Short-term impacts of a pay-it-forward livestock transfer and training program in Nepal

Sarah Janzen[1], Nicholas Magnan[2], and William M. Thompson[2]

[1]Montana State University

[2]University of Georgia

December 18, 2017

Abstract

This study evaluates the short term impacts of Heifer International's livestock transfer and training program in Nepal using a randomized control trial (RCT). The RCT assigned three variations of the program in order to capture differential effects of various program components. We also evaluate a unique "pay it forward" program rule where recipients are encouraged to share what they have learned and even share some of their newly accumulated wealth, in the form of a productive asset, to other households in need. After 1.5 years, we find financial inclusion increases by 0.31 standard deviations, and empowerment increases by 0.24 standard deviations among direct beneficiaries. We observe "pay it forward" effects to indirect beneficiaries in financial inclusion and empowerment that are of approximately the same magnitude as direct beneficiaries. These observed short-term impacts are similar across the different program variations.

1 Introduction

It is often argued that the rural poor largely lack access to the productive assets and human capital necessary to be successful entrepreneurs. Productive asset transfer programs, which typically include a training component, are one way non-governmental organizations (NGOs) and governments try to alleviate these constraints thereby facilitating a permanent transition out of poverty. These programs are popular among donors who subscribe to the well-known "teach a man to fish" maxim. In some cases, they also include a "pay it forward" component where recipients share what they have learned and even share some of their newly accumulated wealth, in the form of a productive asset, to other households in need. Rigorous impact evaluations of combined asset transfer and training programs, particularly evaluations designed to measure impacts on indirect beneficiaries, are few and far between.[1]

In this paper we evaluate the short-term (1.5 year) welfare impacts of Heifer International's (HI) Smallholders in Livestock Value Chain Program in rural Nepal using a randomized controlled trial (RCT). Like similar programs, the program targets poor households in rural areas, particularly women, and seeks to provide a sustainable livelihood and a pathway out of poverty for its beneficiaries. The standard intervention in Nepal provides a package of benefits that includes group formation, livestock (in this case two female goats), technical trainings on improved animal management and entrepreneurship, and values-based trainings. The values-based training encourages beneficiaries to "pay it forward" by providing technical training and giving the first-born female offspring of their received livestock to another poor individual in their community.

This paper contributes to the literature in three important ways. First, we add to a small but growing body of empirical evidence on the *overall* positive impact of livestock transfer programs worldwide (Bandiera et al., 2017; Banerjee et al., 2015; Darrouzet-Nardi et al., 2016; Jodlowski et al., 2016; Kafle, Winter-Nelson, and Goldsmith, 2016; Miller et al.,

[1] Banerjee et al. (2015) test for spillover effects of a program without a "pay it forward" component and find none.

2014; Rawlins et al., 2014). Second, our evaluation is carefully designed to estimate the aforementioned "pay it forward" indirect effects on members of the same targeted community who were not initially targeted by the implementing partner. Measuring the strength and persistence of this element of the program design is crucial to understanding the overall program impacts. Third, our evaluation includes three unique treatments in order to unpack the welfare impacts of different program components. In the first treatment arm, beneficiaries received a complete package that included a livestock transfer, skills-based technical trainings and values-based non-technical trainings. In the second treatment arm, beneficiaries received skills-based technical trainings and values-based non-technical trainings, but *not* livestock. In the third treatment arm beneficiaries received a livestock transfer and skills-based technical trainings, but *not* values-based non-technical trainings. To our knowledge, previous studies in this area do not attempt to disaggregate the impacts of a bundled treatment.

Our total baseline sample is 2,376 women across 60 treatment clusters (village development committees, or VDCs) stratified by region and ethnic composition. Between July and December 2014, HI administered training and delivered goats to randomly selected targeted beneficiaries. Various additional trainings continued throughout 2015. Shortly after HI delivered training and livestock to the original beneficiaries of the project, a devastating earthquake struck Nepal. The earthquake adversely affected ten villages originally included in the evaluation, spread evenly across treatment groups and control. These ten VDCs were dropped from the RCT so that HI could provide disaster relief as deemed appropriate. The remaining sample size comprises 50 villages and 1,828 individuals from which follow-up data was collected in June-July 2016. Although a definite concern, updated power calculations suggest the study remains sufficiently powered to capture overall treatment effects.

Our hypotheses, along with detailed plans for handling the data and analysis, are documented in a registered pre-analysis plan available at `http://www.socialscienceregistry.org/trials/1504`. Because we have a rich dataset informing numerous hypotheses regarding behavioral change and improved welfare across several dimensions, we employ summary indices to capture broad program impacts and reduce the number of hypotheses tested. These include income, asset holdings, expenditures, financial inclusion, food security, physical health, mental health, aspirations, and women's empowerment. Although we greatly reduce the number of hypotheses tested by using these indices, we still account for multiple hypotheses testing by controlling for the false discovery rate using the Benjamini and Hochberg (1995) step-up method.

We estimate intent-to-treat (ITT) effects of the overall program on directly targeted beneficiaries as well as beneficiaries brought into the program through the "pay it forward" (PIF) process, whom we term second generation PIF beneficiaries. In the short-term, we find the intervention causes a statistically significant 0.31 standard deviation increase in our index of financial inclusion, and a 0.24 standard deviation increase in empowerment among direct beneficiaries. Perhaps surprisingly, we do not observe statistically significant differences in outcome indices across treatments for direct beneficiaries, suggesting we are either underpowered to capture small differences between treatments, or the combination of activities is not necessary to achieve the desired impact in the short run (though perhaps in the long run it is). Exploratory analysis suggests these findings stem from increased saving and membership in savings (or other) groups, increased ownership of productive assets, and increased control over use of income.

We also observe substantial PIF effects. Those who live in the same community as direct beneficiaries but who were not targeted as direct beneficiaries themselves experience similar and statistically significant increases in financial inclusion (0.21 standard deviations) and empowerment (0.29 standard deviations) as those observed for directly targeted beneficiaries. These results are impressive given the relatively short time horizon over which to observe an impact on second generation PIF beneficiaries. Notably, this PIF effect vanishes when the values-based trainings are withheld, suggesting the "pay it forward" encouragement (a critical component of the values-based training) helps successfully achieve a broader impact.

Our exploratory analysis reveals additional positive impacts worth highlighting. We observe no impact on our summary measure of mental health, but direct beneficiaries have higher life satisfaction and self esteem, and the intervention may reduce worrying. We observe no impact on our summary measure of aspirations, but we do present evidence that direct beneficiaries positively adjust their aspirations for future income.

This paper focuses on short run impacts. We do not observe statistically significant changes in income, asset holdings, or expenditures in the short run. However, taking into account goat gestational periods and kid growth rates, livestock sales of transferred goats' kids are largely unanticipated within the timeframe of this study. In a program that targets livestock production as an income-generating activity, economic impacts within this short time horizon would be unanticipated. Future work will analyze the long-term impacts of the program on these economic outcomes. Although this planned future work is important, we think the current analysis makes an important contribution of its own. Evaluating short-term impacts is important for understanding causal mechanisms. This paper demonstrates how a multifaceted social protection program that combines trainings with an asset transfer has immediate effects in important dimensions of welfare. We also carefully evaluate the effectiveness of a unique "pay-it-forward" program component; our findings suggest this unique program component could be an important and cost-effective tool for achieving program goals.

2 Background

2.1 Evidence on asset transfer and training programs

Asset transfers, particularly livestock, have been conducted in poor countries since at least 1944, when HI sent 17 cows from Arkansas to Puerto Rico. Since then HI has expanded its reach to over 125 countries. Numerous NGOs and even governments have embraced livestock transfer and training programs as a strategy for fighting poverty (World Vision, BRAC, Save the Children, Oxfam, and the Government of Rwanda are a few examples).

Despite the long history and prevalence of livestock transfer and training programs, until recently there was very little rigorous empirical evidence of their effectiveness (DFID 2014). Recent papers have found that these programs increase income (Bandiera et al., 2017; Banerjee et al., 2015), expenditures (Bandiera et al., 2017; Banerjee et al., 2015; Jodlowski et al., 2016), savings (Bandiera et al., 2017; Banerjee et al., 2015), overall food consumption (Bandiera et al., 2017; Banerjee et al., 2015; Kafle, Winter-Nelson, and Goldsmith, 2016), dairy and meat consumption (Banerjee et al., 2015; Rawlins et al., 2014), dietary diversity (Darrouzet-Nardi et al., 2016; Jodlowski et al., 2016; Kafle, Winter-Nelson, and Goldsmith, 2016; Rawlins et al., 2014), food security (Bandiera et al., 2017; Banerjee et al., 2015), and anthropometrics (Miller et al., 2014; Rawlins et al., 2014). Evidence of impacts on emotional well-being and women's empowerment have been mixed (Bandiera et al., 2017; Banerjee et al., 2015; Roy et al., 2015).

Most notably, Banerjee et al. (2015) evaluate the impact of BRAC's graduation program, a large asset transfer and training program in six countries. Their study finds after three years the graduation program has significantly positive impacts on consumption, food security, assets, finance, time use, income and mental health, but no indirect effects for non-beneficiaries. They observe positive impacts on women's empowerment in the short but not long run.

Several major differences exist between the graduation program and the one evaluated here. First, beneficiaries of the graduation program chose an asset (or bundle of assets) from a list of productive assets. Although livestock was the most common choice, there were alternative options. The value of the productive asset transfer was always higher than the one evaluated here, and beneficiaries in their study also received a regular transfer of food or cash for a few months or even up to a year. In another significant deviation, beneficiaries of the livestock transfer and training program we study here were encouraged to "pay it forward," as described below. This encouragement is a central component of all livestock transfer and training programs implemented globally by Heifer International (HI). To our knowledge, no study has evaluated the impact of a program with this type of encouragement.

2.2 Heifer International's "Pay-it-forward" livestock transfer and training program in Nepal

The intervention we evaluate replicates HI's Smallholders in Livestock Value Chain (SLVC) Program in rural Nepal. Like similar programs, the program targets poor households in rural areas, and seeks to provide a sustainable livelihood and a pathway out of poverty for its primarily women beneficiaries. The standard HI intervention in Nepal provides a package of benefits that includes formation of women's self help and savings groups, technical trainings on improved animal management and entrepreneurship, values-based trainings, a productive asset transfer (in this case goats), and encouragement to "pay it forward".

The process is as follows: After identifying a location to receive the intervention, HI recruits an original group of direct beneficiaries. Direct beneficiary groups typically consist of close neighbors and often include most or all of the households in a given neighborhood. As a rule, HI considers all the households in a targeted area to be objectively poor and therefore eligible for the program, allowing for the possibility that a considerable range of relative wealth and poverty might exist within a group. Once selected, direct beneficiaries within a ward are organized into a self-help group (SHG). Over a period of several months all SHG members participate in a series of trainings. Trainings include (1) technical training on improved animal management, fodder/forage development, entrepreneurship, human and animal nutrition, and home gardening, and (2) HI's values-based training on topics of accountability, sharing and caring, sustainability, self-reliance, income management, environmental stewardship, spirituality, self-help group management, gender justice, and encouragement to pay it forward. The trainings culminate with the beneficiaries receiving a transfer of livestock which includes two doe goats for each beneficiary and a single buck of improved stock (to facilitate a breeding program) for the SHG.

A unique component of HI's model is that it encourages members to "pay it forward" by recruiting additional community members into the program, giving a gift of livestock (of equal value to what was received), and passing down all knowledge that was gained through participation in the programs. HI facilitates values-based empowerment training for both direct and PIF beneficiaries (albeit separately and at different points in time), while all other "pay it forward" trainings are implemented by direct beneficiaries with minimal support from HI. In this way, what might typically be deemed a spillover effect is actually an important program component. The program we evaluate follows an innovation to the basic HI pay-it-forward model, in which each direct beneficiary SHG is tasked with recruiting up to five PIF SHGs, with the goal of full saturation and complete adoption of improved practices and technologies within a community in a relatively short time frame.

3 Experimental design

To establish a causal relationship between the program and changes in outcomes, this study uses a cluster randomized controlled trial (RCT). A cluster design was employed for two reasons. First, group membership is a key component of the program design. Second, PIF effects are anticipated among a second generation of beneficiaries. As described below, we will seek to estimate both direct and second generation effects.

Nepal comprises 75 districts. Districts are further subdivided into village development committees (VDCs), which can be thought of as groupings of villages within a district. Every VDC is split into nine wards, and each ward might include multiple *toles*, or communities. A typical *tole* in the study area has approximately twenty to thirty households; a typical ward has roughly 150 households.

Nepal-based HI staff first identified 60 VDCs in which they had never worked, but that would be good candidates for an asset transfer and training program. Before assigning treatments, HI also identified a central ward and targeted *tole* within the selected central ward for each of the 60 selected VDCs. The expectation was that if assigned to treatment, everyone residing in the targeted *tole* would be targeted by the program, and therefore likely to enroll as a direct benificiary. Through this process, HI pre-identified all targeted beneficiaries (but not necessarily actual beneficiaries) who were later encouraged to form SHGs. Following treatment assignment, these SHGs formed in treated VDCs but not control VDCs. In this way, the individuals in the control arm are directly comparable with those in the treatment arms.

Although indirect PIF effects are expected, we do not anticipate contamination of the control. To an extent, the isolation of rural communities in Nepal provides a natural impediment to such contamination. This is especially true in the Middle Hills (home to about two-thirds of our sample), where lower population density, rugged terrain, poor roads, and inferior cellular connectivity cause communities to be especially cut off. Nevertheless, communities are linked by family and commercial ties. Fewer natural barriers against contamination exist in the Terai, the densely populated plain along the Indian border where about one third of our sample resides. Apart from naturally occurring geographic and social barriers to contamination, we also buffered treated wards from each other and from control VDCs by selecting the 'central' ward within a VDC to be the targeted ward. In this way, we ensure an additional degree of isolation and further reduce the prospect of unintentional spillovers that could bias results.

To improve balance across treatment and control VDCs (and between the various treatment VDCs) we stratified by geography and caste/ethnic composition. First we divided the sample of VDCs into four pools based on district groupings (Hills (2), Middle Hills (1), and Terai (1)). These clusters contained 15, 15, 10, and 20 VDCs respectively. Using administrative data, we then calculated the proportion of residents in each VDC from each of 39 caste/ethnic groups. Within each district grouping we ordered VDCs by the most prevalent caste/ethnic group, then second most prevalent caste/ethnic group, and so on through the ninth most prevalent caste/ethnic group.[2] This created new groups within the district groupings based on rank prevalence of caste/ethnicity. Within these groups, we ordered

[2]Only two of 60 VDCs had more than 9 caste/ethnic groups represented.

VDCs by the proportion of the most prevalent caste/ethnicity, then second most prevalent, and so on. From this ordering we established 16 bins.

Within each stratification bin, we then randomly assigned the 60 VDCs to one of three treatment arms or pure control. All three treatments share some common features. First, HI facilitates the formation of women's SHGs, so all beneficiaries are expected to acquire some level of social capital through group membership and participation. Group members are then encouraged to contribute to group savings accounts with a goal toward increasing financial inclusion. Finally, all beneficiaries are trained on a variety of technical topics including nutrition, home gardening, fodder and forage development, and improved animal management. In addition, all beneficiaries are provided a small amount of cash support for home gardens (approximately \$5) and fodder/forage production (approximately \$10). We'll call these common features the basic intervention.

In order to 'unpack' the benefits of various program components, two additional programmatic elements vary across across treatment arms: a productive asset transfer and additional values-based trainings. The productive asset transfer included two doe goats and cash support (approximately \$40) for goat shed improvement to each individual beneficiary, as well as a shared buck of improved breeding stock for the self-help group. The values-based trainings cover the 12 HI Cornerstones which include: passing the gift; accountability; sharing and caring; sustainability and self-reliance; improved animal management; nutrition and income; gender and focus on the family[3]; genuine need and justice; improving the environment; full participation; training, education, and communication; and spirituality. Notably, the values-based training encourages beneficiaries to "pay it forward" by providing technical training and giving the first-born female offspring of their received livestock to another poor individual in their community.

The treatment arms can be described as follows:

1. *Full Treatment* (FT): basic intervention, values-based training, and a productive asset

[3]Notably, both men and women are encouraged to participate in gender and justice training.

transfer.

2. *No Goats* (NG): Identical to FT, but without the productive asset transfer.

3. *No Values-based Training* (NVT): Identical to FT, but without values-based training.

A fourth arm was randomly selected as pure control. Table 1 summarizes the elements of each treatment arm.

Many of the welfare impacts we consider could be directly affected by either type of training or the asset transfer. For example, women's empowerment could increase as a result of interactions in the group, values-based trainings, technical skills trainings, and/or increased ownership over assets through the transfer. Similarly, income could increase as a direct result of any of these program components in concert or independently. Our experimental design allows us to differentiate between program components.

In addition to variation across treatments, we can also look at differential impacts over time to explore causal channels. Consider income: income could increase early on through improved knowledge about fodder development, livestock nutrition, vaccinations, and breeding, and improved shelters. Any of these factors - a direct result of animal management trainings - might lead to better livestock health of one's current livestock herd, such that the program could increase income in the short run. However, if initial herds are small or if improved practices are concentrated toward the transferred livestock only - then income effects would only be observed in the long run, specifically, after the transferred goat has given birth to a male kid, and once that kid is old enough to be sold.

