September 1, 2017

Note on updated household welfare pre-analysis plan:

In the development of the targeting pre-analysis plan, we realized that own consumption of poultry, beef and eggs was not included in our measure of agricultural production – the survey question referenced in the pre-analysis plan only referenced output that was sold by household, rather than total production. Consistent with crop production, our aim was to capture both sales and own consumption. The targeting pre-analysis plan also contains a method for estimating baseline agricultural production that differs from what was originally included in the household welfare plan. We have amended the household welfare pre-analysis plan to incorporate these updates.

We have also revised our construction of hourly wages to focus on workers' cash salary, rather than after-tax wages and benefits.

Only de-identified data on agricultural and wage variables has been examined to date, following the data management plan filed to the AEA registry on July 24, 2017.

GE Effects of Cash Transfers: Pre-analysis plan for household welfare analysis*

Johannes Haushofer[†], Edward Miguel[‡], Paul Niehaus[§] and Michael Walker[¶]

September 1, 2017

Abstract

This document outlines outcomes and regression specifications for estimating the effects of unconditional cash transfers on the welfare of recipient households as part of the General Equilibrium Effects (GE) project. This project is a randomized evaluation of an unconditional cash transfer program by the NGO *GiveDirectly* (GD) in Kenya. This is a two-level randomized controlled trial where treatment status is randomized at the village level and treatment intensity is randomized at the sublocation level, the administrative unit above a village. This document is part of a series of five pre-analysis plans filed to the AEA trial registry as part of the GE project, and focuses on estimating treatment effects for households eligible to receive transfers from GD using data collected via household surveys. We specify regression equations, primary outcomes, and a catalog of outcomes that we will study, as well as corrections for multiple testing and data checks. We also discuss how the analyses specified here fit in with the other pre-analysis plans filed as part of this project.

Appendix A: Endline household survey instrument

Appendix B: Baseline household survey instrument

Appendix C: Household data management note

---

*AEA Trial Registry: AEARCTR-0000505, `https://www.socialscienceregistry.org/trials/505`
This updates a previously-filed version (from July 6, 2017) to include poultry and livestock own consumption in our measure of household income. This aligns the measure with our household income measure in the targeting pre-analysis plan (Haushofer et al. 2017) and is more consistent with our treatment of crop production. It also adjusts the construction of the hourly wage measure to focus on pre-tax cash salary.
We thank Justin Abraham, Christina Brown, Genevieve Deneoux, and Francis Wong for excellent research assistance, *GiveDirectly* for collaboration, the Busara Center for Behavioral Economics for survey development and piloting, and IPA-Kenya for data collection. This work has been funded by the Private Enterprise Development in Low-Income Countries (PEDL) initiative, the International Growth Centre, the Weiss Family Foundation, and an anonymous donor. Niehaus is a co-founder and president of *GiveDirectly*.

[†]Princeton University, NBER, and Busara Center for Behavioral Economics, haushofer@princeton.edu
[‡]UC Berkeley, emiguel@berkeley.edu
[§]UC San Diego, pniehaus@ucsd.edu
[¶]UC Berkeley, mwwalker@berkeley.edu

1 Introduction

1.1 Summary

This document outlines the analysis plan for endline household survey data collected as part of the General Equilibrium Effects (GE) project, a randomized controlled trial of an unconditional cash transfer program by the NGO *GiveDirectly* (GD). GD makes large unconditional cash transfers to poor households in Kenya. The magnitude of the transfers is large, around USD 1,000 (nominal) per household, about 75% of annual expenditure for recipient households. At the time of this study, GD targeted households living in homes with grass-thatched roofs, a basic means-test for poverty; we find 33% of households eligible in our study area. (GD currently uses a variety of targeting criteria that distributes transfers to a similar share of households). The intervention involves over USD 11 million in transfers and 653 villages in one Kenyan county. Treatment assignment is randomized at the village level, and within treatment villages, all households meeting GD's eligibilty requirement receive the unconditional cash transfer.[1] A second level of randomization provides variation in treatment intensity: sublocations, an administrative unit directly above the village level comprising of an average of ten villages, were randomly assigned to high or low saturation status. In high saturation sublocations, two-thirds of villages were assigned to treatment, while in low saturation sublocations, only one-third of villages were assigned to treatment.[2]

This analysis plan focuses on estimating direct treatment effects for eligible households (transfer recipients). Our experimental design allows us to estimate between-village spillover effects on eligible households, which could contaminate our estimates of the direct treatment effects. This design builds on Haushofer and Shapiro (2016), which allowed for within-village spillovers but assumed no spillovers across villages. This research question also relates to a broad literature on cash transfers, which generally finds positive effects for recipient households (Arnold, Conway, and Greenslade 2011 provide a review of the literature). As GD expects to commit around USD 50 million of cash transfers in 2017, estimating direct effects for transfer recipients is highly relevant.

This document is part of a series of five pre-analysis plans for the GE project. We already filed two pre-analysis plans: one on midline market price and enterprise phone survey data (Haushofer et al. 2016, filed May 19, 2016.), and one on local public finance (Walker 2017, filed February 12, 2017). We plan to file two additional pre-analysis plans in addition to this plan. One will study the targeting of cash transfers, and seeks to address whether targeting the most deprived households leads to the greatest average treatement effects (this will be referred to as the "targeting" PAP). The other, the general equilibrium (GE) PAP, focuses on prices, output and productivity, and looks into spillovers in more detail. The focus of this document is on direct treatment effects for recipient households, as these are highly relevant to policymakers, and the magnitude of treatment effects will provide context for understanding the results pre-specified in the other analysis plans.

1. This follows GD's typical operating procedure for lump sum transfers.
2. More details can be found in section 1.3 on the experimental design.

This pre-analysis plan is not meant to be exhaustive, nor to preclude additional analyses, and we anticipate carrying out additional analyses beyond those described here.

The remainder of this document describes the intervention and experimental design (Sections 1.2 and 1.3), the data collected (Section 2), the empirical strategy (Section 3), the primary outcomes (Section 4), and the full catalog of pre-specified outcomes (Section 5). Section 2.1 discusses data examined to date, and plans for examining data once this pre-analysis plan has been filed but before the rest of these pre-analysis plans are filed. The appendix contains household survey instruments and the associated data management note.

1.2 Intervention[3]

GD provides unconditional cash transfers to poor households in rural Kenya, targeting (for villages in our study) households living in homes with thatched roofs, a basic means-test for poverty. In treatment villages, GD enrolls all households in treatment villages meeting its thatched-roof eligibility criteria ("eligible" households); approximately one-third of all households are eligible. No households in control villages receive transfers. Eligible households enrolled in GD's program receive a series of 3 transfers totaling about USD 1,000[4] via the mobile money system M-Pesa.[5] This is a one-time program and no additional financial assistance is provided to these households after their final large transfer.

GD's enrollment process in treatment villages consists of the following 6 steps:

1. Village meeting (*baraza*): Before beginning work in a village, GD holds a meeting of all households in the village to inform villagers that GD will be working in their village, explain their program and GD as an organization. To prevent gaming, the eligibility criteria are not disclosed.

2. Census: GD staff conduct a household census of the village, collecting information on household names, contact information and housing materials. The information on housing materials are used to determine program eligibilty.

3. Registration: Households identified as eligible based on the household census are visited by the registration team. GD staff confirm the eligibility of the household, inform the household of their eligibility for the program and register the household for the program. This is the point at which households learn they will be receiving transfers, as well as the amount of the transfers, the transfer schedule, and the fact that the transfer is unconditional.[6] Households

3. Some of the text in this section and section 1.3 is reproduced from Walker (2017).

4. The total transfer amount is 87,000 Kenyan Shillings (KES). The average exchange rate from 9/1/14 to 4/30/16 was 97 KES/USD.

5. For more information on M-Pesa, see Mbiti and Weil (2015) and Jack and Suri (2011).

6. To emphasize the unconditional nature of the transfer, households are provided a brochure with many potential uses of the transfer.

are instructed to register for M-Pesa, a prerequisite for receiving the transfer. Households that do not have a mobile phone are given the option to purchase one from GD staff, the cost of which is deducted from the transfer amount.

4. Backcheck: All registered households are backchecked to confirm eligibility in advance of the transfers going out. This is an additional step to prevent gaming by households and field staff, as the census, registration and backcheck teams consist of separate staff members.

5. Transfers: The cash is transferred in a series of three payments via M-Pesa according to the following schedule: (i) the token transfer of KES 7,000 (about USD 70) ensures the system is working properly; (ii) two months afterwards, the first lump sum transfer of KES 40,000 is distributed; (iii) six months after this, the second and final lump sum transfer of KES 40,000 is sent. If households elected to receive a mobile phone from GD, the cost of this is taken out of the second lump sum transfer. Transfers are typically sent at one time per month to all households scheduled to receive transfers.

