Evaluation of the Welfare Impacts of a Livestock Transfer Program in Nepal
Midline Data Pre-analysis Plan

Sarah Janzen[1], Nicholas Magnan[2], Sudhindra Sharma[3], and William M. Thompson[2]

[1]Montana State University
[2]University of Georgia
[3]Nepa School of Social Sciences and Humanities

August 19, 2016

Abstract

This document describes the pre-analysis plan for analysis of midline data from a randomized controlled trial (RCT) evaluating Heifer International's (HI) Program in Nepal. Between September 2014 and September 2015, HI administered productive asset transfers to poor households living in three distinct agro-ecological zones of Nepal. The present RCT assigned three treatments at the village level: Treatment one delivered a full package of benefits to program participants, including physical capital (goats), human capital formation, and social capital formation. Treatment two withheld the physical capital from beneficiaries; treatment three withheld social capital formation. Within treatment groups we further subdivided the sample so as to capture indirect effects, both programmed and unprogrammed. The present document outlines the outcome variables and econometric methods we will use to assess the effect of the program on assets, income, non-food consumption, finance, time use, physical health, mental health, food security, aspirations, and women's empowerment.

Contents

List of Figures

1 Introduction

Social protection policies and programs have been widely heralded as important for addressing persistent poverty. Productive asset transfer programs, often involving livestock, are a particularly popular form of social protection for vulnerable populations, that may become even more vulnerable due to climate change and its effects on agriculture. Such programs are often supplemented with technical training that support human, financial capital development, and social mobilization. The collaborative research project, “Evaluating the Welfare Impacts of a Livestock Transfer Program in Nepal” seeks to disentangle the importance of physical (livestock) assets relative to human and social capital in the provision of social protection designed to permanently increase resiliency and improve nutritional and economic outcomes for the chronically poor in rural Nepal.

In this project, researchers from Montana State University, University of Georgia, International Food Policy and Research Institute, and Nepā School for Social Sciences and Humanities are partnering with the global leader in livestock transfer programs, Heifer International (HI), to evaluate the impact of a multifaceted social protection program developing physical (livestock), human, and social capital. A handful of recent studies have attempted to analyze the effectiveness of livestock transfer programs (Rawlins et al. (2014); Jodlowski et al. (2016); Kafle, Winter-Nelson, and Goldsmith (2016)), but few have been highly rigorous (one important recent exception is Banerjee et al. (2015), which evaluates large transfer programs in six countries). While most livestock transfer programs are multifaceted, similar to the HI program, this is the first study we are aware of that tests the impact of different program components.

This document describes the analysis plan for the randomized controlled trial (RCT) evaluating HI’s Smallholders in Livestock Value Chain (SLVC) Program. In many cases our pre-analysis approach follows the examples provided by the pre-analysis plans of Almeida et al. (2012) and Haushofer and Shapiro (2013).

2 Intervention

HI’s SLVC program targets poor households in rural Nepal, and seeks to provide sustainable livelihoods and a pathway out of poverty for its beneficiaries, focusing on women in particular. This study is a randomized control trial that evaluates an intervention set up to replicate the structure of the SLVC, and is being implemented in seven districts spread across two geographic regions of Nepal.

The standard HI intervention in Nepal consists of a package of benefits consisting of a gift of physical capital (in this case goats), human capital formation, and social capital formation. After identifying a location to receive the intervention, HI recruits an original group of beneficiaries (the “OG”). OG groups are nearly always formed of close neighbors (and often include most or all of the households in a given *tole*, or neighborhood, as described later). As a rule, HI considers all the households in targeted area to be objectively poor and therefore eligible for SLVC, allowing for the possibility that a considerable range of relative wealth and poverty might exist within a group. SLVC targets women between the ages of

20 and 40 in particular and gives them preference within a treated household, implying that SLVC beneficiaries are overwhelmingly female and are mostly young or middle aged.

Once selected, SLVC beneficiaries within a ward are organized into a self-help group (SHG), a crucial element of the social capital formation arm of the intervention. Over a period of months the SHG participates in a series of trainings. Trainings include (1) technical training (improved animal management, fodder/forage development, human and animal nutrition, home gardening, and community animal health worker training for some members), and (2) HI's values-based "Cornerstone" training ("passing on the gift" of livestock and training, accountability, sharing and caring, sustainability, self-reliance, income management, environmental stewardship, spirituality, self-help group management, and empowerment). The trainings culminate with the beneficiaries receiving a gift of livestock, the physical capital component, which in our study included two doe goats for each beneficiary and buck of improved stock (to facilitate a breeding program) for the SHG.

HI intends that the benefits of their programming become viral, spreading quickly and firmly taking root in all corners of a community. To that end, HI requires that members "pass on the gift" by recruiting new beneficiaries, providing them with a gift of livestock, and replicating the technical training. Cornerstone training is always primarily delivered by HI staff. These subsequent generations for beneficiaries are called "POGs". SLVC follows an innovation to the basic POG model, unique to the program and conceived by in-country staff, known as "exponential POG". Each OG SHG is tasked with recruiting up to five new SHGs, with the goal of full saturation and complete adoption of the improved practices and technologies encouraged in the Cornerstone trainings within a relatively short time frame.

While HI is widely credited with pioneering this style of anti-poverty intervention, numerous NGOs (large and small) have embraced the model, devoting enormous resources to productive asset transfers. Even so, rigorous evaluations of productive asset transfers were notably absent from the literature until quite recently (Banerjee et al. (2015) is an important exception), and evidence of their effectiveness and the mechanisms by which they achieve impact is still scant. This study seeks to establish the overall effect of the program and disentangle the relative effects the treatment components on a wide variety of welfare outcomes. Further, the design lets us examine indirect treatment effects.

3 Overview of study

3.1 RCT design

To establish a causal relationship between the program and changes in outcomes, this study uses a randomized control trial (RCT). The RCT design accommodates estimating and comparing treatment effects of several components of the multifaceted HI program. There are three different treatment arms in the evaluation. All three treatments share some common characteristics. The SLVC program provides benefits to groups, which HI helps form, so all beneficiaries are expected to acquire some level of social capital by being in a group. Group members are encouraged to contribute to group savings accounts, and 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 some support for home garden and fodder/forage production.