With these inter-temporal effects in mind, Figure 1 illustrates a timeline of relevant activities. Project implementation began in mid to late 2014 (depending on location) and continued throughout 2015 and 2016. All direct beneficiaries first formed SHGs (shortly after the baseline survey, as described below) and were encouraged to begin saving at this time. Approximately six months later, between March and June 2015, these same direct beneficiaries received livestock if they were assigned to either the FT or NVT treatments. In late 2015 the second generation of beneficiaries, encouraged by direct beneficiaries in their area, enrolled through the "pay-it-forward" program, began to form groups, and participated in the various trainings. Notice that while we know when program activities for these beneficiaries began, it is difficult to know exactly when second generation "pay-it-forward" beneficiaries will receive livestock transfers, because such transfers depend on livestock fertility, which is inherently random. In fact, the program is designed in such a way that the "pay-it-forward" livestock transfer will be staggered, with some receiving livestock transfers within six months of enrolling in the program, while others will wait years before receiving a livestock transfer.

For establishing hypotheses regarding the anticipated timing of impacts, we must carefully consider livestock fertility cycles. We assume a doe can reasonably be impregnated within any given four month window, a five month gestation period, and offspring reach sexual maturity at around seven months (females) or an optimally marketable size at around ten months (males). Depending on breeding cycles and the availability of an improved buck, most direct beneficiaries might be expected to impregnate their does between June and October of 2015, implying the members of a second generation of program goats were typically born near the end of that year and the beginning of 2016. Goats normally experience single births (although multiples aren't uncommon), and the gender of the kid has important implications for impact. The program requires beneficiaries to donate their firstborn female offspring (once it has reached sexual maturity) to another beneficiary through the pay-it-forward mechanism. Alternatively, male kids are expected to be sold on the market.

Taken together these facts imply three noteworthy features of this study, all shown on Figure 1: (1) the earliest pay-it-forward beneficiaries could possibly have received goats was in mid 2016, (2) the earliest possible goat sales (of transferred goat kids) for direct beneficiaries will have taken place in late 2016, and (3) the earliest possible goat sales (of transferred goat kids) for second generation pay-it-forward beneficiaries will take place in early 2018. In section 6, we will return to these features to explain some of our findings.

But first, we link the experimental design to the data collected and discuss our empirical approach.

4 Data

We collected panel household survey data from rural women eligible to participate in the program across three regions of rural Nepal in June-September 2014 and 2016. The main data used for this analysis was collected in June-July 2016, approximately 1.5 years after initial enrollment in the program. Figure 1 shows how the survey timeline fits with program implementation, including surveys planned for future data analysis.

There are two types of respondents in the sample used for this analysis: targeted direct beneficiaries and prospective PIF beneficiaries. Specifically, our sample of targeted direct beneficiaries consists of all households in each of the targeted *toles* (around 25 per ward). In addition, after removing households from the targeted *tole*, we selected a random sample of 15 potential PIF beneficiaries from a complete roster of all households in the central ward. Because of the aggressive nature of the "pay-it-forward" encouragement, we expect that many (if not most) of these households will actually become PIF beneficiaries. Although no intervention took place in control VDCs, sampling in these VDCS occurred in exactly the same manner as in treatment VDCs: 25 individuals from pre-determined targeted *toles*, and 15 individuals from a complete roster of all households in the central ward.

Our total baseline sample is 2,376 women, including 1,286 targeted for direct treatment, and 1,089 households from the central ward likely to receive enter the program through the pay-it-forward mechanism. Shortly after HI delivered training and livestock to the original beneficiaries of the project, a devastating earthquake struck Nepal. The earthquake greatly affected the 10 VDCs belonging to the 'Middle Hills' stratification pool, and were therefore spread evenly across treatment groups and control. We made the decision to drop these from the RCT so that HI could provide earthquake relief in whatever manner they deemed appropriate. Following additional attrition not explicitly related to the earthquake, the remaining sample consists of 50 VDCs and 1,828 households, including 1,031 from targeted *toles* and 797 from the central ward more broadly.

Baseline characteristics and balance

The data includes information on basic household demographics and a variety of outcomes. We group these outcomes across nine dimensions of welfare: women's empowerment, financial inclusion, aspirations, mental health, asset holdings, income, non-food expenditures, physical health, and food security. Details regarding how these outcomes are measured are discussed in sections 5.3 and 6.

We test for balance on observable characteristics between all treated and control groups using the following specification:

$$y_{hvb} = \beta_0 + \beta_1 T^s_{hv} + \varepsilon_{hv} \tag{1}$$

Here, y_{hvb} is a demographic characteristic or outcome of interest for household h residing in village development committee (VDC) v as measured at baseline (b). T^s_{hv} is a binary variable indicating treatment that takes a value of 1 for households assigned any of the three intervention packages and a value of zero otherwise. The superscript s indicates whether a household was categorized as a targeted direct beneficiary ($s = DIR$) or a potential pay-it-forward beneficiary ($s = PIF$). We estimate equation 1 separately for targeted direct beneficiaries and potential pay-it-forwards. ε_{hv} is an idiosyncratic error term clustered at the VDC level. The omitted category is control households. Therefore, β_1 identifies the average difference in baseline demographics and outcomes between any targeted household and control households. In addition to the main specification, we also check for baseline balance between each of the disaggregated treatments and the control group using the following specification:

$$y_{hvm} = \beta_0 + \beta_1 T1^s_{hv} + \beta_2 T2^s_{hv} + \beta_3 T3^s_{hv} + \varepsilon_{hv} \tag{2}$$

We define the outcomes in equation 2 identically to equation 1. The treatment variable from 2 is disaggregated into three separate treatment indicators such that $T1^s_{hv} = 1$ for households slated to receive the full package of benefits ($T1 = FT$), $T2^s_{hv} = 1$ for households selected to receive the no-goats package ($T2 = NG$), and $T3^s_{hv} = 1$ for households offered the no-values-based-training treatment ($T3 = NVT$). As with 1, the omitted category is control households, we cluster ε_{hv} on the VDC level, and we estimate the specification separately for $s = DIR$ and $s = PIF$. Thus, β_1, β_2 and β_3 represent the average differences at baseline between FT-eligible households, NG-eligbile households, NVT-eligible households (respectively) and controls.

We report balance on demographic characteristics in table 2, on caste/ethnicity in table 17, and on levels of outcome summary indices in table 3. Column 1 presents the control means at baseline, column 2 presents the regression coefficients on the treatment variable (T^s_{hv}) from 1. Columns 3-5 contain the regression coefficients on the treatment variables from 2 ($T1^s_{hv}$,$T2^s_{hv}$,$T3^s_{hv}$). In each table panel A presents results for the directly-targeted subset of sample (DIR) compared to direct-target control households, panel B presents presents analogous results for the potential pay-it-forward subsample (PIF).

Overall these balance checks suggest a successful randomization. In cases where imbalance exists for a baseline level of an outcome or a demographic control with respect to any comparisons of interest, we include the baseline level of the imbalanced variables in the econometric specifications described in sections 5.1 and 5.2. Table 2 shows no large or statistically significant differences between the aggregate treatment and controls for DIR households, and only scattered any seemingly non-systematic differences between the disaggregated treatments and controls. Potential pay-it-forwards are approximately two years older than the control mean and own approximately .11 hectares more land. It's worth noting that neither of these difference remain significant at the 10%. after adjusting for multiple inference. We do not expect the age difference to have a material impact on outcomes, and therefore do not control for it in the main econometric specifications. Additional land holdings may impact outcomes, so we control for this in all PIF specifications.

Table 17 presents balance on caste/ethnicity and confirms that our stratification largely succeeded in ensuring a balance of treatments across caste and ethnic lines. The sole exception appears to be that Chheriyas (a subset of high-caste Hindus) appear to be underrepresented in direct-target controls; we therefore include a dummy variable for membership in the Chheriya caste in all specifications for directly-targeted beneficiaries.

Finally, 3 presents the results of balance tests for each of the summary indices. We find no significant differences in the aggregate treatment and controls at baseline for the DIR subset, and no evidence of systematic differences in the PIF. In the disaggregated treatment, however, we do find some evidence of systematically 'worse' baseline outcomes among FT households. The aspirations, financial inclusion, and physical health indices are all lower among DIRs; aspirations and food security are lower among PIFs. To the extent that baseline levels of one indicator affect may effect treatment effects on another, it is appropriate to control for that variable in the treatment effects specifications (e.g., if aspirations affect financial inclusion, one should control for imbalanced aspirations when financial inclusion is the outcome of interest). Where have included such controls we note it in the results presented in section 6.

Attrition

Ignoring the sample purposefully removed following the earthquake, we observed 7.8% attrition (154/1982 households in the post-earthquake sample) between 2014 and 2016. To assess whether the observed attrition is systematic in a way that might bias our results, we employ the same three approaches outlined in the pre-analysis plan of Haushofer and Shapiro (2013), and adapted to our design. For ease of notation, the approach presented below considers the full sample, but we also assess attrition separately for direct beneficiaries and those entering the program through the "pay-it-forward" mechanism.

First, equation 3 estimates whether attrition rates differ across treatment types and control households, where $attrit_{hv}$ is a binary variable indicating that a household was surveyed at baseline but is missing from the endline data set.

$$attrit_{hv} = \beta_0 + \beta_1 T1_{hv} + \beta_1 T2_{hv} + \beta_1 T3_{hv} + \varepsilon_{hv} \tag{3}$$

After estimating equation 3, we do not observe any significant treatment effects on attrition status (table 18). Next, we assess whether attrition rates differ across households with respect to a set of baseline characteristics. To do this we regress a variety of baseline outcomes on attrition status as estimated in equation 4:

$$y_{hv} = \beta_0 + \beta_1 attrit_{hv} + \varepsilon_{hv} \tag{4}$$

We estimate equation 4 for each of the indices (results reported in appendix table 20) as well as a set of demographic variables (appendix table 19). While we do find scattered individual cases where attrition status correlates with a baseline characteristic, these instances do not appear to be systematic or to threaten the integrity of our results. Nevertheless, we err on the side of transparency by reporting treatment effect bounds following Lee (2009), as pre-specified. Finally, equation 5 estimates the extent to which baseline characteristics of treated households differ from control households, after restricting the sample to attrited households:

$$(y_{hvB} | attrit_{hv} = 1) = \beta_0 + \beta_1 T_{hv} + \varepsilon_{hv} \tag{5}$$

In general, we find no noteworthy attrition effects from specification 5 (appendix tables 21 and 22).

5 Estimation

Our main research questions are: (i) what are the short-term welfare impacts of a productive asset transfer and training program? (ii) are all program components necessary for achieving impact? (iii) within a treated village, do treatment effects spillover to subsequent generations of beneficiaries? and (iv) which package of benefits results in the most cost-effective improvements to household and individual well-being in the short-term? We will address questions (i)-(iii) in this and the following section. Section 7 presents our preliminary analysis of question (iv).

Our specific hypotheses, along with detailed plans for handling the data and analysis, are documented in a pre-analysis plan prepared and registered (`http://www.socialscienceregistry.org/trials/1504`) before any analysis took place. In some instances we deviate from this plan, and will specify when this is the case. Notable deviations are discussed in section A of the Appendix.

5.1 Estimation of direct treatment effects

First, we examine program effects for intended direct beneficiaries. For estimating these direct effects, we only include the sample of directly targeted households (both treatment and control), ignoring those targeted to enter the program through the pay-it-forward mechanism. Notationally, this is represented by a superscript DIR.

For this paper, we estimate intent to treat (ITT) effects, noting that the recruitment rates described by Table 4 suggest ITT effects estimated in Equation 6 are likely to be very conservative. We begin with a simple ITT estimation comparing all treatment groups to a common control:[4]

$$y_{hvm} = \beta_0 + \beta_1 T_{hv}^{DIR} + \delta y_{hvb} + \mathbf{X_{hvb}}\gamma + \mathbf{S_{vb}}\rho + \varepsilon_{hv} \tag{6}$$

[4] Equation 6 was not pre-specified in the pre-analysis plan. We include it for a more complete analysis here, noting that its inclusion does not alter our main findings in any way but simplifies exposition.

In this estimation, y_{hvm} is the outcome of interest for household h in village v, measured at midline ($t = m$). The treatment indicator variable (T_{hv}^{DIR}) takes a value of 1 for directly targeted households in wards selected to receive any of the previously described treatments ("treated households"), and a value of 0 otherwise. The counterfactual is households targeted for direct treatment, but located in pure control villages ("control households"). In order to improve statistical power, we condition on baseline ($t = b$) levels of the outcome of interest y_{hvb}, a vector of control variables X_{hvb} for which imbalance at baseline was observed across treatments, and a vector S_{vb} of stratification bin dummies. Finally, ε_{hv} is an idiosyncratic error term. In all specifications we cluster errors at the VDC level, as this is the level of treatment.

In equation 6, β_1 is the average ITT effect across the three treatments. This approach provides a snapshot of program effects, but it does not help to disentangle the impacts due to specific program components. If the marginal impact of particular program components is positive, these average impacts will only tell a partial story. For the full picture, we estimate the intent to treat (ITT) impact for each of the three treatment groups relative to a common control:

$$y_{hvm} = \beta_0 + \beta_1 T1_{hv}^{DIR} + \beta_2 T2_{hv}^{DIR} + \beta_3 T3_{hv}^{DIR} + \delta y_{hvb} + \mathbf{X_{hvb}}\gamma + \mathbf{S_{vb}}\rho + \varepsilon_{hv} \tag{7}$$

In this specification, treatment indicator variables differ by the type of treatment: $T1_{hv}^{DIR}$, $T2_{hv}^{DIR}$, and $T3_{hv}^{DIR}$ each take a value of 1 for directly targeted households in wards selected to receive a particular type of treatment, and a value of 0 otherwise. β_1 represents the ITT treatment effect on households selected to directly receive the full treatment package (T1, FT), β_2 identifies the ITT treatment effect on households selected to directly receive the no-goats package (T2, NG), and β_3 identifies the ITT treatment effect on households selected to directly receive the no-values-based-training treatment package (T3, NVT). As in equation 6, the counterfactual is targeted direct beneficiaries located in pure control areas.

To test whether the treatments effects vary across treatment type, we conduct Wald tests for $\beta_1 = \beta_2$, $\beta_1 = \beta_3$, and $\beta_2 = \beta_3$.

5.2 Estimation of pay-it-forward effects

We also examine the impact of the program on those not directly targeted by the program through its initial recruitment efforts in the community, but eligible to enter the program as a second generation beneficiary through the pay-it-forward (PIF) program component. Superscript PIF thus distinguishes the following from equations 6 and 7, where the samples are quite different. In the analysis of direct treatment effects, the sample consists of those pre-selected for direct benefits (even those in control areas), so that the comparison group is those who would have been targeted for direct benefits had they been selected for treatment. In the analysis of pay-it-forward treatment effects, the sample consists of those pre-selected for possible second-generation benefits (even those in control areas), and the comparison group is those who would have been targeted only as second generation beneficiaries had they been selected for treatment. Following the approach presented in Section 5.1, we estimate β_1, the average ITT PIF impact:

$$y_{hvm} = \beta_0 + \beta_1 T_{hv}^{PIF} + \delta y_{hvb} + \mathbf{X_{hvb}}\gamma + \mathbf{S_{vb}}\rho + \varepsilon_{hv} \tag{8}$$

as well as the differential PIF treatment effects:

$$y_{hvm} = \beta_0 + \beta_1 T1_{hv}^{PIF} + \beta_2 T2_{hv}^{PIF} + \beta_3 T3_{hv}^{PIF} + \delta y_{hvb} + \mathbf{X_{hvb}}\gamma + \mathbf{S_{vb}}\rho + \varepsilon_{hv} \tag{9}$$

In this latter specification, β_1 corresponds to the T1 (FT) ITT effect of being a potential PIF beneficiary in a T1 treatment ward, β_2 captures the T2 (NG) ITT effect of being a potential PIF beneficiary in a T2 treatment ward, and β_3 captures the T3 ITT effect of being a potential PIF beneficiary in a T3 (NVT) treatment ward. We will test for whether the treatments have different PIF effects within wards using Wald tests for $\beta_1 = \beta_2$, $\beta_1 = \beta_3$, and $\beta_2 = \beta_3$.

Notably, pay-it-forward effects could arise through trainings, asset transfers, or through simple observation and replication. If households observe and replicate the behavior of direct beneficiaries, they may benefit indirectly from the HI trainings, even if they don't identify as a second generation program beneficiary. Because this second type of spillover effect is possible, estimation of local average treatment effects (LATE) is not preferable. We therefore estimate ITT effects, keeping in mind that they are likely conservative.

5.3 Defining outcomes and accounting for multiple inference

We have a rich dataset informing numerous hypotheses regarding behavioral change and improved welfare across several dimensions including asset ownership, income, non-food expenditures, financial inclusion, physical health, mental health, aspirations, and women's empowerment. Multiple subindicators exist for each dimension. Subindicators are described and summarized alongside our results in the following section (or in footnotes corresponding to the relevant results table). In almost all cases variables have been coded so that larger values are 'better', therefore positive regression coefficients represent improvements. We will estimate equations 6-9 using each subindicator as a unique outcome of interest y_{hvm}.