6. Follow-up: After transfers go out, GD staff follow up via phone with transfer recipients to ensure no problems have arisen. In addition, there is a GD help line that recipients can contact. If GD staff learn that household conflicts have arisen as a result of the transfers, transfers were sometimes delayed while these problems were worked out.

Existing evidence finds positive benefits of GD's program for recipient households: Haushofer and Shapiro (2016) conducted an impact evaluation in 2012 and found recipient households experienced a 61% increase in the value of assets, a 23% increase in expenditures, as well as improved food security and psychological well-being.

1.3 Experimental Design

This study is one component of a broader investigation into the general equilibrium effects of cash transfers (Haushofer et al. 2014). The GE project takes place in Siaya County, Kenya, a rural area in western Kenya bordering Lake Victoria. Siaya County is predominatly Luo, the second largest ethnic group in Kenya. GD selected both Siaya County and a region within Siaya County[7] based on its high poverty levels and identified target villages for expansion; in practice, these were all villages within the region that a) were not located in peri-urban areas and b) were not part of a previous GD campaign. This gives a final sample of 653 villages, spread across 84 administrative sublocations (the unit above a village), and 3 constituencies.[8]

7. This selection was based on the 2009 Population Census, which occurred prior to devolution and the creation of county governments. Based on 2009 administrative boundaries, the study area consists of 5 of the 7 divisions in Siaya District: Boro, Karemo, Ugunja, Ukwala and Uranga. The 2009 census lists enumeration areas, which we refer to as villages.

8. 5 villages were dropped after randomization: 4 villages, all of which contained the "Town" in the name, were dropped for being too urban for GD to work in. 1 of these was assinged to treatment, the remaining were assigned

We use a two-level randomization in order to generate variation that can be used to identify spillover effects. We randomly assigned sublocations (or in some cases, groups of sublocations) to high or low saturation status. Then, within high saturation groups, we assigned 2/3 of villages to treatment status, while within low saturation groups, we assigned 1/3 of villages to treatment status. As noted above, within treatment villages, all eligible households receive a cash transfer.

The randomization was conducted in two batches based on GD's expansion plans. The first batch included villages in Alego constituency, where GD had previously worked. In Alego, we sought to create saturation groups in which the number of villages in our study was a multiple of 3, if it was possible to combine contiguous sublocations; this also ensured at least 3 villages were in a saturation group. We created 23 saturation groups out of a total of 39 sublocations in Siaya, 11 of which matched directly to a single sublocation. Saturation groups in Siaya had on average 10 villages. We stratified assignment of high and low saturation by the level of exposure within the saturation group (the share of villages involved in a previous GD campaign), splitting the exposure level at the median. We then randomly assigned villages to 3 groups, and randomly assigned these groups to either a) always treatment, b) treatment in high saturation, control in low saturation and c) always control. We randomly generated an order for GD to work in by first randomly ordering the saturation groups and then villages within saturation groups. The second batch included villages in Ugunja and Ugenya constituencies. GD had not previously worked in any villages in these constituencies, so we did not stratify on any variables for these villages. Given the larger number of villages per sublocation, we also took the sublocation to be the saturation group. We assigned villages to one of three groups, pooled the "residual" villages that were not a multiple of 3, and randomly assigned 1/3 of these to the always treatment group, 1/3 to the treatment in high saturation sublocation group, and 1/3 to the always control group. GD worked first in Ugunja and then Ugenya. We generated a random order within these constituencies by first ordering locations (the administrative unit above the sublocation), then sublocations within the location, then villages within the location. Ordering based on location was used in an attempt to limit gaming by households.

Due to the large number of villages and households involved in the study, GD worked on a rolling basis across villages in the study area following the random order described above. The timing of transfers to eligible households within a village may vary for several reasons. GD generally began sending transfers to eligible households within a village once 50% of the eligible households (as identified via the census) completed the enrollment process. Villages that were above this threshold but in which GD was still working on completing the enrollment of other households would see a difference in the timing of transfers to households. If households delayed in signing up for M-Pesa, this would also introduce delays in their transfers and differences across villages. If households reported issues arising due to the transfers (such as marital problems or other conflicts), transfers may be delayed while these problems are worked out.

to control. The boundaries of one control village were unable to be determined by field staff despite repeated efforts. This was an enumeration area created for the 2009 census that did not correspond to existing village boundaries.

2 Data

The primary data source for this analysis are household surveys. In advance of the distribution of transfers to a treatment village, we conducted a baseline household census and household survey. The household census was designed to be comparable to GD's census, but to ensure there was no systematic bias between their censusing methods and ours, we conducted our own censuses in all villages (both treatment and control). The census served as a sampling frame for baseline household surveys. We determined household eligibility based on the census data and targeted 12 households per village for inclusion in the study, 8 eligible households and 4 ineligible households. We randomly ordered households by treatment status, and attempted to survey the first 8 eligible and first 4 ineligible households; we refer to these households as "initially-sampled" households. For couples, we randomly selected either the male or female to be the "target" respondent; if we could not reach the target, but the spouse/partner was available, we surveyed the spouse/partner.

If an initially-sampled household was not available to be surveyed on the day we visited the village for baseline surveys, we replaced this household with the next one on the list in order to ensure that we surveyed 12 households in each village; we refer to these households as "replacement" households. Lastly, we refer to households that were initially-sampled but unable to be surveyed as "missed baseline" households.

Endline surveys target all "initially sampled" and "replacement" households. For households that were baselined, we attempt to survey the same respondent that was surveyed at baseline. Endline surveys began at the end of May 2016 and concluded in June 2017. The median survey date is about 18 months after the baseline surveys and 10 months after the third cash transfer.

2.1 Analysis and data examined to date

A data management note was filed with the AEA trial registry detailing access to the endline household survey data in advance of the filing of this pre-analysis plan.[9] The note stated that members of the research team working on the pre-analysis plan would have limited access to the first weeks of survey data collection (a total of 334 surveys) and deidentified access to up to 10% of the target sample size to ensure survey forms worked properly, improve choice sets for questions, identifying any survey questions taking too much time and developing high frequency and consistency checks to ensure data quality. Field staff working in Kenya would have access to the data to compile, store, and update field tracking information, but would also not estimate treatment effects for any outcomes.

All members of the research team and field teams, as outlined in the note, have followed the

9. This note was filed on June 28, 2016, and updated on August 2, 2016 to include additional individuals as part of the plan. The text of the data management note and electronic signatures from each member of the research team is included in the appendix.

guidelines described in this notes, and no treatment effects have been estimated for the endline household survey datasets in advance of filing this pre-analysis plan. The only use of treatment status information has been to ensure balanced tracking rates across treatment and control villages.

One pre-analysis plan has already been filed describing the analyses to be conducted based on market survey and enterprise phone survey data (Haushofer et al. 2016, filed May 19, 2016.). A second pre-analysis plan on local public finance was filed (Walker 2017, filed February 12, 2017); this local public finance pre-analysis plan covered data collected from local leaders and school head teachers and includes endline household survey data, but access to endline household survey data has not yet been granted. As noted in Section 1.1, there are plans for two additional pre-analysis plans: the targeting pre-analysis plan and the GE pre-analysis plan.

After this, the household welfare pre-analysis plan, is filed, de-identified household survey data for eligible households will be provided to PIs and research assistants to begin the data cleaning process, so that we will not have any information about treatment or saturation status. This will simply allow us to begin variable construction and data cleaning while the other pre-analysis plans are finalized. Identified data will not include not be examined prior to the filing of the targeting and GE pre-analysis plans.

3 Empirical Strategy

3.1 Treatment effect of cash transfers

We use data from all eligible households – "initally-sampled" households (both those that were baselined and missed at baseline) and "replacement" households – as part of our main specifications. We base our classification of eligible households on GE household census data. This is analogous to an intention-to-treat (ITT) analysis.