In order to unpack the benefits of the program, additional benefits received across treatment arms vary. Beneficiaries in the first and second treatment arm received the aforementioned Cornerstone trainings, which focus on values, soft skills, and empowerment. Beneficiaries in the first and third treatment arm received a gift of goats which included a pair of does provided to each OG group member, a shared buck of improved breeding stock for the self-help group (SHG), and support for goat shed improvement. The intervention includes the following treatment arms:

1. **Treatment 1**: group formation, savings encouragement, technical trainings on a variety of topics, small cash support for home gardens and fodder/forage production, values-based training, and a gift of goats.

2. **Treatment 2**: Identical to treatment 1, but without the gift of goats.

3. **Treatment 3**: Identical to treatment 1, but without values-based training.

A fourth arm was randomly selected as pure control.

3.2 Key data sources

We collected baseline and midline data from nearly 3,300 rural women eligible to participate in an asset transfer program across three regions of Nepal from June-September 2014. Owing to the scale, complexity, and logistical challenges associated with the field work, we worked with Interdisciplinary Analysts (IDA), a professional research firm based in Kathmandu with extensive experience with such surveys. Surveys were administered on Android tablets using CSPro survey software. Treatment and control households were interviewed in each round of data collection.

Between July and December 2014, HI administered training and delivered goats to the original beneficiaries. Various additional trainings continued throughout 2015. Midline data was again collected in June-July 2016 for the same households included in the baseline. The midline dataset was not analyzed in any fashion before the registration of this pre-analysis plan.

Endline data collection is anticipated in June-July 2017. This PAP outlines our plan for analysis of program impacts at midline only. At endline, we intend to analyze impacts on the same core outcomes using the same or similar indices. We will likely collect additional data depending on budget and in response to questions that arise from the midline analysis. We will also more carefully consider temporal dynamics of the treatment effects across outcomes. With this in mind, we will post an amended pre-analysis plan before analysis of endline data in 2017.

3.3 Sampling

Understanding the structure of the intervention, and thus the design of our RCT, requires a basic knowledge of how Nepal subdivides into administrative units. Nepal comprises 75 districts, which are relatively large. Districts are further subdivided into village development committees (VDCs), which should be thought of as clusters or groupings of villages within a district (similar to a county in the United States). VDCs are split into nine wards, and wards 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.

With assistance from in-country HI staff, we identified 60 VDCs that had not received any benefits from HI, but that would be good candidates for the SLVC. In each of the 60 VDCs, including controls, HI identified a ward (based on centrality) and *tole* for intended OG treatment. In this way, HI pre-identified potential targeted beneficiaries and SHGs (which were never ultimately formed), so that the individuals in the control arm are comparable with those in the treatment arms.

We then divided the 60 VDCs into three treatment groups and a control group. To achieve balance between 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, Middle Hills, and Terrai). 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.[1] This created new groups within the district groupings based on rank prevalence of caste/ethnic groups. Within these groups, we ordered VDCs by the proportion of the most prevalent caste/ethnic group, then second most prevalent, and so on. From this ordering we established 16 bins and randomly assigned treatment within each.[2]

Within each treatment VDC, we selected three wards to sample: one treated central ward, a second ward in close proximity to the treated ward to remained untreated ("near spillover ward"), and a third ward more distant from the treated ward to remain untreated ("far spillover ward"). Baseline respondents in central wards of treated VDCs were divided into two categories: potential OG members (as identified by HI), and prospective POG members. In each of these central wards, we included all prospective OG members in the sample (around 25 per ward). Because OGs recruit POGs, there was no way to compile a list of potential POG beneficiaries ahead of baseline; all the households in a ward that did not appear on the OG list were potential POGs. Therefore, 15 potential POG households were drawn at random from a complete roster of all the households in the treated ward, after removing households OG households. Because of the aggressive nature of the exponential POG model, we expect that many (if not most) of these households will actually become

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

[2] Because of the unequal number of VDCs in each district grouping there was one bin with two VDCs, two with three VDCs, and 13 with four VDCs. Because of the uneven bin sizes, random treatment assignment within bins resulted in 16 VDCs in treatment 1, 16 VDCs in treatment 2, 15 VDCs in treatment 3, and 13 VDCs in the control group. To obtain equally sized treatment arms we randomly drew one VDC each from treatments 1 and 2 to be placed in the control group.

POG members. Each spillover ward (near and far) contains 10 households selected at random from a full listing of households in the ward.

We sampled from central wards in control VDCs in exactly the same manner as we did from central wards in treatment VDCs: 25 potential OGs (as identified by HI), and 15 potential POGs (drawn at random). We did not, however, sample spillover wards in control VDCs due to budgetary constraints. We do not view this as overly problematic because there is no strong reason to believe that central wards are different from periphery wards in any way (centrality refers purely to geography, not to an economic or political center). In section 5.3 we discuss potential implications of this sampling strategy on our estimation. Figure 1 depicts the sampling structure at the VDC and ward level, and figure 2 depicts the sampling structure within the various ward types.

Our total baseline sample is 3,283 women (960 treatment OGs, 828 treatment POGs, 326 control OGs, 261 control POGs, and 908 households in spillover wards) across 60 treatment clusters (VDCs) stratified by region and ethnic composition. Shortly after HI delivered training and livestock to the original beneficiaries of the project, a devastating earthquake struck Nepal. The earthquake greatly affected 10 VDCs that are part of the evaluation, spread evenly across treatment groups and control. These 10 VDCs were dropped from the RCT, and HI was allowed to intervene in whatever ways they felt were suitable. The remaining sample size is 2,724. Updated power calculations suggest the study remains sufficiently powered. Despite presenting a setback for our original research design, this exogenous shock to a subset of this poor population (for whom we have a rich baseline dataset) provides an opportunity to study the relationship between ex ante empowerment and resiliency to shocks. We continued collecting data in the affected areas for a secondary analysis of resilience, not included in this PAP.

3.4 Mitigating contamination

Treatments were assigned at the VDC level. Because components of the treatment entail human capital and social capital formation, both of which could be transferred to others, we must consider the possibility of contamination. 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 low 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 more 'central' wards within a VDC to be the treated ward. In this way, we ensure an additional degree of isolation and further reduce the prospect of unintentional spillovers that could bias results.

Anticipating the potential for contamination, we included basic information on social, family, and business relationships between neighboring VDCs in each round of data collection. With respect to each of the three VDCs closest to their own, respondents answered questions about the number of contacts in the VDC, and the nature of the contacts (family, social, commercial). We will use these data to identify ex-ante connectivity between VDCs that could result in bias induced by spillovers between treatment arms.