These subindicators will also be aggregated into a primary summary index for each dimension of welfare. These summary indices will be employed as our nine primary outcomes of interest. Where no obvious standard index exists in the literature (asset ownership, financial inclusion, physical health, and mental health), we employ Anderson (2008) to create a standardized weighted average of subindicators. For income, the summary index is the logged sum of total household income. Similarly for non-food expenditures, we take the logged sum of total expenditures. For aspirations, we employ the Bernard and Taffesse (2014) aspirations index. For women's empowerment we employ indicators from the Five Domains of Empowerment (5DE) of the Abbreviated Women's Empowerment in Agriculture Index (A-WEAI) modified to the local context (Alkire et al., 2012; Malapit et al., 2015).

We follow the emerging standard in the program evaluation literature by accounting for multiple hypotheses in two ways. First, the summary index for each welfare dimension consolidates several individual tests into a single test. Second, because we still have multiple outcome dimensions, we report naive p-values and adjusted q-values that control for the false discovery rate (FDR). Specifically, we calculate q-values for multiple hypothesis tests across summary indices, but not across treatments, using the Benjamini and Hochberg (1995) step-up method outlined in Anderson (2008) and applied by Banerjee et al. (2015). To test treatment groups against each other we will conduct Wald tests as described in section 6. For these tests, we will report both naive p-values and q-values that control for FDR. As above, we calculate q-values for a specific hypothesis test across summary indices (and interaction terms when applicable), using the Benjamini and Hochberg (1995) step-up method. When estimating treatment effects on sub-indicators (rather than summary indices) we will report naive p-values. We test for the impact on sub-indicators primarily to identify the mechanism behind impact (or lack thereof) observed for the summary indices. We therefore consider this analysis exploratory, and take a less stringent approach to hypothesis testing.

We prefer controlling for FDR over controlling for the the family-wise error rate (FWER) because we are testing several hypotheses (even after condensing them to summary indices), and FWER adjustments become increasingly severe as the number of tests grow (Anderson, 2008; Benjamini and Hochberg, 1995). Since overall conclusions about program effectiveness depend on many outcomes, the overall conclusion should not be that the intervention is ineffective because of one erroneously rejected null hypothesis. Therefore, it seems reasonable to be more tolerant of Type I error in exchange for greater power. The FDR formalizes this tradeoff between Type I and Type II error (see Benjamini and Hochberg (1995) and Anderson (2008) for a more detailed discussion).

6 Results

6.1 Recruitment rates

Before turning to our empirical strategy for estimating impacts, it is useful to look at participation rates, as shown in Table 4. Across all three treatments, participation of targeted direct beneficiaries was similar at around 65%. In the NG and NVT groups participation rates were slightly lower, although not significantly so, so we do not read much into those differences. Pay-it-forward beneficiary recruitment rates display a greater degree of heterogeneity across treatment types. In the full treatment wards, 67% of potential pay-it-forward households claimed to be actual HI beneficiaries. This indicates that the pay-it-forward aspect of Heifer's program successfully recruits at rates identical to the level achieved for direct beneficiaries. However, in the NG treatment arm, recruitment of PIF beneficiary households was 20% lower at 54%. Comparing these rates to direct beneficiaries across the same two treatments may indicate lower enthusiasm among direct beneficiaries in the absence of the asset transfer - which led to less successful recruitment of second generation beneficiaries. Finally, in the NVT treatment arm, pay-it-forward beneficiary recruitment is much lower at only 14%. This difference suggests that the encouragement to pay-it-forward is vital to the self-propagating nature of the HI intervention. This may explain why studies of other similar programs have not observed spillover effects. Without incorporating spillovers into the program design, others are much less likely to participate.

6.2 Short-run welfare impacts

Our main results are presented in table 5. Panel A shows ITT direct effects for our nine summary indices. Column 1 contains control group means, column 2 contains ITT estimates for the aggregated treatment groups (β_1 in equation 6), and columns 3-5 contain ITT estimates for the three disaggregated treatment groups (β_1, β_2, and β_3 in equation 7). Finally, columns 6-8 presents Wald tests for equal treatment effects. Panel B presents analogous pay-it-forward impacts.

Beginning with the direct impacts shown in 5, we first consider the average direct effect of being in any treatment group (column 2). We find significant impacts on financial inclusion (0.31 SD) and women's empowerment (0.24 SD). The effects on both are still significant to the 0.1 confidence level after controlling for the false discovery rate (FDR). We'll unpack the mechanisms behind these effects when we look at the subindicators below.

We do not observe changes in asset holdings, income or non-food expenditures for direct beneficiaries of the program. Recall that the size of transfer is much smaller than other similar programs (such as BRAC's graduation program evaluated by Banerjee et al. (2015)) and the time spent in the program is also shorter. It seems the asset transfer alone is not large enough to significantly alter total asset holdings (which includes livestock), although we will unpack this further in the subsection below. Of course, total assets are also expected to increase in value through goat production. However, as we note in section 3, if the firstborn goat kid was a female, it's likely the kid was given away as part of the pay-it-forward program component around the time of data collection. If it was given away prior to the survey, the short run return on the asset transfer will be zero. If instead the firstborn kid of the transferred goat (i.e. not a kid of one's pre-existing herd) were male, it would not be ready for sale until late 2016, after the data was collected for this analysis. In this latter case the asset value would increase by the size of the transfer plus the goat kid, but we would not expect any increase in income in this short time frame, at least through this direct channel. Income could increase in the short run through other mechanisms, but these results suggest this is not the case.

We also do not observe changes in physical health, mental health or aspirations, at least in terms of the summary indices. We'll unpack these findings (or lack thereof) in greater detail in the subsection below.

Across welfare dimensions, we do not observe statistically significant differences in outcomes across treatments. This suggests either that our analysis could not capture small differences between treatments or that the combination of activities is not critical for increased empowerment or financial inclusion in the short run. It is too early to say if the different program components are important for improving long-run economic outcomes.

Panel B of table 5 summarizes ITT pay-it-forward impacts. As with the direct effects, we see significant increases in the financial inclusion index and the empowerment index. These results are impressive given the relatively short time horizon in which to observe pay-it-forward impacts. The vast majority of PIF beneficiaries had received extensive training, but most had not yet received an asset transfer. These results remain significant after controlling for FDR at the 0.1 confidence level. Notably, we do not observe pay-it-forward impacts when the values-based trainings that encourage paying it forward are withheld.

These findings are consistent with a narrative that, a little more than one year after the intervention, program beneficiaries show improved levels of welfare outcomes that we might reasonably expect to respond to treatment in the short-term. Financial inclusion and empowerment may lay the groundwork for the ultimate intended program outcomes of improved asset holdings, and income, and subsequent analysis will test for whether these long run impacts are actually observed. Furthermore, the pay-it-forward program component seems to rapidly increase the number of households benefitting from the program.

6.3 Exploratory analysis of subindicators and mechanisms

In this section we carefully examine the components of each index to develop a clearer picture of the driving forces behind the behavior of each index, and to get a sense of potential trajectories of long-term welfare outcomes. Each summary index will be decomposed in separate tables for direct and pay-it-forward impacts. As with the summary indices, column 1 contains control group means, column 2 contains ITT estimates for the aggregated treatment groups, columns 3-5 contain ITT estimates for the three disaggregated treatment groups, and columns 6-8 present results of a Wald test for differential treatment effects. Through this exploratory analysis, we seek to gain insight into the behavior of the outcome indices by an exploratory analysis of the sub-index outcomes. Where significant treatment effects on indices exist, we look to the index components to isolate the driving forces behind the effect. Similarly, where a null effect exists or where large standard errors mute a potentially significant effect, we can examine components to glean a better understanding of the index behavior.

Assets

The asset index is a weighted standardized average of a productive and non-productive asset index, livestock, land, and a housing quality index. The productive asset index is a principle components index of productive assets (including draft animals, plows, grinder, thresher, loom, sewing inventories, and mechanical tools). The non-productive asset index is a principle components index of non-productive assets (including radios, cassette recorders, DVD players, televisions, landline telephones, mobile phones, heater/pressure lamps, electric fans, camera/camcorders, computers, furniture, rugs, clocks, jewelry and watches). Livestock represents owned goats, cattle, water buffalo, swine and chickens aggregated into tropical livestock units (TLUs). Land is measured by total hectares owned. The housing index is a principle components index of indicators of housing quality (roof material, wall material, flooring material, number of rooms/rooms per person, has electricity, kitchen attached/detached, type of cooking stove and fuel, type of toilet, source of drinking water).

We observe no significant effect on the assets summary index, and similarly for productive and non-productive asset indices as well as the housing index. However, we do observe sensible (yet in some cases statistically insignificant) dynamics with respect to livestock and land as reported in table 10. Among both direct and PIF beneficiaries livestock ownership increased by a statistically insignificant 0.208 tropical livestock units (TLUs) on average across treatments. However, among direct and PIF beneficiaries alike, we observe point estimates near zero for the NG treatment effect on livestock holdings. Despite extensive trainings on livestock management, these beneficiaries do not choose to invest more in livestock. Alternatively, direct and PIF beneficiaries who were eligible to receive two free goats under the FT and NVT treatments increase herds by 0.2-0.4 TLUs. Although statistically insignificant, a 0.2 increase in TLUs equals an addition of two goats. Therefore, these point estimates are broadly consistent with the magnitude of the transfers received by FT and NVT beneficiaries, plus a reasonable herd growth rate. Perhaps surprisingly, goat herds have grown at roughly the same rate among intended direct beneficiaries as they have among PIF beneficiaries.

We also see statistically significant increases in landholdings of about 0.05-0.1 hectares (compared to a control mean of 0.47 hectares) for direct beneficiaries belonging to the FT and NVT treatment groups. We do not see an increase for the NG group. This may indicate that households are investing in land to grow fodder for their livestock; field staff report that this is the case. Field staff also report that households often purchase land with additional income generated from raising goats or other businesses. We note, however, that we do not observe increases in total income at this time.

Income

Table 11 reports sub-indicator outcomes for treatment effects on income. Total household income includes livestock income (from sale of livestock and/or livestock products), crop income, permanent salaried income (including salary as a government or education officer, health care provider, NGO or private firm employee, etc.), small business income (including earnings from a small shop, construction business, flea market sales, petty trading, selling cooked food or alcohol, barber, tailor, teashop, etc.) income earned as a day laborer (farm labor, temporary off-farm jobs, driving, construction, housemaid, livestock care, etc.) and other miscellaneous income (including cash transfers, gifts, remittances, pension earnings, lottery winnings, dowries, inheritances, etc.) We find no significant increase in total income, and the point estimates are very small.

While the intervention targets livestock income in particular, we observe no evidence of an increase in livestock income over this period among direct beneficiaries across all treatment types. Surprisingly, we do observe an increase in livestock income for PIF beneficiaries. This disconnect between direct and PIF impacts (also observed for impacts on livestock) highlights the importance of treating our subindicator analysis as exploratory given the wide range of hypotheses being tested. With that in mind, we note a statistically significant increase in business income for direct and PIF NG beneficiaries and a statistically significant decrease in crop income for direct FT beneficiaries. It's possible that NG beneficiaries increased entrepreneurial capacity through the program, which they applied to non-livestock business activities, thereby increasing business income. Similarly, FT beneficiaries may have removed some land from cash crop cultivation in favor of growing fodder for livestock production.

Non-food Expenditures

Table 12 reports impacts on non-food expenditures. Although there is no evidence of a treatment effect on overall expenditures, there is some evidence that participation in the program decreases miscellaneous expenditures, which include some necessities (transportation costs, rent, housing materials) and some expenses that might be considered luxuries (festivals, ceremonies, weddings, funerals, gifts/donations, jewelry, kitchen equipment). This effect is strongest for those in the NG treatment, and somewhat stronger for PIF beneficiaries than direct beneficiaries.[5] We observe no impact on medical expenditures (including medicines and medical supplies, medical consultation and treatment fees, laboratory and diagnostic fees, visits to traditional healers, etc.) or clothing expenditures (including adult womens clothing, childrens clothing including school uniforms and shoes, materials and tailoring items for clothing made at home).

Financial Inclusion

We observe a large 0.31 SD and statistically significant impact on financial inclusion for direct beneficiaries. To interpret the magnitude of this impact, we can consider what it means for a typical direct beneficiary with a median level of financial inclusion as measured by our index at baseline. Access to the program moves this median individual to the 61st

[5]Although not statistically significant, this latter differential impact between direct and PIF beneficiaries could possibly be explained by the costs incurred by direct beneficiaries for a "pay-it-forward ceremony" associated with the PIF program component that took place in many program areas near the time of survey.

percentile in the distribution of financial inclusion indices. Table 7, panel A shows this effect is primarily driven by saving. Membership in a savings group is a requirement of the basic intervention, and we observe membership rates increasing by 17 percentage points on average across treatment types. But Table 7 shows this effect goes beyond simple membership. Those receiving the values-based training (FT and NG treatments) dramatically increase savings by 74-106%, whereas the effect is not significant for NVT direct beneficiaries. Although encouragement to save is included in the basic intervention, it seems the values-based training may have increased the effectiveness of the savings encouragement possibly through acquired social capital among group members or additional training on income management (one of the HI Cornerstones).

Although not significant, the signs on the coefficient for the amount owed to formal lenders (including a bank, development bank, cooperative, finance company, microfinance institution, or savings/credit group) and the amount owed to informal lenders (including a family member, friend, debt to a local shop, or village money lender) suggests a potential shift toward accessing formal credit markets. A statistically significant decrease of 2.3 percent in the discount rate of beneficiaries receiving the full treatment suggests an increased value on future well-being. The sign on the coefficient estimating the impact on the discount rate is consistent across treatments, but not statistically significant for NG or NVT beneficiaries. Counter to this finding, however, we observe no increase in the length of time individuals plan ahead.[6]

Shifting to PIF impacts, panel B of table 7 reveals a statistically significant 0.21 standard deviation increase in financial inclusion for PIF beneficiaries on average across treatments. This implies a typical PIF beneficiary (with a median level of financial inclusion at baseline) moves to the 58th percentile in the distribution of financial inclusion indices. One main reason for the smaller average impact is because NVT direct beneficiaries fail to pay-it-forward. Impacts in the FT and NG PIF groups are 0.22-0.34 standard deviations,

[6]The "planning horizon" variable is an ordered categorical variable following Laajaj (2017) indicating how far individuals plan ahead where (0) = Do not plan ahead, (1) = plan ahead one week, (2) = plan ahead one month, and (3) = plan ahead 6 months.

and not statistically different from each other, whereas no impact is observed among NVT PIF beneficiaries. Subindicator analysis tells a similar story to direct impacts, savings group membership rates increase by 13-19% while the amount saved increases by 9-68% among FT and NG PIF beneficiaries respectively. No impact on savings-related outcomes are observed among NVT beneficiaries. The values-based training appears to be very important for savings group formation and functionality for second generation beneficiaries.

Food Security We observe no treatment effects on the food security index, a weighted standard average of several variables related to meal-skipping and dietary diversity, for either direct of indirect beneficiaries (table 14). Nepali families traditionally eat two main meals (*khana*) per day, with a lighter meal (*khaja*) in between, a paradigm that is clearly visible in the control means of the "meals" and "snacks" variables. We do not see program effect on number of meals in either the aggregate or the disaggregate treatments, and these null effects are largely confirmed by similar null findings for all the meal-skipping variables. Based on discussions with Nepali team members, these results are unsurprising; food-insecure Nepali households may eat smaller, less-diverse, or lower quality meals, but the rarely skip meals altogether. Therefore, we also examine modified versions of the Dietary Diversity Index (DDI), the Household Dietary Diversity Score (HDDS), and the Food Consumption Score (FCS). The DDI is a simple count of the number of distinct food items consumed in the household over the three day period that preceded the survey, HDDS resembles DDI but groups similar food items into categories, and the FCS improves upon the HDDS by applying frequency and quality weights to the categories. We find no statistically significant effects on any of the dietary diversity outcomes in the aggregate treatment for either direct or indirect beneficiaries. Among direct beneficiaries there are two instances of statistically significant negative treatment effects, but neither of these treatment effects are "large" in the sense that they would move a sizeable fraction of the sample into a food-insecure state.

Physical Health

We see no overall effect on the physical health summary index for direct beneficiaries.

When we break it down into its components (table 13, panel A) we do see a significant effect on respondents' subjective opinion of their children's health (around 0.4 points on a ten point scale) for the NG treatment, the NVT treatment, and the aggregate of treated households. We interpret this result with caution given that those receiving benefits may tend to respond more favorably to highly subjective questions like this than those not receiving benefits. We find nothing noteworthy with respect of the physical health of PIF beneficiaries, either for the summary index or for subindicators (table 13, panel B).

Mental Health

We also see no increase in the mental health index, and point estimates are close to zero. We do find some significant effects on individual components of the index (table 9), which includes a measure of depression, locus of control, optimism, life satisfaction, self esteem, happiness, worries, and trust.[7] Most notably, life satisfaction increases by 0.3 points on a ten point scale for households in any treatment. The point estimate for this effect is positive for the distinct treatments, but only significant (and larger) for those receiving the NG package. Self-esteem increases by 0.3 points on an 18 point scale for the aggregated treatments, FT, and NVT. We also find an increase in the worry score (which indicates *less* worrying) for the full treatment group only (0.5 points on a 16 point scale). Given the treatment and the effects we find in other areas, we are surprised at the lack of overall effect on mental health.