For outcomes that were collected at baseline, our main specification is an ANCOVA that conditions on baseline values of the outcome variable; for households that we missed at baseline, we include an indicator that the household was missed at baseline, and include the mean value of the baseline variable in the regression equation.

$$y_{ihvs,t=1} = \beta_0 + \beta_1 T_{vs} + \beta_2 H_s + \delta_1 y_{ihvs,t=0} + \delta_2 M_{ihvs} + \varepsilon_{ihvs} \tag{1}$$

Here, h indexes the household, v indexes the village, s indexes the sublocation and t indicates whether the variable was measured at baseline or endline. We collect data on individual household members using a household roster. This allows us to analyze some outcomes on the individual, rather than the household level. For such outcomes, i indexes members in the household roster. T_{vs} is an indicator for households residing in a treated village. β_1 identifies the effect of receiving a transfer compared to eligible households in control villages. H_s is an indicator for living in a high-saturation sublocation, which we control for as it was part of the research design. β_2 is the effect on $y_{ihvs,t=1}$ of residing in a high-saturation sublocation; this is an average effect across treatment and control villages in high-saturation sublocations. Our focus in this pre-analysis plan is on β_1, and whether there are cross-village spillover effects (β_2) that could influence our interpretation of β_1. This is a very reduced-form approach to spillover effects; we conduct a more detailed examination and explore whether there is an additive effect from being in both a treatment and high-saturation sublocation ($T_{vs} \times H_s$) as part of the separate general equilibrium pre-analysis plan. Following McKenzie (2012), we condition on the baseline values of the outcome variable $y_{hv,t=0}$ to improve statistical power. When $y_{hv,t=0}$ is missing for an observation, we include an indicator term for missingness M_{ihvs} and replace $y_{hv,t=0}$ with its mean. Our primary specification clusters standard errors at the village level, our unit of randomization for treatment status. This provides the most precise estimate of the direct treatment effect, the coefficient on β_1. The tradeoff to added precision on β_1 is that the standard errors on β_2 may not be accurate. To address this possibility, we re-estimate equation (1) clustering standard errors at the saturation group level in order to test if $\beta_2 = 0$. If we find evidence of spillover effects, we will explore the robustness of β_1 and β_2 to alternative parameterizations of spillover effects, including those that may be discussed as part of the general equilibrium PAP.

For outcomes that were not collected as part of the baseline survey (outcomes in section 4 and 5 denoted by ‡), our primary specification is the following:

$$y_{ihvs,t=1} = \beta_0 + \beta_1 T_{vs} + \beta_2 H_s + \varepsilon_{ihvs}, \tag{2}$$

where all variables and standard errors are the same as in Equation 1.

For both equations 1 and 2, our first set of main hypothesis tests focus on the direct effects of cash transfers to recipient households. We test whether $\beta_1 = 0$, and adjust for multiple inference as outlined in Section 3.2 across this direct effect. Next, we test for whether there are cross-village spillovers ($\beta_2 = 0$) for eligible households using equation (1) with standard errors clustered at the saturation group level. We adjust for multiple inference for this test separately from the tests for direct effects. We again note that we will conduct a more detailed spatial investigation into spillover effects that will be pre-specified as part of a separate analysis plan.

We conduct several pre-specified robustness checks that will be included in an appendix. First, we estimate Equation 1 and restrict the sample to a balanced panel of "initally-sampled" eligible households. This avoids the need to use an indicator variable for missing values of the baseline value of an outcome variable, and, though the replacement should be unbiased, avoids the use of these households as well. Second, we estimate Equation 2 restricting the sample to only those eligible households that were initially targeted for surveys.

3.2 Multiple inference adjustments

There is a well-established literature that cash transfers have positive effects on recipient households. We believe the primary question is the dimensions on which cash transfers affect behavior and the magnitude of these effects, rather than the presence of any overall effect. Given that our survey instrument included several items related to a single behavior or dimension, we select ten primary outcomes (in some cases indices of multiple variables). The naive p-values on these outcomes are correct for readers with an a priori interest in one of these particular hypotheses. To control for multiple inference, we calculate sharpened q-values (i) across our 10 primary outcomes and (ii) within each outcome "family" (as defined in section 5) following Benjamini, Krieger, and Yekutieli (2006) to control the false discovery rate (FDR). The FDR controls for the proportion of false positives, which is relevant if one is interested in the proportion of the ten primary outcomes affected by treatment. Rather than specifying a single q, we report the minimum q-value at which each hypothesis is rejected, following Anderson (2008). We will report both standard p-values and minimum q-values in our analysis. We will apply the correction separately for each hypothesis test described in Section 3.1. For readers interested in the hypothesis of whether there is any overall effect on any of these primary outcomes, we estimate a system of seemingly unrelated regressions (SUR) and conduct a Wald test of joint significance for our ten primary outcomes. We note that norms around multiple testing are still evolving in economics, and through the above methods seek to follow current best practices.

3.3 Covariate adjustment

To improve precision, we will estimate two variants of Equations 1 and 2 that incorporate information on household and village characteristics as control variables. The first adjustment method includes theoretically important indicator variables, described in Section 5.13, as additive terms and interactions with the treatment indicator. The set of demeaned indicators partitions our sample so that our estimate remains unbiased for the average treatment effect (Lin 2013).

In a second adjustment method, we begin with the universe of baseline variables in the survey and perform model selection using the LASSO (Tibshirani 1996). Let $y_{k,ihvs}$ denote our primary outcomes measured at endline, where k indexes the outcome, i indexes the household member where available, h indexes the household, v indexes the village, and s indexes the sublocation. For each outcome, we estimate a LASSO regularized regression of the outcome on dichotomized baseline variables and their pairwise interactions. For each k^{th} regression, we perform a repeated 10-fold cross-validation and choose the regularization parameter according to the "one standard error rule"[10] (Krstajic et al. 2014). Denote the sparse set of regressors with non-zero coefficients by $\mathbf{U}_{k,ihvs}$. We obtain covariate-adjusted treatment effect estimates by estimating Equations 1 and

10. We select the most parsimonious model whose error is no more than one standard error above the model with minimum error.

2 including the demeaned vector $\dot{\mathbf{U}}_{k,ihvs}$ as an additive term and interaction with the treatment indicator.

We can compare our primary specification in Equation 1 and both methods of covariate adjustment and examine each specification's capacity to provide precise estimates of the treatment effect. These specifications with covariate adjustment will be included in the appendix. If either specification with covariate adjustment is considerably more precise, we will also include it in the main analysis with our primary specification without covariate adjustment.

3.4 Exact tests of the treatment effect

In addition to the large-sample approach outlined in Section 3.1, we perform Monte Carlo approximations of exact tests of the treatment effect (Fisher 1935). Randomization inference allows us to test the Fisherian sharp null hypothesis that $y_{ihvs}^{(t)} = y_{ihvs}^{(c)}$ for every unit $ihvs$.[11] To remain analogous with our main specification, we use the conventional Wald statistics from Equation 1 as our test statistics.

We calculate exact p-values for the treatment and spillover effects under the null hypothesis using a Fisher permutation test. Specifically, we take 10,000 permutations of the treatment indicator T_{vs} and calculate the Wald statistics for each m^{th} permutation. We hold fixed the treatment-control balance by randomly re-allocating the realized village treatment assignments rather than replicating the underlying distribution. We hold fixed the values of household eligibility E_{hvs} and sublocation saturation status H_s. The exact p-value is

$$\frac{1}{10,000}\sum_{m=1}^{10,000}\mathbf{1}\left[\hat{\boldsymbol{\beta}}_m'\hat{V}(\hat{\boldsymbol{\beta}}_m)^{-1}\hat{\boldsymbol{\beta}}_m > \hat{\boldsymbol{\beta}}_{obs.}'\hat{V}(\hat{\boldsymbol{\beta}}_{obs.})^{-1}\hat{\boldsymbol{\beta}}_{obs.}\right] \tag{3}$$

We will adjust for multiple inference in a fashion analogous to adjustments described in Section 3.2 and calculate minimum q-values to control the FDR. We will report results from our main specification and from randomization inference, but consider the results from the main specification to be primary.

3.5 Heterogeneous treatment effects

We will test whether the impact of the cash transfers varies with household characteristics measured at baseline. This allows us to understand which sub-groups contribute disproportionately to average treatment effects. Pre-specified dimensions of heterogeneity are described in Section 5.13. To investigate heterogeneity by baseline characteristic X_{ihvs}, we estimate a model for each dimension of heterogeneity that interacts the treatment indicator with X_{ihvs}.

11. This hypothesis is more restrictive than a null hypothesis of no average treatment effect.

$$y_{ihvs,t=1} = \beta_0 + \beta_1 T_{vs} + \beta_2 H_s + \gamma_0 X_{ihvs} + \gamma_1 (T_{vs} \times X_{ihvs}) + \delta y_{ihvs,t=0} + \varepsilon_{ihvs} \tag{4}$$

γ_1 represents differential treatment effects across values of X_{ihvs}. We control for the FDR across dimensions of heterogeneity for each outcome and follow the approach outlined for equation (1) for standard errors.