4 Evaluation questions and hypotheses

Our main questions are: (i) across a range of dimensions of household and individual welfare, what are the overall impacts of a productive asset transfer that simultaneously develops human, physical, and social capital, (ii) what are the specific impacts of each aspect of the intervention, (iii) within a treated village, are treatment effects robust or attenuated across generations of beneficiaries (this refers to direct effects and intentionally programmed indirect effects, described in detail below), (iv) what is the magnitude, if any, of geographic spillovers (this refers to indirect effects not explicitly programmed, occurring through social learning and similar channels, described in detail below), (v) are some households more likely than others to benefit from the program, directly or indirectly, and (vi) which package of benefits results in the most cost-effective improvements to household and individual well-being?

The RCT design, coupled with a rich dataset, accommodate testing a number of hypotheses regarding the impact of the SLVC program. Our hypotheses are outlined below.

4.1 Direct impact of each treatment on a variety of outcomes

Each intervention is likely to have a positive impact on a variety of outcomes for selected beneficiaries. We group these outcomes across 10 dimensions:

1. Assets
2. Income
3. Non-food consumption
4. Finance
5. Time use
6. Physical health
7. Mental health
8. Food security
9. Aspirations
10. Empowerment

Details regarding how these outcomes are measured are discussed in section 7. We also anticipate an impact on resilience, but in light of the earthquake discussed above, the analysis of the impact on resilience is considered separately from this PAP. Given the nature of the program and the emphasis on livestock production, we may also be interested in studying livestock production decisions and outcomes. To accommodate a study of the underlying mechanisms regarding differential treatment effects, we will analyze a variety of indicators related to livestock production. A list of sub-indicators related to livestock production are included in Appendix A, but these indicators are not included in the primary welfare analysis.

4.2 Differential impacts by treatment

The impact is likely to vary by treatment type (T1, T2, T3) for many outcomes. Whether impact differs and by how much is important for gauging cost effectiveness because the treatments have substantially different costs.

4.3 Indirect effects

Each intervention is likely to result in indirect effects through two channels:

1. The "Pass on the Gift" requirement insists that OGs transfer benefits to POG beneficiaries
2. Non-beneficiaries may learn from or immitate the actions of beneficiaries based on observation or social contact with treated individuals or groups.

We anticipate possible indirect effects within treatment wards to potential POGs, and to spillover wards both near and far.

4.4 Heterogeneity of impacts

The individual and household characteristics of beneficiaries (and in the case of indirect effects, non-beneficiaries) may determine whether they differentially benefit from the intervention. We hypothesize heterogeneity of impacts across the following dimensions:

1. Assets (quartiles of asset index)
2. Income (quartiles of respondent income and household income)
3. Empowerment of respondent (high/low dummy variable based on a natural break in the women's empowerment summary index, if one exists, or the median)
4. Literacy of respondent and head of household (dummy variable if self-identified as literate)
5. Gender (applicable only for three individual - rather than household - outcomes: child health, aspirations for child education, and individual incomes)
6. Ward population/size (applicable for indirect effects, see below)

It is likely that indirect effects to POG households (in the same ward as treated OG households) will be stronger in smaller wards than larger ones because a greater proportion of sampled individuals will have received training, goats, or beneficial contact with OGs. We will therefore test for heterogenous impacts of indirect effects within wards by quartiles of number of households in the ward.

Details regarding how the first three characteristics are measured are discussed in section 7.

5 Econometric specifications

5.1 Estimation of direct treatment effects

First, we specify our direct treatment effects model to measure the impacts of each of the three treatments. For simplicity, in this equation we consider a sample of only potential OG households (in both treatment and control VDCs), and turn to indirect effects in the following section. The regression specification for direct treatment effects is:

$$y_{hvm} = \beta_0 + \beta_1 T1_{hv}^{OG} + \beta_2 T2_{hv}^{OG} + \beta_3 T3_{hv}^{OG} + \delta y_{hvb} + \mathbf{X_{hvb}}\gamma + \mathbf{S_{vb}}\rho + \varepsilon_{hv} \tag{1}$$

where y_{hvm} is the outcome of interest for household h in village v, measured at midline ($t = m$). Treatment indicator variables ($T1_{hv}^{OG}$, $T2_{hv}^{OG}$, and $T3_{hv}^{OG}$) take a value of 1 for OG households in wards selected to receive any of the previously described treatments ("treated households"), and a value of 0 otherwise. The omitted category is control OG households located in pure control villages ("control households"). In order to improve statistical power, we will 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. We will cluster errors at the VDC level, as this is the level of treatment.

The "intent to treat" (ITT) treatment effects of interest are β_1, β_2, and β_3. β_1 represents the ITT treatment effect on OG households selected to receive the full treatment package (T1) when compared to OG households in pure control VDCs, β_2 identifies the ITT treatment effect on OG households selected to receive the second treatment package (T2), and β_3 identifies the ITT treatment effect on OG households selected to receive the third treatment package (T3).

An important aspect of our evaluation is to test whether the treatments effects vary across treatment type. To do this we will conduct Wald tests for $\beta_1 = \beta_2$, $\beta_1 = \beta_3$, and $\beta_2 = \beta_3$.

5.2 Estimation of indirect effects–treatment wards

Next we test for spillover effects within the ward from OG households to POG households. Under HI's "Passing of the Gift" strategy, HI expects some fraction of potential POG households within the same ward to receive similar, or even identical treatment as the original beneficiaries "pay it forward" through training their peers and giving them the gift of livestock (as well as receiving values-based training from HI staff). For the SLVC program, and the parallel evaluation, HI implemented its novel "Exponential Passing of the Gift" (EPOG) strategy, where each group of OGs is supposed to form five PGO groups. The goal of this strategy is for HI's impact to occur more broadly and rapidly (more households reached in a short period of time) than under the traditional HI (non-exponential) model. However, the exponential model may also be a less intense treatment because households receive benefits over a longer period of time (livestock get passed down much more slowly). Without looking at the midline data, we cannot say to what degree the EPOG succeeded in bringing in more beneficiaries. Given the time between when OG households received benefits and when midline data was collected, it is possible that all potential POG households will not yet have been passed on the gift, and are still waiting to receive the intended POG benefits.