As was the case for directly targeted households, we see no significant change in the mental health summary index for potential pay-it-forward households (table 9, panel B). The observed direct impacts on life satisfaction and self esteem do not exist among PIF beneficiaries. We do observe point estimates of treatment effects on the worries score that are slightly stronger among NVT PIF beneficiaries than those for NVT direct beneficiaries,

[7]Depression is an abbreviated version of the CES-D scale Radloff (1977) with a high value indicating high levels of depression. Locus of control is an abbreviated Rotter (1966) scale where a high value indicates a stronger internal locus of control. Worries questions employ the Penn. State worries questionnaire, and a high value indicates *less* worried. Remaining subindicators are based on aggregated responses to 3-4 questions per subindicator from the 2009 World Values Survey wor (2009), where high values indicate positive welfare (high optimism, more satisfied, high self esteem, happy, and more trusting). A more detailed description of how subindicators are calculated is provided in the pre-analysis plan with a list of specific questions available upon request.

yielding statistically significant aggregate impacts.

Aspirations

A stated goal of the intervention is to alter individual aspirations and increase hope of a better future. We observe no change in the Bernard and Taffesse (2014) aspirations index on average or for individual treatments (table 8). To construct the index, the survey asks about aspirations across income, assets, social status and education, where each category is assigned a weight by the respondent.[8]

Notably, we do observe an increase in (logged) income aspirations, the category of aspirations most directly targeted by the intervention, for direct beneficiaries (panel A). The impact is not statistically distinguishable across treatments (although it is not statistically significant for NVT beneficiaries), and dampened for PIF households. In a separate analysis focused on the formation and failure of aspirations (Janzen et al., 2017), we argue that income aspirations are a better measure of financially-related aspirations than asset aspirations in this context, given the way the survey questions were written. The latter asks only about the "value of home and land" which is an incomplete measure of assets and wealth in this context, and likely leads to substantial measurement error. We do not observe an impact on aspirations regarding social status, although the coefficient estimates are consistently positive. Here again, measurement error is a concern; the survey question used to quantify aspirations in this category asks, "In the future, how many women in this ward do you think might actually seek your advice?" Feedback from the survey team raises concerns that this is not an appropriate or comprehensible measure for the local context. We also do not observe any impact on aspirations for education (measured by the response to "What level of

[8]Following Bernard and Taffesse (2014), the questions related to income aspirations are: (1) "What is the maximum level of income that a person in your community might expect to earn in a year?" (2) "What is the minimum level of income that a person in your community might expect to earn in a year?" (3) "What is your present personal level of income?" and (4) 'What level of yearly income do you personally think you might be able to achieve in the future?" The first two questions are intended to help respondents delineate a realistic range before stating their own current status and their aspirations. The third question records the personal status for that dimension. The fourth question is interpreted as the individual's aspiration. A similar series of questions are asked to assess aspirations in the other three categories. Weights are assigned by respondents through the distribution of 20 tokens across four bins in proportion to how heavily they value a particular category.

education would you like your children to achieve" in general, and disaggregated by gender) of one's children, but we note the program does not directly target improvements in child well-being or education.

Empowerment

An important short-term goal of the program is to increase women's empowerment. We employ indicators (modified to the local context) from the Five Domains of Empowerment (5DE) of the Abbreviated Women's Empowerment in Agriculture Index (A-WEAI) to calculate an empowerment score for all women in the sample (Alkire et al., 2012; Malapit et al., 2015). The index aggregates an empowerment score across decisions about production, access to and decision-making power over productive resources, control over credit and income, leadership in the community and time allocation. Each binary subindicator equals one if the respondent achieves adequate empowerment in that domain, and zero otherwise. Weights and definitions of adequacy are based on the A-WEAI, but adjusted to the local context.[9]

Using this measure, we see an increase of 0.24 SD in empowerment among direct beneficiaries on average across treatments (table 6, panel A). This implies a typical direct beneficiary with a median level of empowerment at baseline moves to the 65th percentile in the distribution of empowerment scores. Looking across these five domains of the A-WEAI, we find a statistically significant average impact in three subindicators. Across treatment types, compared to women in the control group, direct beneficiaries are on average 4.3 percentage points more likely to own productive assets, 4.6 percentage points more likely to have some control over use of income and 15.6 percentage points more likely to belong to a

[9] Specifically, we use the following weights and definitions of adequacy: A respondent is adequately empowered in production decisions if she has at least some input into at least one production decision (weight = 1/5). Adequate ownership means the household owns at least one asset, and the respondent (individually) has at least some ownership of one asset (weight = 2/15). A respondent is adequately empowered in access to and control over credit if the household has at least some credit and the respondent participated to any extent in the decision to borrow (weight = 1/15). Adequacy in control over income means conditional on the household participating in an income-generating activity or expenditure, the respondent participates in decisions regarding at least one non-essential activity or expenditure (weight = 1/5). A respondent is adequately empowered in group membership if she is a member of any group (weight = 1/5). A respondent is adequately empowered for time use if she worked 10.5 hours or less in the previous 24 hours (weight = 1/5).

group. Although the majority of women (65 percent) are already in some kind of group at baseline (for example, a mother's group or savings group), this latter impact demonstrates the importance of group formation for a significant number of women (16%) for whom social capital gains have the most potential.

Since livestock rearing is labor-intensive, we might be concerned the intervention increases time spent working in a way that harms welfare. The A-WEAI uses a detailed accounting of time use to calculate the number of hours spent working (as opposed to leisure or sleeping) on a typical day. Women are considered adequately empowered if they work less than 10.5 hours per day. Table 6 shows no impact (positive or negative) of the intervention in this domain. A more detailed analysis of impacts on time allocated to various activities (reported in Table XX of the Appendix) also reveals no impact. In the short-term, there is no evidence that women work more as a result of the intervention, or even that they spend more time on livestock-rearing.

These results are fairly consistent across treatments. Notably, beneficiaries in the NG treatment arm did not receive livestock and yet they exhibit increased asset ownership of equal magnitude to other treatments, illustrating how this impact is not driven exclusively through the asset transfer. Interestingly, empowerment over production decisions does not increase as much (or significantly) in the NG treatment group, whereas control over income only increases significantly in the NG treatment group, and by a larger amount than other treatments.

With the exception of NVT beneficiaries (who show no increase using the summary index), the intervention increases empowerment among PIF beneficiaries in a similar manner as it does for direct beneficiaries (table 6, panel B). If anything, the results are larger and statistically stronger. For PIF beneficiaries, the empowerment index moves 0.29 SD on average, with demonstrated increases in input regarding production decisions, asset ownership, control over income and group membership.

7 Cost-benefit analysis

Cost per beneficiary in our sample varies by treatment arm as well as direct/pay-it-forward status. Some program costs are common or shared across all treatments. All three treatment types receive the same human capital and technical trainings, for instance, and much of the NGO overhead and administrative expenses are spread evenly across all SHGs regardless of treatment status. Other costs are not incurred at all in certain treatment arms: the NG treatment arm incurs no costs for livestock, while NVT incurs no costs for values-based trainings. In addition to differential costs across treatment arms, recall that direct beneficiaries take on the responsibility for paying it forward - sharing livestock and knowledge to second generation beneficiaries. While HI does conduct some trainings directly, and does provide a limited amount of backstopping to ensure the quality and completeness of the pay-it-forward process, direct beneficiaries shoulder much of the costs associated with pay-it-forward beneficiaries. Therefore, any analysis of the costs of the treatment effects presented here must take into account both dimensions for heterogeneous costs.

We collected detailed cost data on all program activities in each treatment arm, and these amounts can be differentiated between direct and pay-it-forward beneficiaries with a few reasonable assumptions. We present costs per beneficiary broken down by treatment arm and direct/pay-it-forward status in table 1.[10] The table is organized into three panels. We report pooled costs associated with direct and pay-it-forward beneficiaries in the top panel. We tabulate costs associated exclusively with direct beneficiaries in the middle panel. The bottom panel includes HI's costs for provision of benefits to pay-it-forward beneficiaries; these costs would not have been occurred but-for the existence of the pay-it-forward program component. The right-most column aggregates across treatment arms.

Before any disaggregation, we calculate that Heifer spent approximately $137 USD per beneficiary on average. Without considering the distinction between direct and PIF benefi-

[10]The appendix presents full cost data and details the assumptions and methodology used to arrive at the summarized figures presented in table 15 NEED TO UPDATE THIS.

ciaries, the data indicate that FT beneficiaries cost about $118 USD each, NG beneficiaries cost $87, and NVT beneficiaries cost $550. NVT beneficiaries cost so much because the direct NVT members do not pay it forward, and therefore do not take advantage of the economies of scale acquired in the exponential recruiting model. These economies of scale come into sharper focus in the bottom two panels. Costs per direct beneficiary are high relative to the aggregate, ranging from $270 to $465 depending on treatment type. With direct beneficiaries in place, however, treating additional beneficiaries implies decreasing marginal costs: the average PIF beneficiary in our sample costs only $65 USD.

These cost data enter a discussion of effectiveness in two ways, the first (type I cost effectiveness) relates to internal program design and the second (type II) relates to external comparisons with related interventions. First, because we find negligible variation in treatment effects across treatment arms, it stands to reason that HI would be most interested in providing beneficiaries with the lowest-cost combination of benefits that achieves desired outcomes. Second, with the lowest-cost combination of benefits determined, we are interested in how the costs and benefits of the intervention compare to those of comparable interventions, especially other productive asset transfers and cash transfers.

With respect to type I cost effectiveness, the only conclusion that we might make at this point is also the most obvious: values based training helps recruit, retain, and transfer assets and skills to PIF beneficiaries. The two indices where we observe statistically significant program benefits are not in the money-metric, so it's difficult to interpret a cost/benefit ratio. Given the timeline of program activities and goat herd dynamics, we believe it may be too early to assess relative cost effectiveness of the remaining treatment arms (FT and NG), even though per beneficiary cost is lower under the NG regime and the recruitment data suggest that the promise of livestock does not increase uptake. Since it is possible that we will observe differential effects for indices like income, assets, and expenditures (which are in the money metric) in the long run, we will report on relative cost effectiveness in a later analysis.

8 Conclusions

In this study we evaluate the short-term impacts of Heifer International's livestock transfer and training program in Nepal using a randomized controlled trial. We find that in just over one year women beneficiaries are more empowered and connected to financial markets. Exploratory analysis suggests these findings stem from increased saving and membership in savings (or other) groups, increased ownership of productive assets, and increased control over use of income. These impacts are observed not only among households who received livestock and/or training directly from the program, but also for those brought into the program through encouragement to "pay it forward," where other women in the same village are recruited, trained and eventually given livestock by initial Heifer beneficiaries. These findings, combined with the cost-benefit analysis presented in section 7, demonstrate how encouragement to "pay-it-forward" can help achieve a broader impact, at least in terms of women's empowerment and improved access to finance, at low cost.

Our exploratory analysis reveals additional positive impacts that are worth highlighting. We observe no impact on our summary measure of mental health, but direct beneficiaries have higher life satisfaction and self esteem, and there is some evidence the complete multifaceted intervention reduces the time beneficiaries spend worrying. We also present evidence of increased aspirations for future income among direct beneficiaries.

We do not observe statistically significant changes to the longer-run outcomes of income, assets (although herds do increase by approximately 2 goats as expected), and expenditures, but it may be too early to observe these effects. In future work we will measure the strength and persistence of program impacts and the cost effectiveness of the program disaggregated across program components. Measuring the strength and persistence of effects is crucial to understanding the full program impacts.

Although this planned future work is important, we think the current analysis makes an important contribution of its own. Evaluating short-term impacts is important for understanding causal mechanisms. This paper demonstrates how a multifaceted social protection program that combines trainings with an asset transfer has immediate effects. However, the paper also provides suggestive evidence that not all program components are necessary for achieving short-term impacts. Similar impacts are observed among direct beneficiaries who were not allocated an asset transfer or who didn't receive the program's "values-based trainings." Although we cannot yet say if either program component is necessary for achieving long-run impacts (including the important long-run outcomes of income, assets and expenditures), we can say confidently that the basic intervention is sufficient for achieving financial inclusion or empowerment in the short-run. This finding supports evidence provided by Karlan et al. (2017) demonstrating women savings groups in Ghana, Malawi and Uganda increase women's empowerment and financial inclusion over a similar time period, even if they don't affect income, assets, or expenditures.

That being said, these supplemental program components (beyond the basic intervention) may be important for achieving a broader impact through "paying-it-forward," as both components increase recruitment of second generation beneficiaries. The HI values-based training, which explicitly encourages beneficiaries to "pay-it-forward," seems particularly critical for the transmission of PIF impacts. Although "pay-it-forward" is a well-known concept - particularly popular during the holiday season in developed countries with widely publicized examples of paying for a stranger's coffee or leaving an unfathomably large tip at a restaurant - it is rarely incorporated into the design of social protection programs. Yet it could be, and our analysis suggests this unique program component could be an important and cost-effective tool for achieving program goals.

References

2009. *World Values Survey 1981-2008 official aggregate*.

Alkire, S., R. Meinzen-Dick, A. Peterman, A. Quisumbing, and G. Seymour. 2012. "A. Vaz. 2012."The Women's Empowerment in Agriculture Index," Poverty, Health & Nutrition

Division, International Food Policy Research Institute." Working paper, IFPRI Discussion Paper.

Anderson, M.L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, pp. .

Ashraf, N., D. Karlan, and W. Yin. 2006. "Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines." *The Quarterly Journal of Economics*, pp. 635–672.

Bandiera, O., R. Burgess, N. Das, S. Gulesci, I. Rasul, and M. Sulaiman. 2017. "Labor markets and poverty in village economies."

Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Parienté, 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.

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. Series B (Methodological)*, pp. 289–300.

Bernard, T., and A.S. Taffesse. 2014. "Aspirations: An approach to measurement with validation using Ethiopian data." *Journal of African Economies* 23:189–224.

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.

Haushofer, J., and J. Shapiro. 2013. "Welfare Effects of Unconditional Cash Transfers: Pre-analysis Plan." Unpublished.

Janzen, S.A., N.P. Magnan, S. Sharma, and W.M. Thompson. 2017. "Aspirations failure and formation in rural Nepal." *Journal of Economic Behavior and Organization* 139:1–25.

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.

Karlan, D., B. Savonitto, B. Thuysbaert, and C. Udry. 2017. "Impact of savings groups on the lives of the poor." *Proceedings of the National Academy of Sciences* 114:3079–3084.

Laajaj, R. 2017. "Endogenous time horizon and behavioral poverty trap: Theory and evidence from Mozambique." *Journal of Development Economics* 127:187–208.

Lee, D.S. 2009. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *The Review of Economic Studies* 76:1071–1102.

Malapit, H., C. Kovarik, K. Sproule, R. Meinzen-Dick, and A. Quisumbing. 2015. "Instructional Guide on the Abbreviated Women's Empowerment in Agriculture Index (A-WEAI)." Unpublished.

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." *Food and nutrition bulletin* 35:312–326.

Radloff, L.S. 1977. "The CES-D scale a self-report depression scale for research in the general population." *Applied psychological measurement* 1:385–401.

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.

Rotter, J.B. 1966. "Generalized expectancies for internal versus external control of reinforcement." *Psychological monographs: General and applied* 80:1.

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.

Table 1: Description of program components by treatment arm

Description of Program Components	**T1 (FT)**	**T2 (NG)**	**T3 (NVT)**
Basic intervention	x	x	x
SHG formation			
SHG savings encouragement			
training on nutrition			
training on improved animal management			
training and cash support ($5) for home gardening			
training and cash support ($10) for fodder & forage production			
access to community animal health worker			
Productive asset transfer	x		x
2 doe goats			
cash support ($40) for goat shed improvement			
1 shared buck of improved breeding stock (per SHG)			
Values-based trainings	x	x	
encouragement to "pay-it-forward"			
training on SHG management			
training on gender and justice			
*training on remaining HI Cornerstones**			

*The remaining HI Cornerstones not noted elsewhere in this table include: accountability; sharing and caring; sustainability and self-reliance; improving the environment; income; full participation; training, education, and communication; and spirituality.