3.6 Time-dependent treatment effects

Our household survey data contains temporal variation in the delivery of the cash transfers as part of the experimental design. As part of our randomization process, we created a randomized order of both treatment and control villages that determined the order villages begin treatment.[12] In Alego subcounty, we randomly ordered sublocations then randomly ordered villages within these sublocations. In Ugunja and Ukwala subcounties, we first randomly ordered locations (the administrative unit above a sublocation) within each subcounty, then randomly ordered sublocations within locations and villages within sublocations. Clustering by locations in Ugunja and Ukwala was added to help prevent gaming of GD's eligibility criteria; since GD had previously worked in Alego prior to the start of the GE project, this was not implemented for Alego.

The order in which households were surveyed at endline was also randomly determined at the village-level. In theory, survey date is orthogonal to treatment assignment though it is possible that time to endline may correlate with treatment in practice. To assess whether actual endline survey date differs systematically by treatment assignment, we estimate the following equations.

$$\text{Months}_{vs} = \beta_0 + \beta_1 T_{vs} + \beta_2 H_s + \varepsilon_{vs} \tag{5}$$

$$\text{Days}_{hvs} = \beta_0 + \beta_1 T_{vs} + \beta_2 H_s + \varepsilon_{hvs} \tag{6}$$

Equation 5 is estimated at the village level with Months_{vs} as the number of months since the first village is surveyed at endline. Equation 6 is estimated at the household level with Days_{hvs} as the number of days since the first household in the village is surveyed. β_1 in Equation 5 and Equation 6 is the differential delay in survey rollout at the village and household levels, respectively.

We leverage the experimentally-assigned treatment ordering to analyze how treatment effects may depend on time elapsed since the treatment delivery. We define the "experimental last transfer month" for a village as 8 months after the village's "experimental start date" to reflect when villages

12. The first transfer households receive is for USD 72 and is referred to by GD as the "token" transfer. The first transfer is followed by two lump sum transfers of USD 412. The second transfer was sent 2 months after the first transfer and the third transfer sent 8 months after the first transfer.

(if assigned to treatment) could expect the first payments of the third (and final) transfer.[13] We define the month of endline for a village as the first month in which households within that village begin endline surveys. We begin by creating separate indicators for the difference between the month of endline and the experimental last transfer month. Let $R_{vs,q}$ be an indicator which takes a value of 1 for villages in the q^{th} tercile in the distribution of months since the experimental last transfer month. We then estimate the following model with $R_{vs,q=1}$ as the omitted group, clustering standard errors at the village level.

$$\begin{aligned} y_{ihvs,t=1} =& \beta_0 + \beta_1 T_{vs} + \beta_2 H_s + \zeta_0 R_{vs,q=2} + \zeta_1 (T_{vs} \times R_{vs,q=2}) + \eta_0 R_{vs,q=3} + \eta_1 (T_{vs} \times R_{vs,q=3}) \\ &+ \delta y_{ihvs,t=0} + \varepsilon_{ihvs} \end{aligned} \tag{7}$$

We will test $\zeta_1 = 0$, $\eta_1 = 0$, and $\zeta_1 = \eta_1$ to determine whether treatment effects on recipient households differ by transfer timing.

3.7 Analysis of potential attrition bias

To assess whether attrition of households between baseline and endline surveys confounds our results, we conduct the following analyses. Let r_{hvs} be an indicator for whether household h in village v in sublocation s is observed at baseline but not at endline. First, Equation 8 estimates whether the magnitude of attrition varies with treatment status, and we follow equation (1) for our standard errors:

$$r_{hvs} = \beta_0 + \beta_1 T_{vs} + \beta_2 H_s + \varepsilon_{hvs}. \tag{8}$$

Second, Equation 9 assesses whether observation status varies with a vector of baseline characteristics $\mathbf{X}_{hvs,t=0}$ including variables from Sections 4 and 5.13.

$$r_{hvs} = \omega \mathbf{X}_{hvs,t=0} + \varepsilon_{hvs} \tag{9}$$

Third, Equation 10 tests whether there are differences in baseline characteristics across treatment arms for respondents observed at endline.

$$(X_{hvs,t=0} | r_{hvs} = 1) = \beta_0 + \beta_1 T_{vs} + \beta_2 H_s + \varepsilon_{hvs} \tag{10}$$

Equations 8 - 10 will be estimated only for households surveyed at baseline and with standard

13. The construction of the experimental start date is explained in detail in Haushofer et al. (2016). In summary, we assign villages to an experimental start date based on the randomized village ordering and GD's pace working across subcounties.

errors clustered at the village level when looking at β_1 and the saturation gropu level when looking at β_2. If we find worrying levels of differential attrition, we will adjust for potential bias by bounding our parameter of interest (Lee 2009) and by using a weighted least squares estimator with the inverse probability of selection as weights.

3.8 Balance tests

In the appendix, we will include a balance table testing for baseline balance for eligible households across treatment villages and saturation status for all of our primary outcomes for which we have baseline data (those marked with a ‡ in Section 4) and our dimensions of heterogeneity (listed in 5.13), denoted $\mathbf{X}_{hvs,t=0}$ below:

$$\mathbf{X}_{hvs,t=0} = \phi_0 + \phi_1 T_{vs} + \phi_2 H_s + \varepsilon_{ihvs} \tag{11}$$

We are interested in whether ϕ_1 and ϕ_2 are different than zero, implying that there are baseline differences in these outcomes across treatment and control households. Standard errors follow those in equation (1).

4 Primary outcomes of interest

The variables detailed in this section are our primary outcomes of interest will be included in multiple comparison adjustments. In future work examining the long-term effects of cash transfers, we will pool data with the three-year follow-up of Haushofer and Shapiro (2016). Outcomes marked by * do not appear as outcomes in Haushofer and Shapiro (2016). Outcomes marked by † will be analyzed at the individual member level. Outcomes marked by ‡ were not measured at baseline and will be analyzed using Equation 1 omitting $y_{hv,t=0}$.[14] As is common with many income and consumption measures, we take the standard approach of winsorizing the top 1% of our monetary and hours worked variables (outcomes 1, 2, 3, 4 and 10 below). In the appendix we will present unadjusted estimates, as well as estimates trimming the top 1%. If there are large differences between these approaches we will investigate the reasons why and the individuals driving these results.

14. The context should determine when numbers refer to outcomes in the pre-analysis plan versus question numbers (variables) from the household survey. We utilized several versions of the household survey; while all versions were substantively the same, we made minor changes to reflect realized survey times and other issues that may have been encountered over the course of fieldwork. All question numbers refer to those in the last version of the endline survey (denoted v20) and included as an appendix. We will make all other survey versions available as well, and will refer to the analogous questions to those in version 20 in the case of any numbering changes. We construct baseline variables based on these same questions or their analogues, though the numbers may differ. If an index includes both variables that were and were not collected at baseline, we construct the baseline value of the index based on variables that were collected. Cases where baseline variables are constructed in a significantly different manner are noted.

1. **Total value of non-land assets**: Summary measure of asset values (section 5.1); sum of value of asset variables 6.13.a-z, 6.13.aa-hh, and value of loans given (10.8.b) net of total amount of loans taken (sum of variables 10.3.d, 10.4.a, 10.5.a, 10.6.a, and 10.7.b).

2. **Total consumption expenditure in last 12 months**: Summary measure of consumption (section 5.2); sum of total food consumption in last 7 days (12.Q1 for items 12.1-18), frequent purchases in last month (12.19-29) and infrequent purchases in last 12 months (4.4.4, 12.30-38,39b), converted to yearly values.[15]

3. **Total household income in the last 12 months***: Summary index of income and profits (section 5.3); sum of total profits from agriculture and livestock in the last 12 months (outcome 5.3.1) plus total profits from non-ag. business in the last 12 months (outcome 5.3.2) and the total after-tax value of wages, salaries and in-kind transfers earned in the last 12 months (outcome 5.3.3).

4. **Total household business revenue in the last 12 months***: Summary measure of business revenue (section 5.4); sum of total revenue from agriculture and livestock in the last 12 months (outcome 5.4.1) plus sum of total revenue from non-ag business in the last 12 months (outcome 5.4.2).

5. **Subjective well-being index**: Summary measure of subjective well-being (section 5.5); weighted, standardized index of depression, happiness, life satisfaction, and stress, appropriately signed so that positive values indicate greater subjective well-being.

6. **Health status index**: Summary measure of health (section 5.6); weighted, standardized average of self-reported health, index of symptoms, and experienced a major health problem, appropriately signed so that positive values indicate better health outcomes.

7. **Education index**: Summary measure of education (section 5.7): weighted, standardized index of total education expenditure and proportion of school-aged children in school.

8. **Female empowerment index‡**: Summary measure for female empowerment (section 5.8); weighted, standardized index of attitudes index and violence index.

9. **Food security index**: Summary measure of food security (section 5.9); weighted, standardized index of food security outcomes.