Even if potential POG households do not yet receive any benefits, spillover effects may occur through a second channel: households may simply observe or discuss techniques or other concepts learned through the trainings. If households replicate these techniques, they may benefit indirectly from the HI trainings, even if they are receive no goats or formal training. Because this second type of spillover effect is possible, estimation of local average treatment effects (LATE) is not possible. We therefore estimate ITT effects, keeping in mind that they may be very conservative, especially using outcomes at midline. We will also calculate and report the proportion of POG households actually receiving benefits by midline data collection. The regression model for indirect treatment effects within treatment wards compares potential POG treatment households to the corresponding POG control households (and excludes all OG and spillover households). The regression specification is:

$$y_{hvm} = \beta_4 + \beta_5 T1^{POG}_{hv} + \beta_6 T2^{POG}_{hv} + \beta_7 T3^{POG}_{hv} + \delta y_{hvb} + \mathbf{X_{hvb}}\gamma + \mathbf{S_{vb}}\rho + \varepsilon_{hv} \qquad (2)$$

Here again we will condition on baseline levels of the outcome of interest y_{hvb}, a vector of control variables X_{hvb} for which an imbalance at baseline was observed across treatments, and vector S_{vb} of stratification bin dummies. We will cluster standard errors at the VDC level.

In this specification, β_5 corresponds to the T1 ITT effect of being a potential POG in a T1 treatment ward, β_6 captures the T2 ITT effect of being a potential POG in a T2 treatment ward, and β_7 captures the T3 ITT effect of being a potential POG in a T3 treatment ward. We will test for whether the treatments have different indirect effects within wards using Wald tests for $\beta_5 = \beta_6$, $\beta_5 = \beta_7$, and $\beta_6 = \beta_7$.

5.3 Estimation of indirect effects–spillover wards

In discussions with HI staff during the design of the evaluation, HI expressed strong priors that it would be inappropriate to use households in untreated wards within treated VDCs as a comparison group due to the strong spillover effects of their programs. They were interested in testing this prior, so we incorporated "spillover wards" into our study. There are two issues that make estimation of spillover effects to other wards less clean than the estimation of treatment effects on OG and POG households. Both of these issues stem from the fact that the only comparison group we have to compare households in spillover wards against are those in the central ward of control VDCs. Like treated wards in treated VDCs, sampled control wards were chosen due to their central location within a VDC. If these more centrally located wards differ from other wards, it could induce bias in our estimates.[3]

Second, in the central wards in both treatment and control VDCs, HI identified the group of OG households based on natural (primarily geographic) delineations in the ward. This process was costly, and not possible to do in spillover wards. Sampling in spillover wards mimics sampling of POGs (selected from a list of all households in the ward) - except that in the selection of POGs, OG households had been removed from the list. Using an assumption of 150 households per ward, the composition of the sampled households in spillover wards will contain approximately 83 percent households that would have been POG. The samples in treatment and control wards consist of approximately 38 percent POG households. If OG and POG households differ at baseline, we will use probability weighted regression (using a probability weight of $\frac{\text{ward size}-\text{OGs sampled from ward}}{\text{POGs sampled from ward}}$ for POG households in control wards) to make the spillover wards comparable to the control wards in terms of their OG/POG makeup.

The following specification considers the indirect effects on households in spillover wards. For each treatment, we compare households in spillover wards (either close or far) to control households. We then control separately for close spillover wards, to test for differential spillovers based on geography. The econometric specification is:

$$\begin{aligned} y_{hvm} = \beta_8 &+ \beta_9 T1^{S}_{hv} + \beta_{10} T2^{S}_{hv} + \beta_{11} T3^{S}_{hv} \\ &+ \beta_{12} T1^{close}_{hv} + \beta_{13} T2^{close}_{hv} + \beta_{14} T3^{close}_{hv} \\ &+ \delta y_{hvb} + \mathbf{X_{hvb}}\gamma + \mathbf{S_{vb}}\rho + \varepsilon_{hv} \end{aligned} \tag{3}$$

In this specification, β_9 corresponds to the T1 ITT effect of being in any untreated spillover ward (near or far) in a T1 VDC, β_{10} corresponds to the T2 ITT effect of being in any untreated spillover ward in a T2 VDC, and β_{11} corresponds to the T3 ITT effect of being in any untreated spillover ward in a T3 VDC. In addition, β_{12} corresponds to the additional effect of being in the near spillover ward for T1, β_{13} corresponds to the additional effect of being in the near spillover ward for T2, and β_{14} corresponds to the additional effect of being in the spillover ward near to a T3 treatment ward. We will test for whether the treatments have different indirect effects across wards using Wald tests for $\beta_9 = \beta_{10}$, $\beta_9 = \beta_{11}$, and $\beta_{10} = \beta_{11}$ for the effect of being in any spillover wards, and $\beta_9 + \beta_{12} = \beta_{10} + \beta_{13}$, $\beta_9 + \beta_{12} = \beta_{11} + \beta_{14}$, and $\beta_{10} + \beta_{13} = \beta_{11} + \beta_{14}$ for the additional effect of being in a near spillover ward.

Here again we will condition on baseline levels of the outcome of interest y_{hvb}, a vector of control variables X_{hvb} for which an imbalance at baseline was observed across treatments, and vector S_{vb} of stratification bin dummies. We will cluster standard errors at the VDC level. Because of the issues of comparability between households in spillover wards and households in control wards, we view this analysis as exploratory.

[3]We thank Craig McIntosh for bringing this point to our attention.

5.4 Estimation of heterogeneous treatment effects

The individual and household characteristics of beneficiaries (and in the case of spillovers - non-beneficiaries) may determine whether they are differentially impacted by the intervention. We will test whether the effects of each treatment vary with baseline levels of the characteristics specified in section 4.4. For income, assets, and ward size we will consider quartiles. For empowerment, literacy and and gender we will differentiate between two types of individuals (empowered/not, literate/not, and female/male).

We will run the following specification separately for OG households, POG households, and spillover households. We will not differentiate between near and far spillover wards. In this specification, z_{hvb} represents an indicator for one of the binary indicators along which we expect heterogeneous treatment effects. A similar equation can be estimated using multiple quartile dummies for non-binary indicators. We will also estimate the analog of the equation below to accommodate analysis of heterogeneous spillover effects following the same structure employed in equations 2 and 3.

$$\begin{aligned} y_{hvm} = \beta_0 &+ \beta_1 T1^g_{hv} + \beta_2 T2^g_{hv} + \beta_3 T3^g_{hv} \\ &+ \beta_4 (T1^g_{hv} \times z_{hvb}) + \beta_5 (T2^g_{hv} \times z_{hvb}) + \beta_6 (T3^g_{hv} \times z_{hvb}) \\ &+ \lambda z_{hvb} + \delta y_{hvb} + \mathbf{X_{hvb}} \gamma + \mathbf{S_{vb}} \rho + \varepsilon_{hv} \end{aligned} \tag{4}$$

for $g \in [OG, POG, S]$.