Figure 1: Study timeline

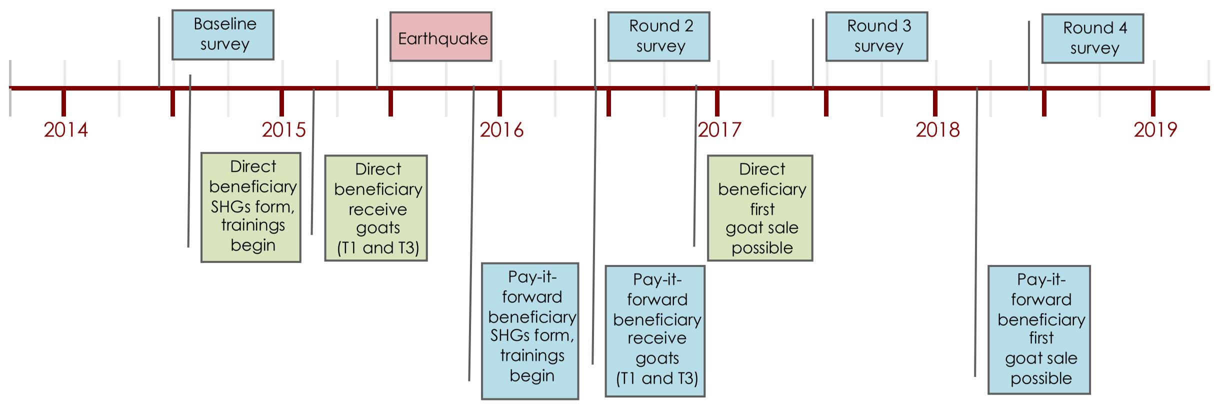

Table 2: Baseline balance: demographic

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
HH Size	5.945	-0.206	-0.414*	-0.198	-0.029	-0.216	-0.385**	-0.169	1,108
	(2.744)	(0.217)	(0.237)	(0.319)	(0.222)	(0.469)	(0.045)	(0.555)	
N Male	2.975	-0.038	-0.168	0.039	0.004	-0.208	-0.172**	0.036	1,087
	(1.553)	(0.107)	(0.108)	(0.152)	(0.114)	(0.118)	(0.045)	(0.791)	
Average Age	31.138	-0.875	-0.004	0.026	-2.523*	-0.030	2.519**	2.549	1,108
	(10.875)	(1.305)	(1.436)	(1.845)	(1.379)	(0.986)	(0.040)	(0.132)	
Resp. Age	41.759	-1.550	-0.661	-1.196	-2.683	0.535	2.022	1.486	1,108
	(13.498)	(1.331)	(1.440)	(1.971)	(1.729)	(0.783)	(0.235)	(0.493)	
Resp. Edu	2.887	0.015	-0.406	0.550	-0.128	-0.957*	-0.278	0.678	1,108
	(4.029)	(0.582)	(0.603)	(0.689)	(0.653)	(0.067)	(0.549)	(0.239)	
Resp. Lit.	0.533	0.023	-0.025	0.070	0.020	-0.095	-0.045	0.050	1,108
	(0.500)	(0.079)	(0.094)	(0.079)	(0.100)	(0.172)	(0.629)	(0.514)	
Income	11.064	0.077	-0.031	0.258	-0.003	-0.289	-0.028	0.261	1,109
	(1.371)	(0.175)	(0.215)	(0.224)	(0.190)	(0.201)	(0.882)	(0.196)	
Land	0.494	-0.048	-0.045	-0.030	-0.069	-0.015	0.024	0.039	1,107
	(0.664)	(0.070)	(0.083)	(0.079)	(0.088)	(0.838)	(0.770)	(0.618)	
TLU	2.465	-0.009	-0.108	-0.237	0.298	0.129	-0.406	-0.535	1,109
	(2.179)	(0.315)	(0.402)	(0.358)	(0.389)	(0.734)	(0.323)	(0.148)	
Has Migrant	0.588	0.050	0.027	0.059	0.063	-0.032	-0.036	-0.004	1,108
	(0.493)	(0.048)	(0.065)	(0.050)	(0.064)	(0.584)	(0.609)	(0.942)	

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
HH Size	5.968	-0.064	-0.529	0.119	0.194	-0.648*	-0.723**	-0.075	873
	(2.713)	(0.292)	(0.320)	(0.368)	(0.339)	(0.054)	(0.018)	(0.829)	
N Male	3.065	0.090	-0.157	0.229	0.188	-0.386*	-0.346**	0.041	855
	(1.652)	(0.144)	(0.167)	(0.204)	(0.164)	(0.065)	(0.041)	(0.841)	
Average Age	29.381	2.038*	3.509**	1.627	1.073	1.882	2.436*	0.554	873
	(10.131)	(1.186)	(1.537)	(1.611)	(1.194)	(0.275)	(0.071)	(0.695)	
Resp. Age	40.272	1.157	2.596	0.175	0.728	2.421	1.868	-0.553	872
	(12.569)	(1.257)	(1.846)	(1.556)	(1.680)	(0.237)	(0.382)	(0.769)	
Resp. Edu	2.806	-0.400	-0.561	-0.278	-0.364	-0.284	-0.197	0.086	872
	(3.947)	(0.448)	(0.537)	(0.510)	(0.540)	(0.554)	(0.699)	(0.858)	
Resp. Lit.	0.516	-0.041	-0.049	0.003	-0.072	-0.052	0.023	0.075	871
	(0.501)	(0.066)	(0.082)	(0.083)	(0.078)	(0.534)	(0.776)	(0.355)	
Income	11.165	-0.062	-0.202	0.142	-0.115	-0.343**	-0.086	0.257*	873
	(1.321)	(0.158)	(0.182)	(0.179)	(0.172)	(0.031)	(0.556)	(0.076)	
Land	0.379	0.109*	0.059	0.112	0.151**	-0.053	-0.092	-0.039	870
	(0.605)	(0.057)	(0.072)	(0.077)	(0.074)	(0.526)	(0.262)	(0.653)	
TLU	2.096	0.319	0.315	0.363	0.283	-0.048	0.032	0.080	873
	(1.941)	(0.290)	(0.313)	(0.424)	(0.377)	(0.905)	(0.930)	(0.862)	
Has Migrant	0.624	-0.002	-0.053	0.016	0.027	-0.069	-0.080	-0.012	873
	(0.486)	(0.057)	(0.065)	(0.057)	(0.083)	(0.186)	(0.315)	(0.873)	

OLS estimates of baseline differences between treatment and control groups. For each outcome, we report the coefficient of interest and clustered (VDC) standard errors in parentheses; Column (1) reports the mean and standard deviation of the outcome variable. Column (2) compares the effect of belonging to any treatment group to the control group. Column (3) measures the effect of belonging to the FT group at baseline. Column (4) measures the effect of belonging to the NG group at baseline. Column (5) measures the effect of belonging to the NVT group at baseline. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Baseline balance: outcome indices

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Empowerment	0.776	-0.018	-0.024	-0.008	-0.022	-0.017	-0.002	0.015	1,101
	(0.187)	(0.020)	(0.024)	(0.023)	(0.027)	(0.472)	(0.942)	(0.577)	
Finance	0.078	-0.134	-0.242*	0.022	-0.188	-0.264	-0.054	0.210	1,109
	(1.031)	(0.111)	(0.124)	(0.160)	(0.140)	(0.103)	(0.701)	(0.226)	
Aspirations	2.046	-1.955	-3.427**	-0.558	-2.000	-2.869**	-1.426	1.443	1,108
	(13.203)	(1.285)	(1.355)	(1.496)	(1.400)	(0.013)	(0.152)	(0.222)	
Mental health	-0.052	-0.065	-0.147	-0.002	-0.055	-0.145	-0.091	0.053	1,109
	(1.013)	(0.101)	(0.147)	(0.114)	(0.139)	(0.340)	(0.594)	(0.709)	
Assets	0.181	-0.115	-0.175	-0.067	-0.109	-0.108	-0.066	0.042	1,109
	(0.889)	(0.150)	(0.199)	(0.173)	(0.200)	(0.586)	(0.767)	(0.833)	
Income (Rs.)	11.064	0.077	-0.031	0.258	-0.003	-0.289	-0.028	0.261	1,109
	(1.371)	(0.175)	(0.215)	(0.224)	(0.190)	(0.201)	(0.882)	(0.196)	
Non-food consumption	8.264	-0.153	-0.187	0.111	-0.377	-0.297	0.190	0.487	1,107
	(1.980)	(0.333)	(0.330)	(0.349)	(0.418)	(0.131)	(0.531)	(0.135)	
Physical health	0.041	-0.005	-0.205**	0.129	0.042	-0.334***	-0.248**	0.087	1,109
	(0.914)	(0.070)	(0.090)	(0.079)	(0.098)	(0.001)	(0.033)	(0.409)	
Food Security	-0.005	-0.064	-0.116	-0.029	-0.052	-0.087	-0.064	0.024	1,109
	(1.028)	(0.126)	(0.144)	(0.159)	(0.165)	(0.566)	(0.688)	(0.891)	

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Empowerment	0.743	0.014	0.011	0.000	0.030	0.010	-0.019	-0.030	867
	(0.183)	(0.020)	(0.022)	(0.028)	(0.025)	(0.675)	(0.383)	(0.281)	
Finance	-0.010	-0.094	-0.045	-0.001	-0.221	-0.044	0.176	0.220	873
	(1.025)	(0.116)	(0.123)	(0.153)	(0.135)	(0.742)	(0.125)	(0.134)	
Aspirations	1.358	-1.434	-2.717**	-0.849	-0.788	-1.868*	-1.929	-0.061	870
	(7.983)	(1.097)	(1.166)	(1.242)	(1.453)	(0.076)	(0.137)	(0.964)	
Mental health	-0.013	-0.111	-0.175	-0.038	-0.119	-0.138	-0.056	0.082	873
	(1.032)	(0.142)	(0.163)	(0.156)	(0.190)	(0.335)	(0.755)	(0.638)	
Assets	-0.041	0.085	0.063	0.223	-0.018	-0.160	0.081	0.241	873
	(0.919)	(0.130)	(0.170)	(0.165)	(0.200)	(0.424)	(0.723)	(0.287)	
Income (Rs.)	11.165	-0.062	-0.202	0.142	-0.115	-0.343**	-0.086	0.257*	873
	(1.321)	(0.158)	(0.182)	(0.179)	(0.172)	(0.031)	(0.556)	(0.076)	
Non-food consumption	8.395	-0.234	-0.355	0.002	-0.335	-0.357*	-0.020	0.337	871
	(2.016)	(0.251)	(0.270)	(0.269)	(0.307)	(0.081)	(0.937)	(0.181)	
Physical health	-0.096	0.048	-0.032	0.188	-0.004	-0.220***	-0.028	0.192*	873
	(0.995)	(0.114)	(0.116)	(0.119)	(0.141)	(0.006)	(0.797)	(0.088)	
Food Security	0.020	-0.177*	-0.321***	-0.041	-0.166	-0.281**	-0.155	0.125	873
	(1.149)	(0.097)	(0.119)	(0.109)	(0.131)	(0.019)	(0.261)	(0.333)	

OLS estimates of baseline differences between treatment and control groups. For each outcome, we report the coefficient of interest and clustered (VDC) standard errors in parentheses; Column (1) reports the mean and standard deviation of the outcome variable. Column (2) compares the effect of belonging to any treatment group to the control group. Column (3) measures the effect of belonging to the FT group at baseline. Column (4) measures the effect of belonging to the TNG group at baseline. Column (5) measures the effect of belonging to the NVT group at baseline. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Treatment compliance by arm and direct/pay-it-forward

	Control mean	FT	NG	NVT	N
Targeted direct beneficiaries	0.122	0.779***	0.687***	0.650***	1,031
	(0.328)	(0.076)	(0.088)	(0.086)	
Potential PIF	0.066	0.739***	0.560***	0.165**	797
	(0.249)	(0.064)	(0.108)	(0.076)	

OLS regressions, clustered (VDC) standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: ITT effects on summary indices

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Empowerment	0.777	0.043**	0.040**	0.042*	0.047**	-0.002	-0.006	-0.005	1,020
	(0.189)	(0.018)†	(0.019)	(0.023)	(0.022)	(0.933)	(0.742)	(0.830)	
Financial Inclusion	0.077	0.341***	0.320**	0.355**	0.343***	-0.035	-0.023	0.012	1,033
	(0.981)	(0.109)†	(0.132)	(0.140)	(0.124)	(0.812)	(0.838)	(0.928)	
Aspirations	-0.810	1.131	-0.154	2.731	0.646	-2.885	-0.800	2.085	1,031
	(12.631)	(1.700)	(1.678)	(2.415)	(1.990)	(0.198)	(0.605)	(0.374)	
Mental health	0.009	0.045	0.148	-0.078	0.076	0.226**	0.072	-0.154	1,032
	(0.970)	(0.076)	(0.091)	(0.101)	(0.087)	(0.038)	(0.426)	(0.130)	
Assets	0.080	0.002	-0.015	0.005	0.012	-0.020	-0.027	-0.007	1,032
	(1.031)	(0.108)	(0.126)	(0.153)	(0.112)	(0.888)	(0.804)	(0.957)	
Income	11.503	-0.027	-0.125	0.094	-0.061	-0.219*	-0.064	0.155	1,031
	(1.228)	(0.141)	(0.175)	(0.147)	(0.164)	(0.087)	(0.678)	(0.218)	
Non-food expenditures	9.641	-0.111	-0.175	-0.078	-0.093	-0.097	-0.082	0.015	1,033
	(1.397)	(0.121)	(0.144)	(0.145)	(0.141)	(0.496)	(0.533)	(0.910)	
Physical health	0.041	0.043	-0.014	0.052	0.078	-0.065	-0.091	-0.026	1,031
	(0.813)	(0.071)	(0.077)	(0.101)	(0.091)	(0.532)	(0.304)	(0.814)	
Food Security	0.004	0.019	0.029	0.151	-0.102	-0.122	0.132	0.253**	1,032
	(0.953)	(0.075)	(0.093)	(0.105)	(0.099)	(0.301)	(0.210)	(0.044)	

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Empowerment	0.749	0.057***	0.067***	0.078***	0.031	-0.011	0.037	0.048**	786
	(0.195)	(0.018)†	(0.023)	(0.021)	(0.021)	(0.649)	(0.140)	(0.032)	
Financial Inclusion	-0.041	0.226**	0.171*	0.377***	0.145	-0.205*	0.026	0.232**	796
	(1.032)	(0.087)†	(0.098)	(0.121)	(0.111)	(0.098)	(0.805)	(0.049)	
Aspirations	-0.845	-0.497	-1.666	0.616	-0.414	-2.281	-1.251	1.030	794
	(14.917)	(1.370)	(2.161)	(1.489)	(1.576)	(0.282)	(0.579)	(0.507)	
Mental health	0.023	0.042	0.070	-0.162	0.190*	0.233***	-0.119	-0.352***	795
	(0.899)	(0.098)	(0.093)	(0.115)	(0.111)	(0.005)	(0.111)	(0.001)	
Assets	-0.037	-0.027	0.094	-0.099	-0.073	0.193*	0.167	-0.026	795
	(1.064)	(0.095)	(0.127)	(0.099)	(0.109)	(0.070)	(0.155)	(0.765)	
Income	11.506	-0.148	-0.168	-0.230	-0.061	0.062	-0.106	-0.168	795
	(1.134)	(0.135)	(0.186)	(0.153)	(0.148)	(0.701)	(0.552)	(0.231)	
Non-food expenditures	9.537	0.021	0.133	-0.033	-0.030	0.166	0.163	-0.003	796
	(1.449)	(0.144)	(0.173)	(0.183)	(0.147)	(0.307)	(0.230)	(0.983)	
Physical health	0.064	-0.017	-0.123	0.158	-0.069	-0.281**	-0.054	0.226**	794
	(0.936)	(0.079)	(0.106)	(0.100)	(0.091)	(0.017)	(0.628)	(0.026)	
Food Security	-0.059	0.041	0.125	0.013	-0.009	0.112	0.133	0.022	795
	(0.874)	(0.073)	(0.089)	(0.097)	(0.098)	(0.296)	(0.206)	(0.845)	

OLS regressions, clustered (VDC) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. N.B. that FT=NG, FT=NVT, NG=NVT represent Wald tests of a null of equal treatment effects, p-values are reported as the sub-statistic. Control variables include baseline dependent variable, stratification bin dummies, and imbalanced variables at baseline. FT: full treatment, NG: no-goats treatment, NVT: no-values-based-training treatment. Indices for assets, financial inclusion, food security, physical health, and mental health are calculated as a standardized weighted average of subindicators following Anderson (2008). Income is the logged sum of total household income. Non-food expenditures is the logged sum of total expenditures. The aspirations index is the Bernard and Taffesse (2014) aspirations index. The empowerment index is a modified 5DE index of the A-WEAI (Alkire et al., 2012; Malapit et al., 2015).

Table 6: ITT effects on women's empowerment

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Empowerment index	0.777	0.043**	0.040**	0.042*	0.047**	-0.002	-0.006	-0.005	1,020
	(0.189)	(0.018)	(0.019)	(0.023)	(0.022)	(0.933)	(0.742)	(0.830)	
Production decisions	0.903	0.027	0.037*	0.005	0.040	0.033	-0.003	-0.035	1,030
	(0.296)	(0.022)	(0.021)	(0.031)	(0.024)	(0.197)	(0.901)	(0.236)	
Asset ownership	0.926	0.042**	0.040**	0.045**	0.040*	-0.005	0.000	0.005	1,030
	(0.262)	(0.016)	(0.017)	(0.019)	(0.020)	(0.764)	(0.987)	(0.787)	
Access to and control over credit	0.416	-0.030	-0.049	-0.025	-0.021	-0.024	-0.028	-0.004	1,025
	(0.494)	(0.046)	(0.052)	(0.066)	(0.056)	(0.712)	(0.623)	(0.960)	
Control over income	0.892	0.047**	0.020	0.078***	0.038	-0.057**	-0.018	0.039	1,026
	(0.311)	(0.023)	(0.025)	(0.026)	(0.030)	(0.028)	(0.542)	(0.148)	
Group membership	0.651	0.159***	0.144***	0.170***	0.161***	-0.026	-0.016	0.009	1,027
	(0.478)	(0.036)	(0.051)	(0.046)	(0.041)	(0.644)	(0.735)	(0.837)	
Works $\leq$ 10.5 hours per day	0.685	-0.017	0.006	-0.062*	0.006	0.068	0.000	-0.068	1,031
	(0.465)	(0.027)	(0.038)	(0.036)	(0.040)	(0.127)	(0.998)	(0.155)	

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Empowerment index	0.749	0.057***	0.067***	0.078***	0.031	-0.011	0.037	0.048**	786
	(0.195)	(0.018)	(0.023)	(0.021)	(0.021)	(0.649)	(0.140)	(0.032)	
Production decisions	0.846	0.104***	0.115***	0.101***	0.096***	0.014	0.019	0.005	794
	(0.362)	(0.029)	(0.033)	(0.029)	(0.032)	(0.510)	(0.475)	(0.809)	
Asset ownership	0.917	0.052***	0.058***	0.062***	0.038*	-0.004	0.020	0.024	793
	(0.276)	(0.019)	(0.021)	(0.022)	(0.020)	(0.825)	(0.181)	(0.111)	
Access to and control over credit	0.323	0.061	0.028	0.097*	0.059	-0.069	-0.032	0.038	791
	(0.469)	(0.040)	(0.046)	(0.053)	(0.050)	(0.182)	(0.514)	(0.491)	
Control over income	0.892	0.055**	0.033	0.084***	0.049*	-0.051**	-0.016	0.035	793
	(0.311)	(0.025)	(0.031)	(0.027)	(0.026)	(0.032)	(0.513)	(0.105)	
Group membership	0.579	0.113**	0.133**	0.186***	0.032	-0.054	0.100	0.154**	793
	(0.495)	(0.048)	(0.056)	(0.064)	(0.059)	(0.415)	(0.118)	(0.020)	
Works $\leq$ 10.5 hours per day	0.703	-0.031	0.005	-0.056	-0.042	0.061	0.047	-0.014	795
	(0.458)	(0.034)	(0.033)	(0.055)	(0.050)	(0.257)	(0.339)	(0.829)	

OLS regressions, clustered (VDC) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. N.B. that FT=NG, FT=NVT, NG=NVT represent Wald tests of a null of equal treatment effects, p-values are reported as the sub-statistic. Control variables include baseline dependent variable, stratification bin dummies, and imbalanced variables at baseline. FT: full treatment, NG: no-goats treatment, NVT: no-values-based-training treatment. The empowerment index is a modified 5DE index of the A-WEAI (Alkire et al., 2012; Malapit et al., 2015). The index aggregates an empowerment score across decisions about production, access to and decision-making power over productive resources, control over credit and income, leadership in the community and time allocation. Each binary subindicator equals one if the respondent achieves empowerment, and zero otherwise. Weights and definitions of adequacy are defined in the main text.

