureau of Economic Research.

Athey, S., J. Tibshirani, and S. Wager. 2019. "Generalized random forests." *The Annals of Statistics* 47:1148–1178.

Attanasio, O., H. Baker-Henningham, R. Bernal, C. Meghir, D. Pineda, and M. Rubio-Codina. 2018. "Early Stimulation and Nutrition: The Impacts of a Scalable Intervention." Working paper No. 25059, National Bureau of Economic Research.

Attanasio, O., C. Meghir, and E. Nix. 2019. "Human Capital Development and Parental Investment in India." Unpublished.

Bandiera, O., R. Burgess, N. Das, S. Gulesci, I. Rasul, and M. Sulaiman. 2013. "Can Basic Entrepreneurship Transform the Economic Lives of the Poor?" IZA Discussion Paper No. 7386.

Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Pariente, J. Shapiro, B. Thuysbaert, and C. Udry. 2015. "A multifaceted program causes lasting progress for the very poor: Evidence from six countries." *Science* 348:1260799–1260799.

Belloni, A., V. Chernozhukov, and C. Hansen. 2014. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives* 28(2):29–50.

Benjamini, Y., and Y. Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society SeriesB* 57:289–300.

Benjamini, Y., and D. Yekutieli. 2005. "False Discovery Rate–Adjusted Multiple Confidence Intervals for Selected Parameters." *Journal of the American Statistical Association* 100:71–81.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45:5.

Chen, C.F., A. Bordas, D. Gourichon, and M. Tixier-Boichard. 2004. "Effect of high ambient temperature and naked neck genotype on performance of dwarf brown-egg layers selected for improved clutch length." *British Poultry Science* 45:346–354.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. 2018a. "Double/debiased machine learning for treatment and structural parameters." *The Econometrics Journal* 21:C1–C68.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernndez-Val. 2018b. "Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments." Working paper, National Bureau of Economic Research.

Choudhury, S., D.D. Headey, and W.A. Masters. 2019. "First foods: Diet quality among infants aged 6–23months in 42 countries." *Food Policy*, sep, pp. 101762.

CNN. 2016. "Can Bill Gates' donation of 100.000 chickens help Africa's poorest?"

Darrouzet-Nardi, A.F., L.C. Miller, N. Joshi, S. Mahato, M. Lohani, and B.L. Rogers. 2016. "Child dietary quality in rural Nepal: Effectiveness of a community-level development intervention." *Food Policy* 61:185–197.

de Carvalho Filho, I.E. 2012. "Household Income as a Determinant of Child Labor and School Enrollment in Brazil: Evidence from a Social Security Reform." *Economic Development and Cultural Change* 60:399–435.

Duflo, E. 2003. "Grandmothers and Granddaughters: Old-Age Pensions and Intrahousehold Allocation in South Africa." *The World Bank Economic Review* 17:1–25.

El Nuevo Diario. 2016. "The Dry Corridor of Central America, Facing its Worst Drought in the last 30 Years (*El Corredor Seco de Centroamrica, ante la Peor Sequa de los ltimos 30 Aos*)." *El Nuevo Diario*, May, pp. .

Emran, M.S., V. Robano, and S.C. Smith. 2014. "Assessing the Frontiers of Ultrapoverty Reduction: Evidence from Challenging the Frontiers of Poverty Reduction/Targeting the Ultra-poor, an Innovative Program in Bangladesh." *Economic Development and Cultural Change* 62:339–380.

FAO. 2017. "Chronology of the Dry Corridor: The Impetus for Resilience in Central America." Food and Agriculture Organization of the United Nations.

Figlio, D., J. Guryan, K. Karbownik, and J. Roth. 2014. "The Effects of Poor Neonatal Health on Childrens Cognitive Development." *The American Economic Review* 104:39213955.

Frongillo, E., and F. Begín. 1993. "Gender Bias in Food Intake Favors Male Preschool Guatemalan Children." *Journal of Nutrition*, pp. .

Gates, B. 2016. "Why I Would Raise Chickens." GatesNotes: The Blog of Bill Gates.

Gillespie, B. 2018. "Sprinkles and Spacing." *Anthropology in Action* 25:24–35.

Glass, N., N.A. Perrin, A. Kohli, J. Campbell, and M.M. Remy. 2017. "Randomised controlled trial of a livestock productive asset transfer programme to improve economic and health outcomes and reduce intimate partner violence in a postconflict setting." *BMJ Global Health* 2:e000165.

Government of Guatemala. 2013. "Impact of the Extended Summer on the Marginal and Subsistence Population of the Dry Corredor in Guatemala (*Impacto de la Cancula Prolongada en La Poblacin de Infra y Subsistencia del Corredor Seco de Guatemala*)."

Guatemala National Institute of Statistics. 2015. "Republic of Guatemala: National Survey of Living Standards 2014." Working paper, Guatemala National Institute of Statistics, Guatemala.