As before, we cluster the standard errors at the VDC level. For gender-disaggregated impacts we only consider the three sub-outcomes for which we have gender-disaggregated data: child health, aspirations for child education, and individual incomes. For all other heterogeneous impacts we do not rule out possible heterogeneous effects across all hypothesized outcome dimensions.

The treatment effects on the omitted category are β_1, β_2, and β_3. Treatment effects for households that take on a value of one for z_{hvb} are $\beta_1 + \beta_4$, $\beta_2 + \beta_5$, and $\beta_3 + \beta_6$. We will test for whether the treatments have different impacts on sub-groups using Wald tests for $\beta_1 = \beta_2$, $\beta_1 = \beta_3$, $\beta_2 = \beta_3$, $\beta_1 + \beta_4 = \beta_2 + \beta_5$, $\beta_1 + \beta_4 = \beta_3 + \beta_6$, and $\beta_2 + \beta_5 = \beta_3 + \beta_6$.

5.5 Accounting for multiple inference

We have a rich dataset and a large number of hypotheses regarding behavioral change and improved welfare across several dimensions. We will account for multiple hypotheses in two ways. First, we will construct one primary summary index for each dimension of welfare described in section 4.1. Each summary index consolidates several individual tests into a single test. We describe how each index will be constructed in section 7. Second, because we still have multiple outcome dimensions, we will report naive p-values and adjusted q-values that control for the false discovery rate (FDR). Specifically, we will 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).

We prefer controlling for FDR over controlling for the the family-wise error rate (FWER) because we are testing a large number of 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). Overall conclusions about the SLVC program effectiveness depend on many outcomes– the overall conclusion should not be that the intervention is ineffective because of one erroneously rejected null hypothesis– so 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).

For gender-disaggregated impacts we only consider thee sub-outcomes for which we have gender-disaggregated data: child health, aspirations for child education, and individual incomes; we will report naive p-values for those regressions. For other heterogenous treatment effect analysis (across income and assets quartiles, as well as empowerment and literacy) we will only examine impacts on summary indices. For these, we will again report naive p-values and q-values that control for the FDR. Specifically, we will will calculate q-values for multiple hypothesis tests across summary indices and interaction terms within treatments, but not across treatments, using the Benjamini and Hochberg (1995) step-up method.

To test treatment groups against each other we will conduct Wald tests as described in sections 5.1 through 5.4. For these tests, we will report both naive p-values and q-values that control for FDR. As above, we will will 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.

5.6 Questions with limited variation

In order to limit noise caused by variables with minimal variation, questions for which 95 percent of observations have the same value within the relevant sample will be omitted from the analysis. If doing so makes it impossible to calculate a proposed indicator, the indicator will be not be calculated.

5.7 Survey attrition

There is some risk of attrition in the sample. When conducting multiple rounds of data collection at similar intervals and of a similar sample, IDA normally expects to observe attrition rates in the 10% range. We will assess the seriousness of any attrition using the same three approaches outlined in the pre-analysis plan by Haushofer and Shapiro (2013), and adapted to our design. The approach presented below considers the full sample, but we can similarly assess attrition for OG, POG, and spillover households separately.

First, equation 5 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.

$$attrit_{hv} = \beta_0 + \beta_1 T1_{hv} + \beta_1 T2_{hv} + \beta_1 T3_{hv} + \varepsilon_{hv} \tag{5}$$

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 stats as estimated in equation 6:

$$y_{hv} = \beta_0 + \beta_1 attrit_{hv} + \varepsilon_{hv} \tag{6}$$

Finally, equation 7 estimates the extent to which baseline characteristics of treated households differ from control households, after restricting the sample to attrited households:

$$(y_{hvB}|attrit_{hv} = 1) = \beta_0 + \beta_1 T_{hv} + \varepsilon_{hv} \tag{7}$$

If attrition is deemed problematic, we will adjust all specifications for attrition using Lee bounds (Lee, 2009).

5.8 Missing data from non-response

For our analysis we will use a combination of sub-indicators and summary indices. The ten summary indices at midline are our main outcomes of interest, and their baseline values will serve as control variables. Sub-indicators from midline will be used as additional outcomes of interest, and their baseline values will likewise serve as control variables. Indicators that show imbalance across treatments will also be used as controls.

We will not impute any index values. In most cases, the index can be computed without the full complement of its components. When this is not possible, we will use median or mode imputation for components necessary to construct the index. We expect this to be necessary for very few cases.

In our analysis using sub-indicators as outcomes, observations with missing values at midline will be dropped from the sample. In these cases we will check whether missing data is correlated with treatment status following the same procedures as for survey attrition. If baseline values are missing, which will be used as control variables, we will use multiple imputation. In general, if any control variables are missing, we will use multiple imputation.

For missing numerical values that we deem are most likely zero, we will impute zero. For missing categorical variables for asset ownership, we will also impute zero.

6 Cost-benefit analysis

We collected detailed information from HI on the costs of implementing the three treatments. Some costs are common across all treatments, while others (such as the cost of the gift of goats) are unique to certain treatments. Income and expenditures for consumption are easily quantifiable benefits, and will be used for the cost-benefit analysis. Because program benefits potentially accrue over a long time horizon, we will conduct the cost/benefit ratio with sensitivity analysis regarding the duration of impact, the discount rate, and the rate of inflation.

The "Passing of the Gift" is a core aspect of HI's programs. Cost-benefit analysis therefore must account for both direct (OG) and indirect (POG) effects. We have a full sample of OG households, and can easily aggregate the benefits to the ward level. For POG households, we have only a partial sample. Because the sample of POG households is random (after removing OG households), we can safely assume that the average benefits are the same for each POG household. We will then multiply the per household by the average ward population to aggregate POG benefits.