10. **Hours worked in the last 7 days***: Summary measure of labor supply (section 5.10); sum of respondent hours worked in agriculture, self-employment and employment. We note that total hours worked is not a welfare measure, and include it as a primary measure given the strong interest in the labor supply responses to cash transfers.

15. Based on aggregate consumption questions that match Haushofer and Shapiro (2016).

4.1 Construction of summary indices

We will follow the procedure proposed by Anderson (2008) to construct indices of subjective well-being, food security, health, education, and female empowerment. First, for each outcome variable y_{jk}, where j indexes the outcome group and k indexes variables within outcome groups, we recode the variable such that high values correspond to positive outcomes. We then compute the covariance matrix $\hat{\mathbf{\Sigma}}_{\mathbf{j}}$ for outcomes in outome group j, which consists of elements:

$$\hat{\Sigma}_{jmn} = \sum_{i=1}^{N_{jmn}} \frac{y_{ijm} - \bar{y}_{jm}}{\sigma^y_{jm}} \frac{y_{ijn} - \bar{y}_{jn}}{\sigma^y_{jn}} \tag{12}$$

Here, N_{jmn} is the number of non-missing observations for outcomes m and n in outcome group j, $\bar{y}_{jm}$ and $\bar{y}_{jn}$ are the means for outcomes m and n, respectively, in outcome group j, and σ^y_{jm} and σ^y_{jn} are the standard deviations in the pure control group for the same outcomes. Next, we invert the covariance matrix, and define weight w_{jk} for each outcome k in outcome group j by summing the entries in the row of the inverted covariance matrix corresponding to that outcome:

$$\hat{\mathbf{\Sigma}}_{\mathbf{j}}^{\mathbf{-1}} = \begin{bmatrix} c_{j11} & c_{j12} & \cdots & c_{j1K} \\ c_{j21} & c_{j22} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots \\ c_{jK1} & \vdots & \ddots & c_{jKK} \end{bmatrix} \tag{13}$$

$$w_{jk} = \sum_{l=1}^{K_j} c_{jkl} \tag{14}$$

Here, K_j is the total number of outcome variables in outcome group j. Finally, we transform each outcome variable by subtracting its mean and dividing by the control group standard deviation, and then weighting it with the weights obtained as described above. We denote the result $\hat{y}_{ij}$ because this transformation yields a generalized least squares estimator.

$$\hat{y}_{ij} = \left(\sum_{k \in \mathbb{K}_{ij}} w_{jk} \right)^{-1} \sum_{k \in \mathbb{K}_{ij}} w_{jk} \frac{y_{ijk} - \bar{y}_{jk}}{\sigma^y_{jk}} \tag{15}$$

4.2 Constructing real values

Our primary specifications use nominal values for monetary outcomes. As part of the study, we collect market prices from 61 weekly markets in our study area, and use this to construct an index of market prices (see Haushofer et al. 2016 for full details). We will check for price effects, and if we find large price effects, we will also report estimates using real values in the main analysis. If we do not find large price effects, we will report real values for monetary outcomes in an appendix. The discussion of price effects will be contained in papers that focus on the results of the midline and GE pre-analysis plans.

We construct real values for monetary outcomes in the following manner. We generate our market price index at a monthly frequency as outlined in Haushofer et al. (2016), and take the mean over the main period of endline data collection (June 2016 to January 2017). For households living in (or near) our study area at endline, we then assign households the market price index of their nearest market, based on the as-the-crow-flies distance between households' location and markets. For households that are missing GPS coordinates and still residing in a GE study village, we use the geometric mean of households within the village from baseline household census data. For households that have migrated outside of the study area to another rural area, we use the mean market price across all markets in our study area as a rural price index. For households that have migrated to an urban area, we inflate our rural price index by the mean urban/rural price difference in market price surveys collected as part of the Kenya Life Panel Survey project, a longitudinal study of nearly 10,000 Kenyan youth involved in two previous randomized controlled trials (Miguel and Kremer 2004; Kremer, Miguel, and Thornton 2009), during our endline survey period.

5 Catalog of variables

We will analyze the treatment effect of cash transfers on a comprehensive list of outcomes, classified below by category, to better understand mechanisms underlying potential effects. We apply multiple inference corrections via FDR as outlined in section 3.2 for each "family" of outcomes, denoted by a separate section. Outcomes marked by * do not appear as outcomes in Haushofer and Shapiro (2016). Outcomes marked by † will be analyzed at the individual member level. Outcomes marked by ‡ were not measured at baseline and will be analyzed using Equation 2.

5.1 Assets

Summary measure – Total value of non-land assets: Sum of value of asset variables 6.13.a-z, 6.13.aa-hh, and value of loans given (10.8.b) net of total amount of loans taken (sum of variables 10.3.d, 10.4.a, 10.5.a, 10.6.a, and 10.7.b).

1. **Total value of livestock**: Sum of value of cattle, goats, sheep, chicken, other birds, and pigs (variables 6.13.aa-ff).
2. **Total value of agricultural tools**: Sum of value of farming tools, hand carts, wheelbarrows and ox plows (variables 6.13.w-z).

3. **Total value of furniture**: Sum of value of beds, mattresses, bednets, tables, sofa pieces, chairs, cupboards/dressers, and clock or watches (variables 6.13.h-o).

4. **Total value of radio / cassette player / CD player or TVs**: Sum of variables 6.13.e and 6.13.q.[16]

5. **House has non-mud floor**: Indicator variable from variable 6.0.a.

6. **House has non-thatched roof**: Indicator variable from variable 6.0.b for roof made of materials other than grass or leaves.

7. **House has non-mud walls**: Indicator variable from variable 6.0.c.

8. **House has electricity**: Indicator for house having electricity from any source (variable 6.1).

9. **House primarily uses an improved toilet**: Indicator variable for toilet or portable toiled from variable 6.2.[17]

10. **Cost of materials and labor to build house***: Variable 6.5.a for households that own their own home (6.5).

11. **Total value of land owned by household***: Total acres of land owned by the household (variable 6.6) multiplied by price per acre of land in the village (6.6.a).

12. **Total amount of loans taken in the last 12 mo.***: Sum of borrowing from merry-go-round/ROSCA, commercial banks/lenders, moneylenders, M-Shwari.[18] and anyone else outside household (variables 10.3.d, 10.4.a, 10.5.a, 10.6.a, and 10.7.b)

13. **Total amount of loans given in the last 12 mo.***: Total amount lent to anyone outside household respondent expects to get back (variable 10.8.b).

As with our primary outcomes, we winsorize the top 1% of the monetary outcomes (the summary measure and outcomes 1-4 and 10-13) as our main specification. We also explore using the unadjusted outcomes and trimming the top 1% of outcomes in an appendix. If we find substantial differences between these approaches, we will investigate the reasons why.

5.2 Consumption

Under the permanent income hypothesis, we would expect to see small changes in consumption expenditure as households consume the annuity value of the transfer. However, in both the US and

16. Haushofer and Shapiro (2016) only includes radios and TVs.
17. Haushofer and Shapiro (2016) use an indicator for ownership of an improved toilet/latrine. Our survey question asked about most frequent use rather than ownership, so we have adjusted this outcome.
18. M-Shwari is a mobile lending service tied to M-Pesa and Safaricom.

in developing countries, there is evidence of strong consumption expenditure responses to transfers (e.g. Parker et al. 2013; Haushofer and Shapiro 2016; Jappelli and Pistaferri 2010).

Summary measure – Total consumption expenditure in last 12 months:‡ Sum of total food consumption in last 7 days (items 12.1-18), frequent purchases in last month (12.19-29) and infrequent purchases in last 12 months (4.4.4, 12.30-38,39b), converted to yearly values and all based on aggregate consumption questions.

1. **Total food consumption in the last 12 months***‡: For 23 food items including staples, vegetables, meat, fruits, and other consumption (items 12.C1-C23), expenditure in a typical week (12.Q8) for months purchased times the number of months purchased (12.Q7) plus total consumption value in a typical week for months produced (12.Q6) times months produced (12.Q5) plus gift consumption (12.Q9).
2. **Marginal utilities of consumption expenditure ("neediness")*‡**: Index of 23 food items including staples, vegetables, meat, fruits, and other consumption (12.C1-C23) as calculated in Ligon (2016).
3. **Total expenditure on temptation goods in the last mo.‡**: Sum of alcoholic drinks and tobacco products in last 7 days (12.Q1 for items 12.14 and 12.15) times 4 plus lottery tickets/gambling in the last month (12.22).
4. **Total housing expenditure in the last 12 mo.‡**: Sum of house rent/mortgage, home repair/maintenance, and improving/expending home (variables 12.30-32).
5. **Total education expenditure in the last 12 mo.**: Variable 4.4.4.[19]
6. **Total medical expenditure in the last 12 mo.‡**: Variable 12.36.
7. **Total social expenditure in the last 12 mo.‡**: Sum of recreation/entertainment expenses in the last month (12.24) times 12 plus religious expenses, charitable donations, weddings and funerals, and dowry/bride price in the last 12 months (variables 12.33-35 and 12.38).
8. **Total expenditure on durables in the last 12 mo.‡**: Variable 12.37.
9. **Total flow value of durables in the last 12 months‡**: see Section 5.2.1.
10. **Total consumption in the past 12 months, including flow value of durables‡**: Total consumption summary measure plus flow value of durables (see Section 5.2.1).