Table 7: ITT effects on financial inclusion

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Financial index	0.077	0.341***	0.320**	0.355**	0.343***	-0.035	-0.023	0.012	1,033
	(0.981)	(0.109)	(0.132)	(0.140)	(0.124)	(0.812)	(0.838)	(0.928)	
Amount saved	3.824	0.700**	0.708**	1.055***	0.379	-0.348	0.329	0.677**	1,032
	(2.930)	(0.282)	(0.335)	(0.318)	(0.347)	(0.268)	(0.277)	(0.036)	
Savings group	0.539	0.168***	0.191***	0.183***	0.139**	0.008	0.052	0.044	1,026
	(0.499)	(0.052)	(0.060)	(0.065)	(0.064)	(0.904)	(0.367)	(0.488)	
Owe formal lender	0.937	0.370	0.151	0.331	0.564	-0.180	-0.413	-0.233	1,032
	(3.242)	(0.377)	(0.481)	(0.506)	(0.424)	(0.723)	(0.378)	(0.645)	
Owe informal lender	3.352	-0.571	-0.454	-0.942*	-0.336	0.488	-0.118	-0.606	1,032
	(5.129)	(0.415)	(0.482)	(0.557)	(0.479)	(0.364)	(0.800)	(0.247)	
Discount rate	0.053	-0.014	-0.027**	-0.005	-0.013	-0.022*	-0.014	0.008	749
	(0.079)	(0.011)	(0.011)	(0.014)	(0.013)	(0.070)	(0.144)	(0.505)	
Planning horizon	1.829	0.086	0.082	-0.053	0.209	0.134	-0.127	-0.262*	1,029
	(0.943)	(0.126)	(0.148)	(0.155)	(0.147)	(0.384)	(0.357)	(0.090)	

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Financial index	-0.041	0.226**	0.171*	0.377***	0.145	-0.205*	0.026	0.232**	796
	(1.032)	(0.087)	(0.098)	(0.121)	(0.111)	(0.098)	(0.805)	(0.049)	
Amount saved	3.545	0.137	0.092	0.628**	-0.238	-0.535	0.330	0.866**	795
	(2.981)	(0.256)	(0.308)	(0.310)	(0.340)	(0.117)	(0.387)	(0.021)	
Savings group	0.492	0.101**	0.136**	0.181***	0.002	-0.046	0.134*	0.180***	793
	(0.501)	(0.047)	(0.062)	(0.058)	(0.061)	(0.474)	(0.058)	(0.004)	
Owe formal lender	0.646	0.253	-0.325	0.594	0.474	-0.920**	-0.799**	0.120	795
	(2.802)	(0.330)	(0.346)	(0.456)	(0.405)	(0.035)	(0.025)	(0.806)	
Owe informal lender	3.316	-0.304	-0.288	-0.702	0.019	0.414	-0.307	-0.721*	795
	(5.183)	(0.366)	(0.423)	(0.474)	(0.396)	(0.305)	(0.405)	(0.079)	
Discount rate	0.043	0.002	-0.012	0.011	0.005	-0.023**	-0.017	0.006	585
	(0.064)	(0.010)	(0.009)	(0.012)	(0.013)	(0.032)	(0.124)	(0.629)	
Planning horizon	1.759	0.163	0.164	0.080	0.234	0.084	-0.069	-0.153	794
	(0.962)	(0.105)	(0.111)	(0.126)	(0.140)	(0.516)	(0.586)	(0.260)	

OLS regressions, clustered (VDC) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. N.B. that FT=NG, FT=NVT, NG=NVT represent Wald tests of a null of equal treatment effects, p-values are reported as the sub-statistic. Control variables include baseline dependent variable, stratification bin dummies, and imbalanced variables at baseline. FT: full treatment, NG: no-goats treatment, NVT: no-values-based-training treatment. The financial inclusion index is a weighted standardized average of each subindicator presented in this table. Subindicators include the logged amount saved in the last month, a dummy variable equal to one if the respondent belongs to a savings group, the logged amount owed to formal/informal lenders, the calculated discount rate following Ashraf, Karlan, and Yin (2006), and an ordered categorical variable indicating how far individuals plan ahead following Laajaj (2017).

Table 8: ITT effects on aspirations

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Aspirations index	-0.810	1.131	-0.154	2.731	0.646	-2.885	-0.800	2.085	1,031
	(12.631)	(1.700)	(1.678)	(2.415)	(1.990)	(0.198)	(0.605)	(0.374)	
Income aspirations	11.043	0.530**	0.380	0.785**	0.419	-0.405	-0.039	0.366	1,031
	(3.038)	(0.239)	(0.240)	(0.340)	(0.279)	(0.193)	(0.873)	(0.268)	
Asset aspirations	13.794	-0.041	-0.253	0.271	-0.163	-0.523	-0.090	0.433	1,031
	(2.754)	(0.271)	(0.285)	(0.387)	(0.371)	(0.188)	(0.800)	(0.333)	
Children's education aspirations	14.885	-0.141	-0.769	0.058	0.139	-0.827	-0.908**	-0.081	1,031
	(3.581)	(0.485)	(0.543)	(0.608)	(0.555)	(0.165)	(0.033)	(0.890)	
Daughters' education aspirations	14.185	-0.388	-1.227**	-0.007	-0.107	-1.219*	-1.119**	0.100	1,031
	(4.086)	(0.560)	(0.603)	(0.746)	(0.612)	(0.082)	(0.022)	(0.886)	
Sons' education aspirations	14.581	-0.189	-0.574	0.145	-0.201	-0.719	-0.373	0.346	1,031
	(3.810)	(0.432)	(0.469)	(0.580)	(0.557)	(0.228)	(0.431)	(0.591)	
Status aspirations	15.567	2.788	4.027	4.515	0.366	-0.488	3.660	4.148	1,031
	(19.168)	(2.187)	(2.541)	(2.886)	(2.490)	(0.865)	(0.115)	(0.130)	

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Aspirations index	-0.845	-0.497	-1.666	0.616	-0.414	-2.281	-1.251	1.030	794
	(14.917)	(1.370)	(2.161)	(1.489)	(1.576)	(0.282)	(0.579)	(0.507)	
Income aspirations	11.045	0.186	0.044	0.532	0.021	-0.488	0.023	0.510	795
	(3.131)	(0.321)	(0.397)	(0.356)	(0.355)	(0.131)	(0.950)	(0.103)	
Asset aspirations	13.814	-0.211	-0.443	0.081	-0.260	-0.524	-0.183	0.340	795
	(2.478)	(0.235)	(0.378)	(0.234)	(0.294)	(0.163)	(0.654)	(0.237)	
Children's education aspirations	14.692	-0.015	-0.059	-0.450	0.411	0.391	-0.470	-0.861**	795
	(3.612)	(0.374)	(0.499)	(0.415)	(0.422)	(0.373)	(0.338)	(0.031)	
Daughters' education aspirations	13.626	-0.093	-0.253	-0.307	0.244	0.054	-0.497	-0.551	794
	(4.330)	(0.388)	(0.492)	(0.483)	(0.457)	(0.915)	(0.317)	(0.257)	
Sons' education aspirations	14.574	-0.302	-0.441	-0.488	-0.009	0.047	-0.432	-0.478	794
	(3.507)	(0.309)	(0.474)	(0.338)	(0.373)	(0.919)	(0.391)	(0.214)	
Status aspirations	15.641	0.545	-0.203	1.422	0.469	-1.625	-0.672	0.953	795
	(20.629)	(2.188)	(2.637)	(2.610)	(2.575)	(0.535)	(0.788)	(0.677)	

OLS regressions, clustered (VDC) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. N.B. that FT=NG, FT=NVT, NG=NVT represent Wald tests of a null of equal treatment effects, p-values are reported as the sub-statistic. Control variables include baseline dependent variable, stratification bin dummies, and imbalanced variables at baseline. FT: full treatment, NG: no-goats treatment, NVT: no-values-based-training treatment. The aspirations index is the Bernard and Taffesse (2014) aspirations index, a weighted average of aspirations for income, assets, education (not disaggregated by gender) and social status. Weights are assigned through the distribution of 20 tokens across four bins in proportion to how heavily the respondent values a particular category. Specific questions used to elicit aspirations are described in the main text.

Table 9: ITT effects on mental health

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Mental health index	0.009	0.045	0.148	-0.078	0.076	0.226**	0.072	-0.154	1,032
	(0.970)	(0.076)	(0.091)	(0.101)	(0.087)	(0.038)	(0.426)	(0.130)	
Depression score	6.541	0.096	0.273	-0.102	0.138	0.374*	0.134	-0.240	1,032
	(1.906)	(0.205)	(0.245)	(0.227)	(0.231)	(0.066)	(0.536)	(0.209)	
Locus of control	2.915	-0.083	-0.242	0.035	-0.070	-0.277*	-0.173	0.105	1,032
	(1.482)	(0.133)	(0.160)	(0.161)	(0.166)	(0.088)	(0.308)	(0.562)	
Optimism	6.404	-0.015	0.069	-0.058	-0.040	0.127	0.109	-0.018	1,032
	(1.184)	(0.128)	(0.165)	(0.175)	(0.143)	(0.503)	(0.471)	(0.909)	
Life Satisfaction	6.361	0.256	0.165	0.448**	0.150	-0.282	0.015	0.298	1,030
	(1.904)	(0.173)	(0.208)	(0.207)	(0.191)	(0.164)	(0.935)	(0.109)	
Self-esteem	9.604	0.219*	0.271	0.179	0.215	0.092	0.056	-0.036	1,032
	(1.655)	(0.113)	(0.171)	(0.137)	(0.140)	(0.607)	(0.754)	(0.803)	
Happiness	2.070	-0.044	-0.007	-0.084	-0.036	0.078	0.029	-0.049	1,032
	(0.530)	(0.054)	(0.066)	(0.060)	(0.064)	(0.168)	(0.650)	(0.388)	
Worry score	9.075	0.217	0.503**	-0.049	0.246	0.551***	0.257	-0.294	1,017
	(2.091)	(0.197)	(0.206)	(0.232)	(0.226)	(0.004)	(0.171)	(0.163)	
Trust score	1.528	-0.002	0.108	-0.248	0.130	0.356*	-0.022	-0.378*	1,029
	(1.334)	(0.160)	(0.171)	(0.222)	(0.181)	(0.077)	(0.901)	(0.079)	

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Mental health index	0.023	0.042	0.070	-0.162	0.190*	0.233***	-0.119	-0.352***	795
	(0.899)	(0.098)	(0.093)	(0.115)	(0.111)	(0.005)	(0.111)	(0.001)	
Depression score	6.610	-0.005	0.268	-0.220	-0.071	0.488*	0.339*	-0.150	795
	(1.889)	(0.176)	(0.173)	(0.261)	(0.201)	(0.051)	(0.078)	(0.560)	
Locus of control	2.990	-0.143	-0.094	-0.257*	-0.090	0.163	-0.004	-0.167	795
	(1.388)	(0.116)	(0.121)	(0.149)	(0.165)	(0.227)	(0.977)	(0.350)	
Optimism	6.354	0.023	0.018	-0.085	0.117	0.103	-0.099	-0.201	795
	(1.185)	(0.142)	(0.145)	(0.158)	(0.172)	(0.437)	(0.424)	(0.167)	
Life Satisfaction	6.431	0.076	-0.046	0.232	0.051	-0.279	-0.098	0.181	794
	(1.940)	(0.177)	(0.221)	(0.206)	(0.185)	(0.162)	(0.610)	(0.297)	
Self-esteem	9.841	-0.047	-0.007	-0.086	-0.050	0.079	0.043	-0.036	795
	(1.819)	(0.153)	(0.198)	(0.199)	(0.230)	(0.725)	(0.872)	(0.887)	
Happiness	2.051	-0.008	-0.049	-0.043	0.058	-0.007	-0.107**	-0.101*	795
	(0.525)	(0.055)	(0.056)	(0.066)	(0.060)	(0.890)	(0.021)	(0.060)	
Worry score	8.917	0.329	0.534**	0.050	0.381	0.484	0.153	-0.331	787
	(1.951)	(0.220)	(0.230)	(0.366)	(0.238)	(0.180)	(0.474)	(0.339)	
Trust score	1.528	0.087	0.327	-0.294	0.196	0.621**	0.131	-0.489*	794
	(1.352)	(0.197)	(0.210)	(0.267)	(0.243)	(0.017)	(0.575)	(0.070)	

OLS regressions, clustered (VDC) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. N.B. that FT=NG, FT=NVT, NG=NVT represent Wald tests of a null of equal treatment effects, p-values are reported as the sub-statistic. Control variables include baseline dependent variable, stratification bin dummies, and imbalanced variables at baseline. FT: full treatment, NG: no-goats treatment, NVT: no-values-based-training treatment. The mental health index is a weighted standardized average of each subindicator presented in this table. Depression is an abbreviated version of the CES-D scale Radloff (1977) with a high value indicating high levels of depression. Locus of control is an abbreviated Rotter (1966) scale where a high value indicates a stronger internal locus of control. Remaining subindicators are based on aggregated responses to 3-4 questions per subindicator from the 2009 World Values Survey, where high values indicate positive welfare.

Table 10: ITT effects on household asset holdings

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Asset index	0.080	0.002	-0.015	0.005	0.012	-0.020	-0.027	-0.007	1,032
	(1.031)	(0.108)	(0.126)	(0.153)	(0.112)	(0.888)	(0.804)	(0.957)	
Productive asset index	0.134	-0.000	0.002	0.009	-0.009	-0.007	0.011	0.018*	1,029
	(0.068)	(0.008)	(0.008)	(0.009)	(0.010)	(0.381)	(0.200)	(0.071)	
Non-productive asset index	0.475	-0.011	-0.022	-0.005	-0.008	-0.016	-0.013	0.003	1,030
	(0.083)	(0.013)	(0.013)	(0.015)	(0.013)	(0.166)	(0.156)	(0.786)	
Livestock (TLU)	2.576	0.196	0.244	0.043	0.292	0.201	-0.048	-0.249	1,032
	(2.353)	(0.165)	(0.170)	(0.229)	(0.217)	(0.359)	(0.825)	(0.326)	
Land (hectares)	0.470	0.060**	0.059*	0.012	0.103***	0.047	-0.044	-0.091**	1,029
	(0.538)	(0.030)	(0.030)	(0.038)	(0.037)	(0.216)	(0.238)	(0.039)	
Housing index	2.534	-0.009	0.074	-0.104	0.020	0.178	0.053	-0.125	1,032
	(0.824)	(0.091)	(0.135)	(0.111)	(0.097)	(0.168)	(0.642)	(0.233)	

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Asset index	-0.037	-0.027	0.094	-0.099	-0.073	0.193*	0.167	-0.026	795
	(1.064)	(0.095)	(0.127)	(0.099)	(0.109)	(0.070)	(0.155)	(0.765)	
Productive asset index	0.135	0.003	0.002	0.012	-0.003	-0.009	0.005	0.015	794
	(0.069)	(0.007)	(0.007)	(0.010)	(0.008)	(0.264)	(0.431)	(0.141)	
Non-productive asset index	0.472	-0.016	-0.012	-0.021*	-0.015	0.008	0.002	-0.006	794
	(0.083)	(0.010)	(0.011)	(0.012)	(0.011)	(0.375)	(0.790)	(0.535)	
Livestock (TLU)	2.382	0.209	0.435**	-0.017	0.202	0.452**	0.234	-0.218	795
	(2.050)	(0.136)	(0.187)	(0.165)	(0.162)	(0.021)	(0.232)	(0.240)	
Land (hectares)	0.392	0.024	0.030	0.022	0.022	0.008	0.008	-0.000	794
	(0.475)	(0.033)	(0.038)	(0.040)	(0.039)	(0.817)	(0.835)	(0.989)	
Housing index	2.435	0.040	0.160	-0.089	0.043	0.249**	0.116	-0.132*	795
	(0.927)	(0.107)	(0.149)	(0.102)	(0.114)	(0.033)	(0.342)	(0.094)	

OLS regressions, clustered (VDC) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. N.B. that FT=NG, FT=NVT, NG=NVT represent Wald tests of a null of equal treatment effects, p-values are reported as the sub-statistic. Control variables include baseline dependent variable, stratification bin dummies, and imbalanced variables at baseline. FT: full treatment, NG: no-goats treatment, NVT: no-values-based-training treatment. The asset index is a weighted standardized average of each subindicator presented in this table. The productive and non-productive asset indices are each a principle components index of owned productive and non-productive assets respectively. Livestock represents owned goats, cattle, water buffalo, swine and chickens aggregated into tropical livestock units (TLUs). Land is measured by total hectares owned. The housing index is a principle components index of various indicators of housing quality.