Whether or not to include spillover benefits is a nuanced question. The SLVC program as currently administered by Heifer is rolled out in every ward of a VDC enrolled in the program (which would have increased the risk of contamination of controls in this study). Therefore, spillover households within a VDC do not exist under the SLVC. However, if spillover effects are strong, then Heifer may be able to improve their cost effectiveness by rolling out their programs in fewer wards per VDC and across more VDCs (less intense with wider geographical span). Therefore we will conduct cost benefit analysis both with and without the inclusion of spillover households. If spillover households do increase their income and/or consumption as a result of the program, then we need to multiply the per household benefits by the number of total spillover households in a VDC, which will be quite large (8 non-central wards with approximately 150 households each). This should not artificially inflate the benefits because households were sampled at random from spillover wards.

It is possible that HI's program will take time to result in increases in income and expenditures as herds grow, behaviors changes, and financial inclusion improves. Therefore we anticipate being able to conduct a more accurate and complete cost-benefit analysis at endline.

7 Indicators for each outcome of interest

In this section we list the indicators for each hypothesized outcome, grouped across ten dimensions of welfare. For each dimension we include a variety of sub-indicators (grouped by categories), one or more summary indices, and one primary summary index. The primary summary index is listed in bold. Some, but not all, sub-indicators are used to construct the summary index for each dimension. Where we indicate that an index will be a "weighted standardized average of variables," we follow the method outlined in Anderson (2008).

7.1 Assets

Within the assets dimension of welfare, we consider 5 sub-categories with 5 total sub-indicators. All variables included in the construction of sub-indicators are also listed below.

1. Productive assets (draft animals, plows, grinder, thresher, loom, sewing inventories, mechanical tools, and other productive assets.)
    (a) Principle components index of productive assets
2. Non-productive assets (radios, cassette recorders, DVD players, televisions, telephones (landline), mobile phones, heater/pressure lamps, electric fans, camera/camcorder, computer, furniture, rugs, clocks, jewelry and watches)
    (a) Principal components index of non-productive assets
3. Livestock (goats, cattle, water buffalo, swine, chickens)
    (a) Number of tropical livestock units
4. Land
    (a) Hectares of land
5. Housing (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)
    (a) Principal components index of housing

***Asset Summary Index*: weighted standardized average of indicators (1a), (2a), (3a), (4a) and (5a)**.

7.2 Income

Within the income dimension of welfare, we consider 6 sub-categories with 11 total sub-indicators. 2 summary indices are proposed: one for individual income and one for household income. Individual income (including the individual income summary index) will only be considered in the analysis of heterogeneous treatment effects across gender. Only household-level income is included in the primary summary index for income. Logged values of income sub-indicators and summary indices will be used in regression analysis.

1. Income from livestock and livestock products
    (a) Household income from livestock and livestock products
    (b) Individual income from livestock and livestock products
2. Income from crops
    (a) Household income from crops
    (b) Individual income from crops
3. Income from permanent employment (salaried, public sector, formal employment)

   (a) Household income from permanent employment
   (b) Individual income from permanent employment

4. Income from enterprise activities
   (a) Household income from enterprise activities
   (b) Individual income from enterprise activities

5. Other income including day labor
   (a) Other household income
   (b) Other individual income

6. Remittances
   (a) Household remittances

Individual Income Summary Index: sum of total individual income (1b), (2b), (3b), (4b), and (5b).

***Income Summary Index*: sum of total household income (1a), (2a), (3a), (4a), and (5a)**.

7.3 Non-food consumption

Within the non-food consumption dimension of welfare, we consider 3 sub-categories with 6 total sub-indicators. For consumption categories that were not reported annually, we multiply the monthly or quarterly figures by the appropriate factor to achieve an annualized amount. 2 summary indices are proposed: one for household consumption, and one for per capita consumption. For per capita consumption sub-indicators (and summary index) we divide all annualized amounts by household size. Given the nature of the program, we also consider total livestock related expenditures (not per capita). Logged values of non-food consumption sub-indicators and the summary index will be used in regression analysis.

1. Medical (includes medicines and medical supplies, medical consultation and treatment fees, laboratory and diagnostic fees, visits to traditional healers)
   (a) Annual medical expenditures per capita
   (b) Annual medical expenditures

2. Clothing (includes adult women's clothing, children's clothing including school uniforms and shoes, materials and tailoring items for clothing made at home)
   (a) Annual clothing expenditures per capita
   (b) Annual clothing expenditures

3. Miscellaneous (includes transportation to/from school, rent, kitchen equipment (cutlery, pots, plates, etc), jewelry and ceremonial items, roofing, walls, painting, gifts and donations, weddings, funerals, ceremonies, and festivals)

(a) Annual miscellaneous expenses per capita

(b) Annual miscellaneous expenses

Non-food Household Consumption Summary Index: sum of annual non-food consumption (1b), (2b), and (3b).

***Non-food Consumption Summary Index*: sum of annual non-food consumption per capita (1a), (2a), and (3a)**.

7.4 Finance (borrowing, saving, investment, planning

Within the finance dimension of welfare, we consider 3 sub-categories with 6 total sub-indicators.

1. Savings
   (a) Logged amount saved last month
   (b) Belongs to a savings group (dummy variable)
2. Credit
   (a) Logged amount outstanding debt, formal lender
   (b) Logged amount outstanding debt, informal lender
3. Future-oriented preferences
   (a) Discount rate, following Ashraf, Karlan, and Yin (2006)
   (b) Planning horizon, following Laajaj et al. (2012)[4]

***Finance Summary Index*: weighted standardized average of variables (1a), (1b), (2a), (2b), (3a), and (3b)**.

7.5 Time use

Within the time use dimension of welfare, we consider 2 sub-categories with 4 total sub-indicators. Time use data was collected differently at baseline and midline so they are not directly comparable.

1. Time spent working
   (a) Time spent working in agriculture
   (b) Time spent working in livestock
   (c) Time spent working on other productive activities

[4] Ordered categorical variables for (0) Do not plan ahead, (1) plan ahead one week, (2) plan ahead one month, (3) plan ahead 6 months.

2. Time spent at leisure
    (a) Time spent at leisure

***Time Use Summary Index*: sum of total time spent working (1a), (1b), and (1c)**.

7.6 Physical health

Within the physical health dimension of welfare, we consider 3 sub-categories with 9 total sub-indicators. 5 sub-indicators are child-specific and not directly included in the primary summary index. Overall, 3 summary indices are proposed. Child-specific indicators, including one child-specific index, will only be considered in the analysis of heterogeneous treatment effects across gender. Only household-level outcomes are included in the primary summary index for physical health.