As with our primary outcomes, we winsorize the top 1% of the monetary outcomes (the summary measure and outcomes 1, 3-10) as our main specification. We also explore using the unadjusted outcomes and trimming the top 1% of outcomes in an appendix. If we find substantial differences between these approaches, we will investigate the reasons why.

19. We construct baseline values of educational expenditure based on total school fees, supplies, and other school contributions in the 3 most recent school terms across all children.

5.2.1 Estimating the flow value of consumer durables

In addition to using the purchase price of durable goods, we estimate the value of services flowing from durables as a measure of household welfare. This is important as household investments in durables could actually decrease household expenditure, for instance if it results in lower maintenance costs. We take two approaches to estimating the flow value of household durables, first measuring rental equivalence and then measuring the user cost. We note in advance that we only have this information for a smaller set of households, particularly once we look only at eligible households within treatment and control villages. Because of this, these analyses for consumption outcomes 9 and 10 may be more suggestive, and if we find that these measures are too imprecise to draw meaningful inferences, we may exclude these outcomes from our main analysis, though we will still report them in an appendix.

If competitive rental markets exist for a durable good and the economy is in equilibrium, the market rental value measures the flow value from that durable. To measure rental equivalence, we surveyed both market vendors (as part of a subset of rounds of our market surveys) and a subsample of households to ask about the price at which they would rent durables. For households, we only ask the rental value if the household owns the item. Our preferred measure uses the data collected from households, though we again note that we only have this for a small randomly-selected subset of households. We take the mean annualized rental price of a single unit the good separately for treatment and control villages, as well as eligible and ineligible households. We then multiply the applicable mean rental value by the number of units of a good that each household owns in order to have a measure of the flow value for all households in our sample.[20]

We will compare household values of rental equivalence with those collected from market surveys. With market surveys, we are unable to test for differences across treatment and control villages or eligible and ineligible households, which may mask potential quality differences between groups. We can test for differences in rental prices based on the density of recipient households within a certain radii.

Because rental markets are not common in many of these goods, we also follow the user cost approach outlined in Deaton and Zaidi (2002) to serve as a sensitivity and consistency check. This approach takes the opportunity cost of owning the good for one period as its flow value. The flow value V of a durable good over period t is given by

20. There are several items (beds and mattresses in particular) for which it is not customarily appropriate to ask about renting. We do not observe rental prices for these and thus omit them from our measure of consumption value when using rental equivalence. We are able to calculate user costs for these items.

$$V_t = p_t(i_t - \pi_t + \delta_t)$$

$$\delta_t = 1 - \pi_t - \left(\frac{p_t}{p_{t-T}}\right)^{\frac{1}{T}}$$

$i_t - \pi_t$ is the Fisherian real interest rate in period t and δ_t is the geometric depreciation rate in period t calculated using the vintage T, original purchase price p_{t-T}, and current market price p_t. We set our reference period t over 12 months to remain analogous with our other consumption measures. The user cost calculated in this way can be interpreted as the sum of the good's net return and possible capital gains over the 12 month period.

We obtain data on the real rate of interest taken as an average over several years and use that rate for all durable goods. We observe for all households the stock of durables and current market price p_t. We will use median values from the subset of households surveyed about user costs (the same households asked about rental costs) to impute T and p_{t-T} of each good for all households. We will test for differences between the flow value of a unit of a good calculated via the rental equivalence measure and the user cost measure.

5.3 Income and profits

Summary measure – Total household income in the last 12 months*: Sum of total profits from agriculture and livestock in the last 12 months (outcome 1 below) plus total profits from non-ag. business in the last 12 months (outcome 2 below) and the total after-tax value of wages, salaries and in-kind transfers earned in the last 12 months (outcome 3 below).

1. **Total profits from agriculture and livestock in the last 12 mo.***: Sum of crop output (7.16) valued at market prices if not reported in monetary units,[21] plus pastoral output sold (7.6aa) plus own production of poultry, eggs and beef in the last 12 months, calculated as number of months consuming own production (12.Q5) times monetary value of typical weekly consumption when consuming own production (12.Q6) times 4 for chicken/duck/poultry, beef and eggs)), net of salaries paid to workers outside the household (7.12), agricultural inputs (7.13.a-f), and the rental cost of land for agricultural purposes (acres rented for agriculture (6.8.b) times months rented (6.8.c) times monthly rent (6.8.d)).[22,23]

21. Commodity prices obtained from market surveys. We use the price of the crop at the nearest market over the course of the endline survey. For any crops that are not included in our market survey, we use the median unit price for households within the same sublocation.

22. We do not subtract off unpaid labor. It is not clear that respondents are doing this when we ask about self-employment profits, so we do not subtract from ag. profits for consistency.

23. Due to time constraints in our baseline survey, we did not collect information on crop-by-crop production or own consumption of poultry, eggs and beef. To construct baseline measures, we generate predicted a) crop profits, b) poultry profits and c) livestock profit as a flexible function of quartic polynomials of: i) amount of crop/poultry/livestock sold; ii) input costs (sum of spending on tools, fertilizer, irrigation, animal medicine, improved/hybrid seeds, agricul-

2. **Total profits from non-ag. business in the last 12 mo.***: Sum of self-reported profits (8.11b) for all businesses owned.[24]

3. **Total after-tax value of wages, salaries and in-kind transfers earned last 12 mo.***: Sum of cash salary last month (9.10) plus total value of benefits last month (9.12) net of income tax paid last month (9.11), annualized to a yearly measure by multiplying by the number of months worked in last 12 months (calculated as month of survey minus employment start date (9.3) if employment working patterns (9.7) are full-time or part-time, and based on 9.7a if employment working patterns are seasonal.) for all workers in the household. We set this to zero if no household members are working for wages.[25]

4. **Hourly wage rate for those employed/working for wages***†: Cash salary (variable 9.10) last month times (7/30) to get a measure of weekly earnings, divided by total hours worked in the last week (9.8), for each individual in the household working for wages. We investigate this at an individual level and we restrict this analysis to individuals working for wages. When looking at labor supply and time use (section 5.10), we look at selection into working for wages as an outcome.

As with our primary outcomes, we winsorize the top 1% of the monetary outcomes (the summary measure and outcomes 1-4) as our main specification. We also explore using the unadjusted outcomes and trimming the top 1% of outcomes in an appendix. If we find substantial differences between these approaches, we will investigate the reasons why.

5.4 Business revenue

Summary measure – Total household business revenue in the last 12 months[26]: Sum of total revenue from agriculture and livestock in the last 12 months (outcome 1 below) plus total self-employment earnings in the last 12 months for all businesses owned (outcome 2 below).

1. **Total revenue from agriculture and livestock in the last 12 mo.**: Sum of crop output (7.16) valued at market prices if not reported in monetary units, plus pastoral output (7.6aa)

tural insurance); iii) (crops only) land used in agriculture; iv) (crops only) amount paid on land rental for agriculture; v) number of workers working in each agricultural activity in the last 12 months. The listed variables were all collected at baseline. We use out-of-sample data from households surveyed as part of the Kenya Life Panel Survey, Round 3 (KLPS-3) located in Busia and Siaya counties to estimate the relationship between these variables, as the KLPS-3 survey collected all of these variables. We then generate predicted agricultural and livestock profits for households at baseline in our sample. This procedure matches that described in the targeting analysis plan. We will also look at a measure of total agricultural and pastoral output sales, as included in the original version of this pre-analysis plan.

24. We will also analyze this variable as a monthly measure using 8.11a. Our primary measure will use self-reported profits. We will also construct a measure where we replace missing values of self-reported profits with self-employment revenue (8.7b) net of total costs (sum of 8.15, 8.16.a-h, and 8.17.a-g times 12 plus annualized business license costs (8.8a times 12 divided by 8.8b, the number of months license valid) for all businesses owned. If we find high levels of missing values for 8.11b we may replace our primary measure with this constructed measure.