Guerrant, R.L., M.D. DeBoer, S.R. Moore, R.J. Scharf, and A.A.M. Lima. 2012. "The Impoverished Gut—a Triple Burden of Diarrhoea, Stunting and Chronic Disease." *Nature Reviews Gastroenterology & Hepatology* 10:220–229.

Headey, D., and K. Hirvonen. 2016. "Is Exposure to Poultry Harmful to Child Nutrition? An Observational Analysis for Rural Ethiopia." *PLOS ONE* 11:1–16.

Headey, D., K. Hirvonen, and J. Hoddinott. 2018. "Animal Sourced Foods and Child Stunting." *American Journal of Agricultural Economics* 100:1302–1319, IFPRI Discussion Paper 01695.

Headey, D., R. Rawat, S. Kim, P. Menon, M. Ruel, and P. Nguyen. 2017. "Is Exposure to Animal Feces Harmful to Child Nutrition and Health Outcomes? A Multicountry Observational Analysis." *The American Journal of Tropical Medicine and Hygiene* 96:961–969.

Hoddinott, J., J. Maluccio, J. Behrman, R. Flores, and R. Martorell. 2008. "Effect of a Nutrition Intervention During Early Childhood on Economic Productivity in Guatemalan Adults." *The Lancet* 371:411–416.

Iannotti, L.L., C.K. Lutter, C.P. Stewart, C.A.G. Riofrío, C. Malo, G. Reinhart, A. Palacios, C. Karp, M. Chapnick, K. Cox, and W.F. Waters. 2017a. "Eggs in Early Complementary Feeding and Child Growth: A Randomized Controlled Trial." *Pediatrics* 140:e20163459.

Iannotti, L.L., C.K. Lutter, W.F. Waters, C.A.G. Riofrío, C. Malo, G. Reinhart, A. Palacios, C. Karp, M. Chapnick, K. Cox, S. Aguirre, L. Narvaez, F. López, R. Sidhu, P. Kell, X. Jiang, H. Fujiwara, D.S. Ory, R. Young, and C.P. Stewart. 2017b. "Eggs Early in

Complementary Feeding Increase Choline Pathway Biomarkers and DHA: a Randomized Controlled Trial in Ecuador." *The American Journal of Clinical Nutrition* 106:1482–1489.

Imbens, G., and J. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62:467–475.

INCAP. 2012. "Food Composition Table for Central America and Panama (Tabla de Composición de Alimentos de Centroamérica)." techreport, Nutrition Institute of Central America and Panama (INCAP).

INE. 2015. "Republic of Guatemala: National Living Standards Survey of 2014 (*República de Guatemala: Encuesta Nacional de Condiciones de Vida 2014*)." National Institute of Statistics (*Instituto Nacional de Estadística*).

Janzen, S., N. Magnan, S. Sharma, and W. Thompson. 2018. "Short-Term Impacts of a Pay-It-Forward Livestock Transfer and Training Program in Nepal." *AEA Papers and Proceedings* 108:422–25.

Jin, M., and L.L. Iannotti. 2014. "Livestock production, animal source food intake, and young child growth: The role of gender for ensuring nutrition impacts." *Social Science & Medicine* 105:16–21.

Jodlowski, M., A. Winter-Nelson, K. Baylis, and P.D. Goldsmith. 2016. "Milk in the Data: Food Security Impacts from a Livestock Field Experiment in Zambia." *World Development* 77:99–114.

Kafle, K., A. Winter-Nelson, and P. Goldsmith. 2016. "Does 25 cents more per day make a difference? The impact of livestock transfer and development in rural Zambia." *Food Policy* 63:62–72.

Krishna, A., M. Poghosyan, and N. Das. 2012. "How Much Can Asset Transfers Help the

Poorest? Evaluating the Results of BRAC's Ultra-Poor Programme (2002–2008)." *Journal of Development Studies* 48:254–267.

Lakens, D. 2016. "Why you don't need to adjust your alpha level for all tests you'll do in your lifetime." Web site.

Lee, D. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *The Review of Economic Studies* 76:1071–1102.

Lobell, D.B., M. Bnziger, C. Magorokosho, and B. Vivek. 2011. "Nonlinear Heat Effects on African Maize as Evidenced by Historical Yield Trials." *Nature Climate Change* 1:42–45.

MacKinnon, J., and H. White. 1985. "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics* 29:305–325.

Matz, J.A., and G. Narciso. 2010. "Does Reinforcing Spouses Land Rights Improve Childrens Outcomes? Evidence from a Quasi-Natural Experiment in Rural Vietnam." Institute for International Integration Studies Discussion Paper 348.

Miller, L.C., N. Joshi, M. Lohani, B. Rogers, M. Kershaw, R. Houser, S. Ghosh, J.K. Griffiths, S. Mahato, and P. Webb. 2016. "Duration of programme exposure is associated with improved outcomes in nutrition and health: the case for longer project cycles from intervention experience in rural Nepal." *Journal of Development Effectiveness* 9:101–119.