1. Respondent health
    (a) Days of work missed due to illness in the past month (respondent-only)
    (b) General personal health respondent (Answer to: "How healthy is the respondent" on a 1-10 scale, self-assessed, respondent-only)
2. Child health
    (a) Days of school missed by school-age child due to illness in the past month (child-level)
    (b) Average days of school missed per school-age child due to illness in the past month (household-level)
    (c) General child health (Answer to "How healthy are the children in the household" on a 1-10 scale, assessed by respondent for all children, household-level)
3. Child anthropometrics
    (a) Body Mass Index, (child-level, under age 5 only)
    (b) Stunting (WAZ), (child-level, under age 5 only)
    (c) Wasting (HAZ), (child-level, under age 5 only)
    (d) Mid-upper arm circumference (MUAC), (child-level, under age 5 only)[5]

Children's Anthropometrics Summary Index - Child-level: weighted standardized average of variables (3a), (3b), (3c), (3d), constructed for every child under age 5

Children's Anthropometrics Summary Index - Household-level: average of all child-level children's anthropometrics indices within a household, constructed for every household with at least one child under age 5

***Physical Health Summary Index*: weighted standardized average of variables (1a), (1b), (2b), (2c) and the Household-level Children's Anthropometrics Index**.

[5] not available at baseline

7.7 Mental health

Within the mental health dimension of welfare, we consider no sub-categories and 9 total sub-indicators.

1. Mental health
   (a) Depression (modified CES-D)
   (b) Locus of control (modified Rotter)
   (c) Worries (WVS)
   (d) Optimism (WVS)
   (e) Trust (WVS)
   (f) Happiness (WVS)
   (g) Self esteem (WVS)
   (h) Life Satisfaction (WVS)

We measure depression with an abbreviated version of the CES-D scale (Radloff, 1977), abstracting a cross-section of four questions from the full questionnaire and adapting them to the local context. Questions are meant to probe for syptoms of depression, and have four possible answers which are scored between zero and three. The respondent's CES-D score is an aggregate of their responses,hence the range of possible scores is the interval (0,12), where higher scores are more depressed.

We measure locus of control with an abbreviated Rotter (1966) scale. We ask a series of six binary-response questions that elicit locus of control. Responses that suggest a strong internal locus receive one point, responses suggesting a weak internal locus receive zero. Therefore, the range of possible score is the interval (0,6), with higher scores indicating a stronger internal locus of control.

We include relevant questions from the the World Values Survey (2009) to quantify measurements of optimism, trust, happiness, life satisfaction, worries and self esteem. Optimism, worries and trust are captured by similar mechanisms: we ask three or four questions with Likert scale responses coded zero through three, and sum across the responses with higher scores indicating more optimistic, less worried, more trusting respondents. Happiness is captured in a single question with responses coded one through four, where one is 'not happy at all' and four is 'very happy'. Respondents report life satisfaction on scale of one to ten.

***Mental Health Summary Index*: weighted standardized average of variables (1a), (1b), (1c), (1d), (1e), (1f), (1g), (1h)**.

7.8 Food security

Within the food security dimension of welfare, we consider 3 sub-categories and 8 total sub-indicators.

1. Meals eaten
    (a) Normal number of meals per day, family
    (b) Normal number of snacks per day, family
2. Meals skipped
    (a) Skipped at least one meal in past week, respondent
    (b) Skipped at least one meal in past week, children
    (c) Went a full day without eating in past week, respondent
    (d) Went a full day without eating in past week, children
3. Dietary diversity
    (a) Household dietary diversity score (HDDS)
    (b) Children dietary diversity score (CDDS)

To measure dietary diversity, we calculate a modified version of a Household Dietary Diversity Score (HDDS). A typical HDDS sums the number of categories (out a of a list of 12) from which a household consumed at least one food item over the past 24 hours. The dietary diversity data we collected spans nine categories (cereals, tubers, vegetables, fruit, meat, eggs, fish, legumes/nuts/seeds, dairy), and asks for the total number of meals where that category was consumed over the past three days. Since the typical Nepali family consumes two main meals per day, the range of possible values is (0, 54). We calculate a modified HDDS outcome for the full household as well as for children only.

***Food Security Summary Index*: weighted standardized average of variables (1a), (1b), (2a), (2b), (2c), (2d), (3a), and (3b)**.

7.9 Aspirations

Within the aspirations dimension of welfare, we employ the aspirations index of Bernard and Taffesse (2014). We consider 4 sub-categories and 6 total sub-indicators. For the construction of our summary index, we employ individualized weights elicited for each aspirational dimension following Bernard and Taffesse (2014). Weights are assigned by respondents through the distribution of 20 tokens across four bins in proportion to how heavily they value a particular dimension.

1. Income
    (a) stated aspiration for annual income
2. Assets and wealth
    (a) stated aspiration for value of home and land
3. Education

(a) stated aspiration in number of years of education for children
(b) stated aspiration in number of years of education for male children
(c) stated aspiration in number of years of education for female children

4. Status

(a) stated aspiration for number of people in community whom would seek one's advice

***Aspirations Summary Index*: standardized weighted average of (1a), (2a), (3a), and (4a) following Bernard and Taffesse (2014).**

7.10 Women's empowerment

Within the women's empowerment dimension of welfare, we employ the Women's Empowerment in Agriculture Index (WEAI) and the related "abbreviated" WEAI (A-WEAI) as described in Alkire et al. (2012) and Malapit and Quisumbing (2015). The WEAI and A-WEAI are composed of five sub-categories[6], and six sub-indicators. For the construction of our summary index, we weight each sub-indicator according to the A-WEAI assigned weight. Each binary sub-indicator is one if the respondent achieves "adequate" empowerment, and zero otherwise. Below we list each sub-category, the corresponding sub-indicators, our modified definitions for adequacy, and sub-indicator weights. Each definition of adequacy is adjusted for our data, but based on the A-WEAI. Some sub-indicators related to women's empowerment were collected differently at baseline and midline so these sub-indicators and the summary index are not directly comparable over time. This is particularly true for the sub-categories of production and time.