25. We will also analyze this variable as a monthly measure.

26. Haushofer and Shapiro (2016) looks at business revenue outcomes at a monthly frequency

for livestock/pastoral activities.[27,28]

2. **Total revenue from non-ag. business in the last 12 mo.**: Sum of variables 8.8b for all businesses owned.[29]

3. **Non-ag. business owned by household**: Indicator for variable 8.1.

4. **Total costs in the last 12 mo.***: Sum of costs in agriculture and livestock, and non-ag. business (see outcomes 5 and 6 below).

5. **Total costs in agriculture and livestock in the last 12 mo.***: Sum of wages paid to agricultural/pastoral workers outside the household (7.12) and agricultural/pastoral inputs (7.13.a-f) for all agricultural/pastoral activities plus the rental cost of land from acres rented for agricultural purposes (6.8.b) times months rented (6.8.c) times monthly rent (6.8.d, converted to monthly rate if reported in other units).

6. **Total costs in non-ag. business in the last 12 mo.***: Sum of total wage bill last month (8.6c) plus monthly rent (8.15a) plus spending on operating costs (8.16.a-h) for all businesses owned plus government taxes, fees and bribes (8.17a-f), multiplied by 12, plus annual cost of county license (8.8a times 12 divided by number of months valid (8.8b)) if business licensed with the county government.

As with our primary outcomes, we winsorize the top 1% of the monetary outcomes (the summary measure and outcomes 1-6) as our main specification. We also explore using the unadjusted outcomes and trimming the top 1% of outcomes in an appendix. If we find substantial differences between these approaches, we will investigate the reasons why.

5.5 Subjective well-being

Summary measure – Subjective well-being index: Weighted, standardized average of depression, happiness, life satisfaction, and perceived stress, appropriately signed so that positive values represent better subjective well-being.

1. **Depression**: Scale score calculated from variables 14.1-10.

2. **Happiness**: Variable 14.23, reverse-coded.

3. **Life satisfaction**: Variable 14.24.

27. Commodity prices obtained from market surveys. We use the price of the crop at the nearest market over the course of the endline survey. For any crops that are not included in our market survey, we use the median unit price for households within the same sublocation.

28. Due to time constraints at baseline, we did not collect crop-specific measures of output, and instead focused on total agricultural and pastoral output sales, which will be used as the measure of ag. and livestock revenue.

29. We will also analyze this variable as a monthly measure.

4. **Perceived stress‡**: Scale score calculated from variables 15.7.1-4.

5. **Aspirations**[30]: Scale score calculated from variables 4.4.5a-d, 4.4.6a-d, 4.4.7-10, and variables in sections 15.1-15.3.

6. **Self-efficacy**: Scale score calculated from variables 15.4.1-10.

7. **Internal locus of control**: Scale score calculated from variables 15.5.1-5.

8. **Hope‡**: Scale score calculated from variables 15.6.1-8.

5.6 Health

Summary measure – Health status index[31]: Weighted, standardized average of self-reported health, index of symptoms, and experienced a major health problem, appropriately signed so that positive values indicate better health outcomes.

1. **Self-reported level of general health***: If respondent reports health is "very good" in 11.8, coded as 5. Otherwise based on answer to 11.9, where "good" is coded as 4, "fair" is coded as 3, "poor" is coded as 2, and "very poor" is coded as 1.

2. **Index of recent symptoms*‡**: Weighted, standardized average of variables 11.13 for health conditions.

3. **Days of work, school missed due to poor health in the last 4 weeks***: Variable 11.7.

4. **Experiences major health problem that affected work or life*‡**: Variable 11.18.

5. **Major health problem resolved*‡**: Indicator variable using 11.21 for whether any listed health problems have been resolved, conditional on having a major health problem (11.18).

6. **No. of visits to hospital or clinic in the last 4 weeks*‡**: Variable 11.14.

7. **Total expenditure on medical treatments and medicine in the last 4 weeks*‡**: Sum of variables 11.15.a-c.

We note that for health outcomes, greater utilization of health services in terms of the number of visits or total expenditure could occur as households are now able to afford additional health care, or due to a decline in health status. Similarly, cash transfers could allow respondents in poor health to miss work when they otherwise would have felt the need to work. Care should thus be taken in interpreting these measures, and interpretation will depend in part on our findings for other outcomes. We include these outcomes as we believe these are still important to understanding

30. Aspirations baseline data are unavailable for households in Alego subcounty.
31. We note that the components of this index differ from Haushofer and Shapiro (2016).

household health. Our summary index focuses on outcomes for which increases are more easily interpretable as improvements in health status.

As with our primary outcomes, we winsorize the top 1% of the monetary outcomes (outcome 7) as our main specification. We also explore using the unadjusted outcomes and trimming the top 1% of outcomes in an appendix. If we find substantial differences between these approaches, we will investigate the reasons why.

5.7 Education

Summary measure – Education index: Weighted, standardized average of total education expenditure and proportion of school-aged children in school, appropriately signed so that higher values represent better education outcomes.

1. **Total education expenditure in the last 12 months**: Variable 4.4.4. We also explore calculating this as the total school-related expenditure in the 3 most recent school terms (outcome 5), summing across all children.

2. **Proportion of school-aged children in school**: Number of school-aged children (6-16) in the household roster attending school in 2016 (4.2.11), divided by total number of school-aged children on the household roster.[32]

3. **Undertaken new form of education or training***: Variable 5.1.7 for respondents.

4. **No. of days attended school in the last five days school was in session*†**: Variable 4.1.9 for household members between 6-16 years old or currently enrolled in school (set as zero for household members between 6-16 and not in school).

5. **Per child school-related expenditures in the 3 most recent school terms*†**: Sum of school/activity fees (variables 4.2.6,4.2.17,4.2.27) for the three most recent completed terms plus school-related expenses by year (4.2.7, 4.2.18, 4.2.28) and school development projects by year (4.2.9, 4.2.20, 4.2.30), assigned proportionally over the school year for completed terms on the basis of household survey date, for the last 3 completed school terms, for household members enrolled in the current school year, where analysis is conducted at the individual student level.

6. **Days of school missed due to being sent home because of school fees, most recent school term*†**: Variable 4.2.10a,4.2.21a,4.2.31a for household members enrolled in the current school year, depending on school year. Set to zero if child sent home zero times (variable 4.2.10,4.2.21,4.2.31, depending on school year).

32. We will also explore this based on the current school year, where the current school year is defined as 2016 and 2017 for households surveyed at endline in 2016 and 2017, respectively.

As with our primary outcomes, we winsorize the top 1% of the monetary outcomes (outcomes 1 and 5) as our main specification. We also explore using the unadjusted outcomes and trimming the top 1% of outcomes in an appendix. If we find substantial differences between these approaches, we will investigate the reasons why.

5.8 Female empowerment

Summary measure – Female empowerment index‡: Weighted, standardized average of violence and attitudes index, appropriately signed so that positive values reflect more female empowerment/less domestic violence.

1. **Violence index‡**: Weighted standardized average of frequencies of physical, emotional, and sexual violence (outcomes 3, 5 and 7).
2. **Attitudes index‡**: Weighted standardized average of male-oriented attitudes and justifiability of domestic violence (outcomes 10 and 11).
3. **Spouse pushed, twisted the arm of, punched, kicked, choked, or used a weapon on the respondent in the last 6 mo.‡**: Indicator variable for at least one positive value of variables 18.22a-g.
4. **Frequency of physical violence in the last 6 mo.‡**: Sum of instances of variables 18.22a-g(B).
5. **Spouse said or did something to humiliate you in front of others, threatened to hurt or harm you or someone close to you, or insulted you or make you feel bad about yourself in the last 6 mo.‡**: Indicator variable for at least one positive value of variables 18.21a-c.
6. **Frequency of emotional violence in the last 6 mo.‡**: Sum of instances of variables 18.21a-c(B).
7. **Spouse raped or performed non-consensual sexual acts on the respondent‡**: Indicator variable for at least one positive value of variables 18.22h-i.
8. **Frequency of sexual violence in the last 6 mo.‡**: Sum of instances of variables 18.22h-i(B).
9. **Marital control‡**: Sum of instances of variables 18.20a-e(B).
10. **Male-oriented attitudes (respondent)‡**: Sum of indicators for respondent agreeing with male-oriented attitudes, variables 13.3.1-7(a), where 13.3.2 is reverse-coded.
11. **Justifiability of domestic violence (respondent)‡**: Sum of indicators for agreeing domestic violence justified, variables 13.3.8i-v.

12. **Male-oriented attitudes (community)‡**: Sum of indicators for respondent believing community agrees with male-oriented attitudes, variables 13.3.1-7(b), where 13.3.2 is reverse-coded.

5.9 Food security

Summary measure – Food security index: Weighted, standardized index of food security outcomes, appropriately signed so that higher values represent greater food security.