Table 11: ITT effects on household income

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Total income	11.503	-0.027	-0.125	0.094	-0.061	-0.219*	-0.064	0.155	1,031
	(1.228)	(0.141)	(0.175)	(0.147)	(0.164)	(0.087)	(0.678)	(0.218)	
Livestock income	4.908	0.409	-0.093	0.585	0.623*	-0.679	-0.716*	-0.038	1,031
	(5.100)	(0.270)	(0.379)	(0.395)	(0.318)	(0.126)	(0.096)	(0.930)	
Crop income	3.274	-0.562	-1.055**	-0.486	-0.261	-0.569	-0.794	-0.225	1,031
	(4.902)	(0.415)	(0.485)	(0.461)	(0.491)	(0.236)	(0.109)	(0.594)	
Permanent income	2.317	0.070	-0.444	0.621	-0.030	-1.065	-0.415	0.651	1,031
	(4.699)	(0.431)	(0.499)	(0.661)	(0.543)	(0.120)	(0.478)	(0.369)	
Business income	2.750	0.537	0.235	1.136*	0.240	-0.901	-0.005	0.896*	1,031
	(4.633)	(0.508)	(0.693)	(0.591)	(0.524)	(0.120)	(0.993)	(0.062)	
Cash income	2.827	-0.173	-0.434	0.078	-0.201	-0.512	-0.233	0.279	1,031
	(5.012)	(0.679)	(0.696)	(0.797)	(0.782)	(0.357)	(0.703)	(0.683)	
Other income	2.241	0.597	0.470	0.947	0.386	-0.477	0.083	0.560	1,031
	(4.505)	(0.502)	(0.624)	(0.614)	(0.523)	(0.384)	(0.873)	(0.276)	

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Total income	11.506	-0.148	-0.168	-0.230	-0.061	0.062	-0.106	-0.168	795
	(1.134)	(0.135)	(0.186)	(0.153)	(0.148)	(0.701)	(0.552)	(0.231)	
Livestock income	4.427	1.497***	2.011***	1.227*	1.269**	0.784	0.742	-0.042	795
	(4.983)	(0.469)	(0.626)	(0.622)	(0.492)	(0.275)	(0.185)	(0.941)	
Crop income	2.834	0.189	0.486	0.252	-0.139	0.234	0.625	0.391	795
	(4.628)	(0.428)	(0.549)	(0.389)	(0.668)	(0.628)	(0.381)	(0.532)	
Permanent income	2.584	-0.252	-0.387	0.159	-0.479	-0.546	0.092	0.638	795
	(4.897)	(0.492)	(0.522)	(0.659)	(0.604)	(0.348)	(0.859)	(0.330)	
Business income	2.750	0.642	0.128	1.284**	0.557	-1.157	-0.430	0.727*	795
	(4.663)	(0.549)	(0.808)	(0.570)	(0.555)	(0.106)	(0.536)	(0.096)	
Day labor income	2.452	-0.133	-0.631	-0.303	0.457	-0.328	-1.088*	-0.759	795
	(4.785)	(0.568)	(0.621)	(0.697)	(0.631)	(0.572)	(0.051)	(0.212)	
Other income	2.748	0.489	0.249	0.697	0.528	-0.448	-0.278	0.169	795
	(4.855)	(0.533)	(0.675)	(0.603)	(0.573)	(0.450)	(0.620)	(0.718)	

OLS regressions, clustered (VDC) standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$. N.B. that FT=NG, FT=NVT, NG=NVT represent Wald tests of a null of equal treatment effects, p-values are reported as the sub-statistic. Control variables include baseline dependent variable, stratification bin dummies, and imbalanced variables at baseline. FT: full treatment, NG: no-goats treatment, NVT: no-values-based-training treatment. Total income is the logged sum of total household income. Each subindicator is also a logged sum. Total household income includes livestock income, crop income, permanent salaried income, small business (entrepreneurial) income, income earned as a day laborer and other miscellaneous income including remittances.

Table 12: ITT effects on non-food expenditures

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Total expenditure	9.641	-0.111	-0.175	-0.078	-0.093	-0.097	-0.082	0.015	1,033
	(1.397)	(0.121)	(0.144)	(0.145)	(0.141)	(0.496)	(0.533)	(0.910)	
Medical expenditures	6.713	-0.264	-0.225	0.109	-0.615*	-0.333	0.391	0.724*	1,033
	(3.578)	(0.293)	(0.380)	(0.391)	(0.333)	(0.415)	(0.308)	(0.060)	
Clothing expenditures	7.358	0.053	-0.209	0.174	0.137	-0.383*	-0.346*	0.037	1,033
	(1.805)	(0.180)	(0.237)	(0.193)	(0.192)	(0.068)	(0.072)	(0.788)	
Misc. expenditures	6.084	-0.688*	-0.696	-0.861*	-0.528	0.165	-0.168	-0.333	1,033
	(3.468)	(0.395)	(0.501)	(0.455)	(0.462)	(0.737)	(0.711)	(0.436)	

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Total expenditure	9.537	0.021	0.133	-0.033	-0.030	0.166	0.163	-0.003	796
	(1.449)	(0.144)	(0.173)	(0.183)	(0.147)	(0.307)	(0.230)	(0.983)	
Medical expenditures	6.650	-0.212	0.192	-0.375	-0.433	0.567**	0.625*	0.058	796
	(3.766)	(0.327)	(0.323)	(0.352)	(0.398)	(0.039)	(0.057)	(0.860)	
Clothing expenditures	7.478	-0.078	-0.072	-0.069	-0.091	-0.004	0.019	0.023	796
	(1.666)	(0.145)	(0.174)	(0.188)	(0.150)	(0.984)	(0.889)	(0.881)	
Misc. expenditures	6.178	-0.816*	-0.253	-1.213**	-0.959*	0.960*	0.706	-0.254	796
	(3.110)	(0.435)	(0.576)	(0.475)	(0.517)	(0.075)	(0.210)	(0.585)	

OLS regressions, clustered (VDC) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. N.B. that FT=NG, FT=NVT, NG=NVT represent Wald tests of a null of equal treatment effects, p-values are reported as the sub-statistic. Control variables include baseline dependent variable, stratification bin dummies, and imbalanced variables at baseline. FT: full treatment, NG: no-goats treatment, NVT: no-values-based-training treatment. Non-food expenditures is the logged sum of total annual expenditures. Each subindicator is also a logged sum. For expenditure categories that were not reported annually, we multiply the monthly or quarterly figures by the appropriate factor to achieve an annualized amount. Total expenditures include medical expenditures, clothing expenditures and miscellaneous expenditures.

Table 13: ITT effects on physical health

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Physical health index	0.041	0.043	-0.014	0.052	0.078	-0.065	-0.091	-0.026	1,031
	(0.813)	(0.071)	(0.077)	(0.101)	(0.091)	(0.532)	(0.304)	(0.814)	
Days work missed	1.554	0.144	0.020	0.345	0.061	-0.325	-0.041	0.284	1,030
	(2.985)	(0.276)	(0.361)	(0.403)	(0.343)	(0.481)	(0.920)	(0.517)	
Subjective own health	6.468	0.036	-0.145	0.195	0.028	-0.340	-0.173	0.167	1,030
	(1.886)	(0.129)	(0.165)	(0.180)	(0.149)	(0.108)	(0.312)	(0.384)	
Subjective child health	7.054	0.327**	0.132	0.434**	0.379*	-0.303	-0.247	0.055	680
	(1.596)	(0.162)	(0.186)	(0.208)	(0.195)	(0.175)	(0.204)	(0.796)	

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Physical health index	0.064	-0.017	-0.123	0.158	-0.069	-0.281**	-0.054	0.226**	794
	(0.936)	(0.079)	(0.106)	(0.100)	(0.091)	(0.017)	(0.628)	(0.026)	
Days work missed	1.590	0.389	0.536	0.264	0.363	0.272	0.172	-0.099	794
	(3.540)	(0.314)	(0.411)	(0.462)	(0.329)	(0.592)	(0.641)	(0.811)	
Subjective own health	6.518	0.029	-0.109	0.501**	-0.250	-0.610**	0.141	0.751***	794
	(1.845)	(0.176)	(0.239)	(0.222)	(0.188)	(0.018)	(0.555)	(0.001)	
Subjective child health	7.134	0.217	-0.008	0.440	0.212	-0.449	-0.220	0.228	517
	(1.797)	(0.268)	(0.342)	(0.312)	(0.291)	(0.178)	(0.489)	(0.341)	

OLS regressions, clustered (VDC) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. N.B. that FT=NG, FT=NVT, NG=NVT represent Wald tests of a null of equal treatment effects, p-values are reported as the sub-statistic. Control variables include baseline dependent variable, stratification bin dummies, and imbalanced variables at baseline. FT: full treatment, NG: no-goats treatment, NVT: no-values-based-training treatment. The physical health index is a weighted standardized average of each subindicator presented in this table. Subindicators include the number of days of work missed due to illness in the past month by the respondent, and the response to the following two questions: and

Table 14: ITT effects on food security

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Food Security Index	0.004	0.019	0.029	0.151	-0.102	-0.122	0.132	0.253**	1,032
	(0.953)	(0.075)	(0.093)	(0.105)	(0.099)	(0.301)	(0.210)	(0.044)	
Meals	2.019	0.017	0.043*	0.010	0.005	0.033	0.038*	0.004	1,031
	(0.235)	(0.012)	(0.021)	(0.019)	(0.010)	(0.199)	(0.061)	(0.801)	
Snacks	1.296	-0.008	-0.006	0.055	-0.065	-0.061	0.060	0.121	1,031
	(0.511)	(0.062)	(0.076)	(0.088)	(0.068)	(0.504)	(0.394)	(0.143)	
Skipped Meals (Resp)	0.019	0.009	0.000	0.005	0.018	-0.005	-0.017	-0.012	1,031
	(0.135)	(0.008)	(0.010)	(0.010)	(0.011)	(0.685)	(0.209)	(0.356)	
Skipped Day (Resp)	0.022	0.004	0.000	-0.005	0.015*	0.005	-0.015	-0.020**	1,031
	(0.148)	(0.008)	(0.011)	(0.008)	(0.008)	(0.651)	(0.189)	(0.014)	
Skipped Meal (Child)	0.016	0.011	0.027*	-0.004	0.010	0.030*	0.016	-0.014	483
	(0.128)	(0.010)	(0.015)	(0.012)	(0.012)	(0.072)	(0.289)	(0.326)	
Skipped Day (Child)	0.000	0.014*	0.005	0.011	0.025**	-0.007	-0.020**	-0.013	483
	(0.000)	(0.007)	(0.006)	(0.008)	(0.012)	(0.441)	(0.049)	(0.289)	
Enough to eat	0.759	0.071	0.075	0.070	0.068	0.006	0.007	0.001	1,031
	(0.428)	(0.051)	(0.061)	(0.059)	(0.058)	(0.921)	(0.893)	(0.978)	
DDI	9.248	-0.214	-0.673**	0.344	-0.380	-1.016***	-0.293	0.724**	1,032
	(2.876)	(0.243)	(0.286)	(0.315)	(0.312)	(0.005)	(0.323)	(0.049)	
HDDS	4.789	-0.046	-0.174	0.055	-0.042	-0.229*	-0.132	0.097	1,032
	(1.155)	(0.126)	(0.135)	(0.147)	(0.166)	(0.096)	(0.354)	(0.556)	
FCS	45.933	-1.942	-3.409**	-0.297	-2.346	-3.112	-1.063	2.049	1,032
	(15.254)	(1.486)	(1.672)	(1.908)	(1.884)	(0.106)	(0.550)	(0.327)	

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Food Security Index	-0.059	0.041	0.125	0.013	-0.009	0.112	0.133	0.022	795
	(0.874)	(0.073)	(0.089)	(0.097)	(0.098)	(0.296)	(0.206)	(0.845)	
Meals	2.026	0.005	0.035	-0.006	-0.011	0.040	0.045	0.005	794
	(0.188)	(0.014)	(0.021)	(0.019)	(0.016)	(0.078)	(0.031)	(0.790)	
Snacks	1.231	0.074	0.114	0.096	0.021	0.018	0.093	0.075	794
	(0.480)	(0.051)	(0.059)	(0.065)	(0.064)	(0.766)	(0.132)	(0.272)	
Skipped Meals (Resp)	0.021	0.013	0.023	-0.007	0.022	0.030	0.002	-0.029	794
	(0.142)	(0.013)	(0.018)	(0.018)	(0.016)	(0.164)	(0.931)	(0.123)	
Skipped Day (Resp)	0.010	0.009	0.021	0.005	0.003	0.016	0.018	0.002	794
	(0.101)	(0.010)	(0.018)	(0.014)	(0.011)	(0.464)	(0.354)	(0.883)	
Skipped Meal (Child)	0.011	0.001	0.040	-0.014	-0.011	0.054	0.051	-0.003	381
	(0.103)	(0.008)	(0.022)	(0.010)	(0.009)	(0.037)	(0.035)	(0.794)	
Skipped Day (Child)	0.011	-0.010	-0.011	-0.009	-0.010	-0.002	-0.002	0.000	381
	(0.103)	(0.007)	(0.008)	(0.006)	(0.007)	(0.590)	(0.717)	(0.875)	
Enough to eat	0.749	0.039	0.066	-0.011	0.059	0.077	0.007	-0.069	794
	(0.435)	(0.048)	(0.061)	(0.058)	(0.055)	(0.198)	(0.897)	(0.211)	
DDI	9.205	-0.258	-0.197	-0.120	-0.431	-0.078	0.233	0.311	795
	(3.000)	(0.361)	(0.491)	(0.446)	(0.398)	(0.878)	(0.605)	(0.425)	
HDDS	4.759	-0.055	-0.031	-0.080	-0.057	0.049	0.026	-0.023	795
	(1.121)	(0.137)	(0.157)	(0.176)	(0.155)	(0.770)	(0.849)	(0.884)	
FCS	45.633	-2.902	-2.015	-3.331	-3.323	1.316	1.308	-0.008	795
	(16.201)	(1.637)	(2.002)	(2.100)	(1.714)	(0.525)	(0.444)	(0.996)	

OLS regressions, clustered (VDC) standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$. N.B. that FT=NG, FT=NVT, NG=NVT represent Wald tests of a null of equal treatment effects, p-values are reported as the sub-statistic. Control variables include baseline dependent variable, stratification bin dummies, and imbalanced variables at baseline. FT: full treatment, NG: no-goats treatment, NVT: no-values-based-training treatment. The food security index is a weighted standardized average of each subindicator presented in this table, excluding child-specific indicators. Subindicators include the normal number of meals/snacks eaten per day, a dummy variable equal to one if the respondent/children skipped at least one meal in the past week, a dummy variable equal to one if the respondent/children went a full day without eating in the past week, and three modified indicators of dietary quality (described in the main text): a dietary diversity index (DDI), a household dietary diversity score (HDDS), and a food consumption score (FCS).

Table 15: Costs by treatment arm and direct/pay-it-forward

		Treatment Arm			
	OG+POG	*T1*	*T2*	*T3*	*All T*
Operations	Livestock	7,926,795	18,000	7,109,931	15,054,725
	Horticulture	2,388,577	2,310,583	1,187,457	5,886,617
	Equipment & Supply	3,409,281	3,772,544	1,975,263	9,157,088
	Trainings	4,639,482	5,167,277	2,043,355	11,850,115
Administrative	Tech Services & Eval	2,083,117	2,148,510	2,095,112	6,326,738
	Personnel	5,056,910	5,056,910	5,056,903	15,170,723
	Office Expenses	1,156,125	1,129,859	1,154,781	3,440,765
	Total	26,660,286	19,603,683	20,622,802	66,886,771
	Per Beneficiary	11,849	8,713	54,994	13,720
	OG				
Operations	Livestock	7,926,795	18,000	7,109,931	15,054,725
	Horticulture	2,388,577	2,310,583	1,187,457	5,886,617
	Equipment & Supply	3,409,281	3,772,544	1,975,263	9,157,088
	Trainings	2,088,512	2,356,270	910,377	5,355,160
Administrative	Tech Services & Eval	416,623	429,702	419,022	1,265,348
	Personnel	1,011,382	1,011,382	1,011,381	3,034,145
	Office Expenses	231,225	225,972	230,956	688,153
	Total	17,472,395	10,124,453	12,844,387	40,441,235
	Per Beneficiary	46,593	26,999	34,252	35,948
	POG				
Operations	Livestock	0	0	0	0
	Horticulture	0	0	0	0
	Equipment & Supply	0	0	0	0
	Trainings	2,020,692	2,020,692	0	4,041,385
Administrative	Tech Services & Eval	2,504,538	2,556,852	0	5,061,390
	Personnel	6,068,289	6,068,289	0	12,136,578
	Office Expenses	1,386,812	1,365,800	0	2,752,612
	Total	11,980,332	12,011,634	0	23,991,965
	Per Beneficiary	6,390	6,406	0	6,398

Nepali rupees, exchange rate roughly 100 NPR per USD.