1. Production

(a) Input in productive decisions: A respondent is adequately empowered in 1(a) if she has at least some input into at least one production decision. For the baseline indicator, productive decisions include those regarding livestock and productive assets management and care. Due to survey changes at midline, productive decisions at midline include those regarding livestock and productive activities. (1/5)

2. Resources

(a) Ownership of assets: Adequate ownership means that the household owns at least one asset, and that the respondent (individually) has at least some ownership of one asset. (2/15)
(b) Access to and decisions on credit: A respondent is adequate in this sub-indicator if the household has at least some credit and the respondent participated to any extent in the decision to borrow. (1/15)

3. Income

[6] WEAI refers to them as dimensions

   (a) Control over use of income: Adequacy in control over income means that 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. (1/5)

4. Leadership

   (a) Group membership: A respondent is adequately empowered in this sub-indicator if she is a member of any group. (1/5)

5. Time

   (a) Workload: A respondent is adequate for 5(a) if she worked 10.5 hours or less in the previous 24 hours. Note that the survey questions about time use are different between baseline and midline, but we can aggregate the midline time use responses such that the data are similar. (1/5)

***Women's Empowerment Summary Index*: weighted sum of indicators (1a), (2a), (2b), (3a), (4a), and (5a) following the Abbreviated Women's Empowerment in Agriculture Index**.

References

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.

Almeida, R., S. Hirshleifer, D. McKenzie, C. Ridao-Cano, and A.L. Yener. 2012. "Impact of vocational training on the unemployed in Turkey: preanalysis plan." Unpublished.

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.

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.

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

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.

Laajaj, R., et al. 2012. "Closing the eyes on a gloomy future: Psychological causes and economic consequences." In *Pacific Development Economics Conference, University of California, Davis*.

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.J.L., and A.R. Quisumbing. 2015. "What dimensions of women's empowerment in agriculture matter for nutrition in Ghana?" *Food Policy* 52:54–63.

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.

Figure 1: VDC and ward sampling structure

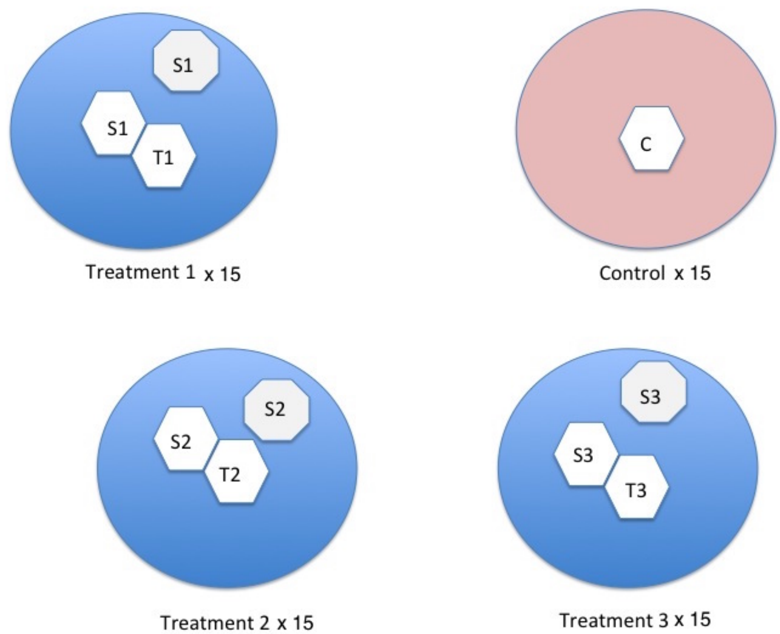
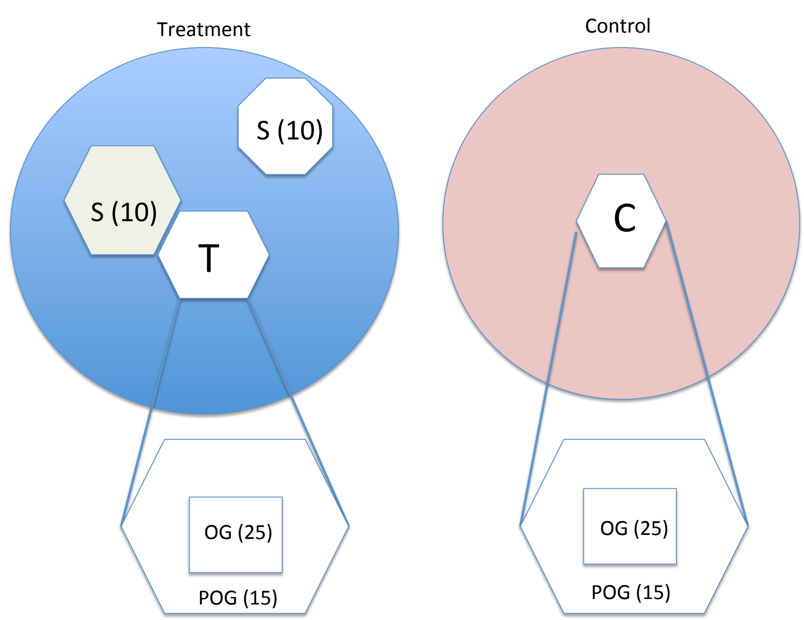


Figure 2: Household sampling structure



7.11 Appendix A: Livestock production indicators

To accommodate a study of the underlying mechanisms regarding differential treatment effects, we will consider the following sub-categories and sub-indicators related to livestock production. This outcome is not included in the primary welfare analysis.

1. Livestock expenditures
   (a) total annual livestock-related expenditures
   (b) expenditures for transportation for livestock feed
   (c) livestock breeding expenditures
   (d) livestock fodder and water expenditures
   (e) improved livestock shelter expenditures
2. Livestock herd dynamics
   (a) total livestock (TLU)
   (b) total goats
   (c) livestock purchases (TLU)
   (d) livestock gifts received (TLU)
   (e) livestock gifts given (TLU)
   (f) livestock births (TLU)
   (g) livestock deaths (TLU)
   (h) livestock sales (TLU)
3. Goat practices
   (a) Improved/non-improved pen
   (b) frequency of goat manure removal
   (c) goat manure used as fertilizer
   (d) use of free range grass
   (e) use of wild-grown fodder for goat feed
   (f) use of home-grown fodder for goat feed
   (g) use of grains/foodstuff available in the house for goat feed
   (h) use of mineral block for goat feed
   (i) use of medicine against leech
   (j) use of medicine against lice
   (k) use of other medicine
   (l) use of vaccination against PPR
   (m) use of other vaccination

4. Access to extension
    (a) awareness of access to trained animal health care specialist
    (b) any formal training in livestock management practices
5. Decision-making in livestock
    (a) some livestock ownership
    (b) some input into decisions regarding care and maintenance of livestock
    (c) some input into decisions regarding renting/selling livestock
    (d) some input into decisions regarding livestock income