Miller, L.C., N. Joshi, M. Lohani, B. Rogers, M. Loraditch, R. Houser, P. Singh, and S. Mahato. 2014. "Community development and livestock promotion in rural Nepal: Effects on child growth and health.", pp. .

Ministry of Public Health and Social Assistance. 2017. "National Survey of Maternal and Child Health 2014-2015: Final Report." Working paper, Ministry of Public Health and Assistance, National Institute of Statistics, and ICF International.

Misha, F.A., W.A. Raza, J. Ara, and E. van de Poel. 2019. “How Far Does a Big Push Really Push? Long-Term Effects of an Asset Transfer Program on Employment Trajectories.” *Economic Development and Cultural Change*, sep, pp. 000–000.

Multilateral Investment Fund. 2018. “Project Status Report (Reporte de Estado del Proyecto).”

National Institute of Population Research and Training. 2016. “Bangladesh Demographic and Health Survey 2014.” Working paper, NIPORT, Mitra and Associates, and ICF International, Dhaka, Bangladesh.

National Institute of Statistics of Rwanda. 2015. “2014-2015 RDHS Key Findings.” Working paper, NISR, Ministry of Health, and ICF International, Rockville, MD.

Nepal Ministry of Health. 2016. “Nepal: 2016 Demographic and Health Survey Key Findings.” Working paper, Ministry of Health, Nepal, Kathmandu.

Ngure, F., B. Reid, J. Humphrey, M. Mbuya, G. Pelto, and R. Stoltzfus. 2014. “Water, sanitation, and hygiene (WASH), environmental enteropathy, nutrition, and early child development: making the links.” *Annals of the New York Academy of Sciences* 1308:118–128.

Padilla, V.F. 2017. “Comsultancy for the Description of Regional Mechanisms for the Marketing of Basic Grains and the Efficiency of the Production Systems of CUNORI: Final Report.” Working paper, Opportunities and New Businesses.

Qian, N. 2008. “Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance.” *Quarterly Journal of Economics* 123:1251–1285.

Rawlins, R., S. Pimkina, C.B. Barrett, S. Pedersen, and B. Wydick. 2014. “Got milk? The impact of Heifer International’s livestock donation programs in Rwanda on nutritional outcomes.” *Food Policy* 44:202–213.

Raza, W.A., E.V. de Poel, and T.V. Ourti. 2018. "Impact and spill-over effects of an asset transfer program on child undernutrition: Evidence from a randomized control trial in Bangladesh." *Journal of Health Economics* 62:105–120.

Roodman, D., M.Ø. Nielsen, J.G. MacKinnon, and M.D. Webb. 2019. "Fast and wild: Bootstrap inference in Stata using boottest." *The Stata Journal: Promoting communications on statistics and Stata* 19:4–60.

Roy, S., J. Ara, N. Das, and A.R. Quisumbing. 2015. ""Flypaper effects" in transfers targeted to women: Evidence from BRAC's "Targeting the Ultra Poor" program in Bangladesh." *Journal of Development Economics* 117:1–19.

Sonaiya, E., and S. Swan. 2004. *Small-Scale Poultry Poultry Production: Technical Guide*. Food and Agriculture Organization of the United Nations.

Tofail, F., L. Fernald, K. Das, M. Rahman, T. Ahmed, K. Jannat, L. Unicomb, B. Arnold, S. Ashraf, P. Winch, P. Kariger, C. Stewart, J. Colford, and S. Luby. 2018. "Effect of Water Quality, Sanitation, Hand Washing, and Nutritional Interventions on Child Development in Rural Bangladesh (WASH Benefits Bangladesh): a Cluster-Randomised Controlled Trial." *The Lancet Child & Adolescent Health* 2:255–268.

Wong, J., J. de Bruyn, B. Bagnol, H. Grieve, M. Li, R. Pym, and R. Alders. 2017. "Small-scale poultry and food security in resource-poor settings: A review." *Global Food Security* 15:43–52.

World Bank. 2019a. "GDP per capita, PPP." Accessed October 1, 2019.

—. 2019b. "GINI index (World Bank estimate)." Accessed October 1, 2019.

—. 2019c. "Rural population (% of total population)." Accessed October 1, 2019.

World Food Program. 2008. "Food Consumption Analysis: Calculation and Use of the Food Consumption Socre in Food Security Analysis." Working paper, World Food Program.

—. 2018. "Guatemala."

World Health Organization. 2011. "WHO Anthro version 3.2.2."

Young, A. 2016. "Improved, Nearly Exact, Statistical Inference with Robust and Clustered Covariance Matrices using Effective Degrees of Freedom Corrections." Unpublished.

Zou, H., and T. Hastie. 2005. "Regularization and variable selection via the elastic net." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67:301–320.