Appendix

A Deviations from the pre-analysis plan

In this section we note material deviations from our original pre-analysis plan. First, the pre-analysis plan describes an analysis of spillover effects beyond the pay-it-forward targeted areas. In addition to sampling from central targeted wards (within the targeted VDC), our original sampling design selected households at random from a neighboring ward and a second additional ward within the targeted VDC. To reduce costs, we only collected data on households in these "spillover wards" in VDCs assigned to a treatment, but not in control VDCs. At the time, our assumption was that central wards would be statistically indistinguishable from non-central wards, so that a random sample from central wards in control VDCs would be an appropriate counterfactual to a random sample from non-central wards in treatment VDCs. Unfortunately, balance checks revealed some important differences between these two samples at baseline. Given these issues, we did not proceed with the initially proposed analysis of broader spillover effects.

Second, the pre-analysis plan also hypothesizes heterogeneity of impacts across several dimensions, including assets (wealth), income, empowerment, literacy levels of respondent and the head of household, and gender (the latter being applicable only for a small number of individual-level outcomes). In addition, for analysis of PIF impacts, we originally planned to consider heterogeneity by ward-level population since it's likely PIF impacts will be stronger in smaller wards than larger ones because a greater proportion of sampled individuals will have been treated. We do not conduct the originally proposed heterogeneity analysis due to a lack of power. We admit the heterogeneity analysis was overly ambitious, and even more so after removing 10/60 clusters from the sample following the aforementioned earthquake.

Third, we consider only nine of the originally proposed ten outcome dimensions, omitting time use. There are two reasons we do not include the pre-specified analysis for time use in the main paper. First, time use was included in the original list of outcomes as a potential unintended consequence of the intervention - that is, we do not expect the program to increase time spent in leisure activities, rather, we are worried the intervention will increase time spent working to the detriment of individual welfare. Given that, we should not include it in our estimations that control for the FDR; doing so penalizes us for a null result that should be interpreted positively. Second, time use is a subindicator of our empowerment summary index. Including it as a main summary index and a subindicator seems duplicative. For completeness, we do report the pre-specified results on time use in 16 with some discussion of the findings included in the sub-section on empowerment.

Finally, two of the summary indices (food security and physical health) were not calculated in strict accordance with the pre-analysis plan because of poor data quality. In the case of food security, the PAP specified that the summary index would consist of six subindicators related to meals eaten/skipped (number of meals/day eaten in the household hold, number of snacks/day eaten in the household, skipped at least one meal in the past week (binary, respondent/child), went a full day without eating in the past week (binary, respondent/child)), as well as separate household dietary diversity scores (HDDS) for the household as whole and for the children in the household. First, analysis showed that there was virtually no difference between the respondent and children for the meal skipping variables. Therefore, because this information is redundant and because calculating the index with child-specific sub-indicators included would have required us to impute a large number of values to childless households, we omit the child-specific subindicators. Second, the data suggest that meal skipping is extremely uncommon and that meals per day is very stable. This is consistent with the guidance of our Nepali partners, who point out that food insecure households in rural Nepal eat smaller quantities of a less diverse or nurtritious diet, but virtually never skip a meal entirely. Therefore, in addition to the pre-specified HDDS, we add two more measurements of dietary diversity: modified versions of the dietary diversity index (DDI) and the Food Consumption Score (FCS). To the extent that our data permit it we calculate DDI as suggested by (citation) and FCS as suggested by (citation), but in both cases our the structure of the data required us to deviate somewhat from established norms.

The physical health dimension as pre-specified included several measurements of child anthropometrics (HAZ, WAZ, MUAC) as sub-indicator components of the physical health index. An initial review of the descriptive statistics of the raw measurements and the calculated anthropometrics revealed implausible means and extremely high variances, suggesting very high levels of measurement error in the collected data. After follow-up with the field team and consultation with anthropometrics experts, we concluded that the equipment used was inadequate and protocols employed insufficient to accurately conduct anthropometric measurements under difficult field conditions. We therefore omit these subindicators from physical health.

Table 16: ITT effects on time use

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Time working	8.673	0.230	-0.167	0.483	0.290	-0.651	-0.457	0.194	974
	(3.258)	(0.294)	(0.422)	(0.382)	(0.384)	(0.167)	(0.362)	(0.659)	
Ag Work	2.037	-0.152	-0.468	0.214	-0.248	-0.682*	-0.221	0.461	974
	(2.929)	(0.297)	(0.346)	(0.382)	(0.335)	(0.052)	(0.506)	(0.197)	
LST Work	2.166	0.234	0.352	-0.152	0.489*	0.504*	-0.138	-0.642**	974
	(2.394)	(0.205)	(0.280)	(0.245)	(0.251)	(0.086)	(0.667)	(0.029)	
Other work	4.471	0.186	-0.042	0.442	0.122	-0.484	-0.165	0.319	974
	(2.808)	(0.263)	(0.295)	(0.352)	(0.306)	(0.136)	(0.591)	(0.352)	
Leisure	14.173	-0.087	0.287	-0.372	-0.103	0.659	0.389	-0.270	974
	(2.902)	(0.200)	(0.321)	(0.308)	(0.280)	(0.123)	(0.342)	(0.476)	

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Time working	9.057	-0.116	-0.333	0.253	-0.249	-0.586	-0.084	0.502	757
	(3.174)	(0.307)	(0.324)	(0.407)	(0.435)	(0.120)	(0.837)	(0.288)	
Ag Work	2.266	0.102	-0.235	0.607	-0.035	-0.842	-0.199	0.643	757
	(3.025)	(0.379)	(0.448)	(0.524)	(0.477)	(0.122)	(0.661)	(0.250)	
LST Work	2.008	-0.077	0.191	-0.052	-0.337	0.243	0.528*	0.285	757
	(2.334)	(0.198)	(0.272)	(0.252)	(0.238)	(0.390)	(0.067)	(0.288)	
Other work	4.822	-0.231	-0.517	-0.333	0.126	-0.184	-0.643**	-0.459	757
	(3.224)	(0.287)	(0.333)	(0.343)	(0.361)	(0.566)	(0.042)	(0.228)	
Leisure	14.112	0.033	0.297	-0.202	-0.000	0.498*	0.297	-0.201	757
	(2.790)	(0.251)	(0.257)	(0.325)	(0.324)	(0.072)	(0.311)	(0.552)	

OLS regressions, clustered (VDC) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. N.B. that FT=NG, FT=NVT, NG=NVT represent Wald tests of a null of equal treatment effects, p-values are reported as the sub-statistic. Control variables include baseline dependent variable, stratification bin dummies, and imbalanced variables at baseline. FT: full treatment, NG: no-goats treatment, NVT: no-values-based-training treatment.

B Additional Balance Checks

Table 17: Baseline balance: caste/ethnicity

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Brahmin	0.192	-0.041	-0.063	0.031	-0.091	-0.094	0.029	0.123	1,109
	(0.395)	(0.106)	(0.121)	(0.142)	(0.116)	(0.481)	(0.783)	(0.341)	
Chheriya	0.024	0.098**	0.149*	0.081	0.070	0.069	0.079	0.011	1,109
	(0.153)	(0.039)	(0.074)	(0.055)	(0.057)	(0.447)	(0.386)	(0.889)	
Newar	0.086	-0.063	-0.062	-0.079	-0.048	0.016	-0.015	-0.031	1,109
	(0.281)	(0.068)	(0.069)	(0.068)	(0.072)	(0.253)	(0.594)	(0.207)	
Hill Janajati	0.464	-0.013	-0.066	-0.005	0.027	-0.061	-0.094	-0.033	1,109
	(0.500)	(0.120)	(0.143)	(0.159)	(0.144)	(0.700)	(0.511)	(0.836)	
Tharu	0.000	0.059	0.031	0.076	0.066	-0.044	-0.035	0.010	1,109
	(0.000)	(0.035)	(0.031)	(0.073)	(0.064)	(0.580)	(0.628)	(0.922)	
Madeshi	0.079	-0.070	-0.051	-0.079*	-0.079*	0.028	0.028	-0.000	1,109
	(0.270)	(0.042)	(0.049)	(0.041)	(0.041)	(0.303)	(0.303)		
M. Dalit	0.010	0.009	-0.010	-0.007	0.042	-0.004	-0.052	-0.049	1,109
	(0.101)	(0.015)	(0.007)	(0.008)	(0.038)	(0.305)	(0.173)	(0.205)	
Gen. Dalit	0.086	-0.009	-0.015	0.012	-0.023	-0.027	0.008	0.035	1,109
	(0.281)	(0.049)	(0.045)	(0.085)	(0.047)	(0.731)	(0.791)	(0.660)	
Muslim	0.000	0.007	0.024	0.000**	0.000	0.024	0.024	0.000	1,109
	(0.000)	(0.007)	(0.023)	(0.000)	(0.000)	(0.310)	(0.310)		
Indigenous	0.058	0.022	0.064	-0.030	0.036	0.093*	0.028	-0.065	1,109
	(0.235)	(0.036)	(0.057)	(0.030)	(0.051)	(0.075)	(0.673)	(0.147)	

	Control mean	Any	FT	NG	NVT	FT=NG	FT=NVT	NG=NVT	N
Brahmin	0.110	0.012	-0.025	0.063	0.001	-0.088	-0.026	0.062	873
	(0.314)	(0.055)	(0.059)	(0.077)	(0.066)	(0.225)	(0.664)	(0.424)	
Chheriya	0.142	-0.025	0.037	-0.032	-0.074	0.069	0.111	0.042	873
	(0.350)	(0.060)	(0.072)	(0.067)	(0.078)	(0.302)	(0.155)	(0.562)	
Newar	0.018	0.006	-0.004	0.001	0.020	-0.005	-0.024	-0.019	873
	(0.135)	(0.013)	(0.012)	(0.017)	(0.020)	(0.748)	(0.215)	(0.406)	
Hill Janajati	0.399	0.031	0.030	0.034	0.031	-0.003	-0.001	0.003	873
	(0.491)	(0.094)	(0.109)	(0.126)	(0.129)	(0.979)	(0.997)	(0.984)	
Tharu	0.023	0.018	0.010	-0.023	0.062	0.033	-0.052	-0.085	873
	(0.150)	(0.034)	(0.035)	(0.014)	(0.079)	(0.305)	(0.539)	(0.281)	
Madeshi	0.073	-0.028	-0.012	-0.040	-0.031	0.028	0.019	-0.009	873
	(0.261)	(0.062)	(0.067)	(0.065)	(0.072)	(0.516)	(0.721)	(0.859)	
M. Dalit	0.005	0.027	0.000	0.005	0.072	-0.005	-0.072	-0.067	873
	(0.068)	(0.019)	(0.006)	(0.010)	(0.047)	(0.636)	(0.136)	(0.170)	
Gen. Dalit	0.142	-0.046	-0.067	-0.008	-0.061	-0.059	-0.005	0.054	873
	(0.350)	(0.053)	(0.056)	(0.074)	(0.059)	(0.358)	(0.908)	(0.425)	
Muslim	0.037	-0.020	0.015	-0.037**	-0.037**	0.052	0.052	0.000	873
	(0.188)	(0.024)	(0.053)	(0.017)	(0.017)	(0.305)	(0.305)	(1.000)	
Indigenous	0.050	0.023	0.016	0.036	0.018	-0.021	-0.002	0.018	873
	(0.219)	(0.027)	(0.036)	(0.039)	(0.041)	(0.656)	(0.965)	(0.710)	

OLS estimates of baseline differences between treatment and control groups. For each outcome, we report the coefficient of interest and clustered (VDC) standard errors in parentheses; Column (1) reports the mean and standard deviation of the outcome variable. Column (2) compares the effect of belonging to any treatment group to the control group. Column (3) measures the effect of belonging to the FT group at baseline. Column (4) measures the effect of belonging to the NG group at baseline. Column (5) measures the effect of belonging to the NVT group at baseline. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

C Results of Attrition Analysis

Table 18: Differential attrition by treatment

	Control mean (SD)	Any	FT	NG	NVT	N
Attrition	0.000	-0.006	0.031	-0.031	-0.017	1,982
	(0.000)	(0.029)	(0.034)	(0.033)	(0.030)	

Table 19: Differential attrition by baseline characteristics: demographic

	Attrited	N
HH Size	-0.141	1,981
	(0.290)	
N Male	0.005	1,942
	(0.159)	
Average Age	0.535	1,981
	(1.214)	
Resp. Age	1.490	1,980
	(1.052)	
Resp. Edu	0.114	1,980
	(0.370)	
Resp. Lit.	-0.041	1,979
	(0.054)	
Income	0.046	1,982
	(0.117)	
Land	0.075	1,977
	(0.073)	
TLU	-0.392	1,982
	(0.165)**	
Has Migrant	0.004	1,981
	(0.037)	

Table 20: Differential attrition by baseline characteristics: outcome indices

	Attrited	N
Empowerment	0.012	1,968
	(0.016)	
Finance	0.194	1,982
	(0.087)**	
Aspirations	-1.297	1,978
	(0.516)**	
Mental health	-0.075	1,982
	(0.098)	
Asset index	-0.050	1,982
	(0.111)	
Income (Rs.)	0.046	1,982
	(0.117)	
Non-food consumption	0.310	1,978
	(0.166)*	
Physical health	-0.130	1,982
	(0.082)	
Food Security	-0.103	1,982
	(0.116)	

Table 21: Attrited households: demographic

	Treatment mean	Any	FT	NG	NSC	N
HH Size	6.048	-0.453	-0.633	-0.488	-0.139	153
	(3.715)	(0.590)	(0.730)	(0.623)	(0.797)	
N Male	2.951	0.114	-0.010	0.174	0.268	148
	(2.133)	(0.348)	(0.428)	(0.355)	(0.438)	
Average Age	32.496	-1.828	-1.732	-0.275	-3.159	153
	(14.885)	(2.839)	(3.107)	(3.738)	(3.559)	
Resp. Age	43.643	-1.958	-1.699	-0.603	-3.400	153
	(13.842)	(2.109)	(2.302)	(2.932)	(3.316)	
Resp. Edu	3.071	-0.333	-0.637	0.609	-0.556	153
	(4.069)	(0.998)	(1.053)	(1.039)	(1.346)	
Resp. Lit.	0.524	-0.055	-0.165	0.196	-0.069	153
	(0.505)	(0.125)	(0.134)	(0.122)	(0.171)	
Income	11.221	-0.080	-0.297	0.178	0.066	154
	(1.156)	(0.260)	(0.340)	(0.325)	(0.282)	
Land	0.629	-0.138	-0.245	-0.036	-0.040	152
	(0.880)	(0.181)	(0.186)	(0.222)	(0.244)	
TLU	2.010	0.045	0.093	0.051	-0.033	154
	(2.065)	(0.465)	(0.477)	(0.592)	(0.627)	
Has Migrant	0.690	-0.087	-0.143	-0.010	-0.054	153
	(0.468)	(0.072)	(0.073)*	(0.118)	(0.109)	

Table 22: Attrited households: outcome indices

	Treatment mean	Any	FT	NG	NSC	N
HH Size	6.048	-0.453	-0.633	-0.488	-0.139	153
	(3.715)	(0.590)	(0.730)	(0.623)	(0.797)	
N Male	2.951	0.114	-0.010	0.174	0.268	148
	(2.133)	(0.348)	(0.428)	(0.355)	(0.438)	
Average Age	32.496	-1.828	-1.732	-0.275	-3.159	153
	(14.885)	(2.839)	(3.107)	(3.738)	(3.559)	
Resp. Age	43.643	-1.958	-1.699	-0.603	-3.400	153
	(13.842)	(2.109)	(2.302)	(2.932)	(3.316)	
Resp. Edu	3.071	-0.333	-0.637	0.609	-0.556	153
	(4.069)	(0.998)	(1.053)	(1.039)	(1.346)	
Resp. Lit.	0.524	-0.055	-0.165	0.196	-0.069	153
	(0.505)	(0.125)	(0.134)	(0.122)	(0.171)	
Income	11.221	-0.080	-0.297	0.178	0.066	154
	(1.156)	(0.260)	(0.340)	(0.325)	(0.282)	
Land	0.629	-0.138	-0.245	-0.036	-0.040	152
	(0.880)	(0.181)	(0.186)	(0.222)	(0.244)	
TLU	2.010	0.045	0.093	0.051	-0.033	154
	(2.065)	(0.465)	(0.477)	(0.592)	(0.627)	
Has Migrant	0.690	-0.087	-0.143	-0.010	-0.054	153
	(0.468)	(0.072)	(0.073)*	(0.118)	(0.109)	