1. **No. of days adults in household skipped meals or cut the amount of meals in the last 7 days**: Variable 11.11a.
2. **No. of days children in household skipped meals or cut the amount of meals in the last 7 days**: Variable 11.11b.
3. **No. of days adults in household gone entire days without food in the last 7 days**: Variable 11.12a.
4. **No. of days children in household gone entire days without food in the last 7 days**: Variable 11.12b.
5. **No. of days children in household gone to bed hungry in the last 7 days**: Variable 11.10b.
6. **No. of days adults in household gone to bed hungry in the last 7 days**: Variable 11.10a.
7. **No. of meals eaten yesterday that included meat, fish, or eggs**: Sum of variable 11.2 and 11.3.

5.10 Labor supply and time use

Summary measure – Total hours worked in the last 7 days*: Sum of respondent's hours worked in agriculture (7.7), self-employment (8.4 when respondent is main decision-maker) and employment (9.8 for respondent) in the last 7 days.

1. **No. of months respondent worked in self-employment or employment the last 12 mo.***: Sum of number of months worked in self-employment enterprise (8.5) if respondent is main decision-maker (0 if respondent not self-employed) plus number of months respondent worked in employment, calculated as month of survey minus employment start date (9.3) if employment working patterns (9.7) full-time or part-time, and based on 9.7a if employment working patterns seasonal.

2. **Respondent currently self-employed or employed/working for pay***: Indicator for respondent main decision-maker for self-employed enterprise (8.3b) or respondent employed/working for pay (variable 9.2 for respondent).

3. **Respondent's total hours worked in employment or self-employment in last 7 days***: Sum of hours worked in employment (9.7 when respondent is household member in employment in 9.2) and self-employment (8.4 when respondent is main decision-maker) in the last 7 days.

4. **Proportion of working-age adult household members working in self-employment or employment***: Number of household members with an occupation in self-employment or employment (4.1.8), divided by number of adults in the household aged 18 to 65 from household roster.[33]

5. **Hours household spent actively searching for jobs, applying for jobs, or in interviews in the last 7 days*‡**: Variable 9.21.

6. **Respondent's hours spent on household chores in the last 7 days**: Variable 6.14.

7. **Respondent's hours spent performing leisurely activities in the last 24 hours*‡**: Hours calculated from time use module (section 17 of the survey), where leisure activities are defined as the following codes: sleep (1), eat (2), bathe/dress (3), religious activity (4-5), rest (6), play with children (13), visit/entertain friends (14), play sports (17), spend time with spouse/partner (18).

As with our primary outcomes, we winsorize the top 1% of the hours outcomes (summary measure and outcomes 3,5,6,7) as our main specification. We also explore using the unadjusted outcomes and trimming the top 1% of outcomes in an appendix. If we find substantial differences between these approaches, we will investigate the reasons why.

5.11 Migration and remittances

1. **Respondent lived in a different (administrative) location for more than 4 months since baseline*‡**: Indicator, variable 16.1.2.

2. **Respondent moved to new (administrative) location for work-related reasons*‡**: Indicator, variable 16.1.2 conditional on variable 16.1.6 reporting work-related reasons.

3. **Respondent lived in an urban area for more than 4 months since baseline***‡: Indicator, where urban areas are classified following Hamory Hicks et al. (2017), based on question 16.1.3d (lives in town/city). As a robustness check, we will also classify this based on

33. We can also construct this based on responses in sections 8 and sections 9 about household members working in self-employment and employment.

whether respondents lived in one of the following cities: Nairobi, Mombasa, Kisumu, Eldoret, Nakuru (the five largest cities in Kenya) or Kampala (Uganda).

4. **Net change in number of household members since baseline**: Number of household members currently in household (other than respondent) (4.1) minus number of household members as of baseline date (4.2).

5. **Any baseline household member migrated to an urban area**: Indicator based on 4.3.4 being migration and on 4.3.7d being an urban area, as defined in outcome 3.

6. **Net value of remittances and goods sent in the last 12 mo.‡**: Sum of variables 16.2.6a or 16.2.9 (depending on number of transfers received) net of 16.2.15a or 16.2.18 (depending on number of transfers sent) across all transfer relationships.

As with our primary outcomes, we winsorize the top 1% of the monetary outcomes (outcome 6) as our main specification. We also explore using the unadjusted outcomes and trimming the top 1% of outcomes in an appendix. If we find substantial differences between these approaches, we will investigate the reasons why.

5.12 Crime and safety

1. **No. of times victimized by theft in the last 12 mo.*‡**: Sum of variables 10.1.28-30(a) (number of times).

2. **No. of times victimized by assault, arson, or witchcraft in the last 12 mo.*‡**: Sum of variables 10.1.31-34 (a) (number of times).

3. **Indicator for unreported crimes in the last 12 mo.*‡**: Indicator for response of "no" to any of variables 10.1.28-35(b) .

4. **Worry about crime or safety in the neighborhood*‡** Indicator (variable 10.1.36) for somewhat worried or very worried about crime or safety.

5.13 Baseline covariates and dimensions of heterogeneity

1. Respondent is female

2. Respondent is 25 years or older

3. Respondent is married

4. Respondent completed primary school

5. Respondent has at least one child in the household

6. Above median value of subjective well-being index
7. Respondent operates a non-agricultural business
8. Respondent earns income from employment

References

Anderson, Michael 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* 103 (484): 1481–1495.

Arnold, Catherine, Tim Conway, and Michael Greenslade. 2011. *Cash Transfers.* Evidence Paper. DFID.

Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli. 2006. "Adaptive Linear Step-up Procedures That Control the False Discovery Rate." *Biometrika:* 491–507.

Deaton, Angus, and Salman Zaidi. 2002. *Guidelines for constructing consumption aggregates for welfare analysis.* Vol. 135. World Bank Publications.

Fisher, Ronald Aylmer. 1935. *The Design of Experiments.* Includes index. Bibliography: p. 245. Edinburgh: Oliver & Boyd.

Hamory Hicks, Joan, Marieke Kleemans, Nicholas Y. Li, and Edward Miguel. 2017. "Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata." NBER Working Papers No. 23253.

Haushofer, Johannes, Edward Miguel, Paul Niehaus, and Michael Walker. 2014. "General Equilibrium Effects of Cash Transfers in Kenya." AEA Trial Registry. November. `https://www.socialscienceregistry.org/trials/505/history/3031`.

———. 2016. "Pre-analysis Plan for Midline Data: General Equilibrium Effects of Cash Transfers." May.

———. 2017. "GE Effects of Cash Transfers: Pre-analysis plan for targeting analysis." September.

Haushofer, Johannes, and Jeremy Shapiro. 2016. "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya" [in en]. *The Quarterly Journal of Economics* (July): qjw025. ISSN: 0033-5533, 1531-4650.

Jack, William, and Tavneet Suri. 2011. "Mobile Money: The Economics of M-PESA." NBER Working Paper No. 16721, January.

Jappelli, Tullio, and Luigi Pistaferri. 2010. "The Consumption Response to Income Changes." *Annual Review of Economics* 2 (1): 479–506.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. "Incentives to Learn." *The Review of Economics and Statistics* 91 (3): 437–456.

Krstajic, Damjan, Ljubomir J Buturovic, David E Leahy, and Simon Thomas. 2014. "Cross-validation pitfalls when selecting and assessing regression and classification models." *Journal of Cheminformatics* 6 (March): 10. ISSN: 1758-2946.

Lee, David S. 2009. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *The Review of Economic Studies* 76 (3): 1071–1102.

Ligon, Ethan. 2016. "Estimating household neediness from disaggregate expenditures."

Lin, Winston. 2013. "Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique" [in EN]. *The Annals of Applied Statistics* 7, no. 1 (March): 295–318. ISSN: 1932-6157, 1941-7330.

Mbiti, Isaac, and David N. Weil. 2015. "Mobile Banking: The Impact of M-Pesa in Kenya." In *African Successes, Volume III: Modernization and Development,* 247–293. NBER Chapters. National Bureau of Economic Research, Inc, March.

McKenzie, David. 2012. "Beyond baseline and follow-up: The case for more T in experiments." *Journal of Development Economics* 99 (2): 210–221.

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1): 159–217.

Parker, Jonathan A., Nicholas S. Souleles, David S. Johnson, and Robert McClelland. 2013. "Consumer Spending and the Economic Stimulus Payments of 2008." *American Economic Review* 103 (6): 2530–2553.

Tibshirani, Robert. 1996. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological):* 267–288.

Walker, Michael. 2017. "Pre-Analysis Plan: Local Public Finance and Unconditional Cash Transfers in Kenya." February.