The General Equilibrium Effects of Large Universal Cash Transfers in Malawi:

Pre-Analysis Plan for Short-Run Household Welfare Impacts using Midline Data

Dennis Egger, David Henning, Dean Karlan, Kanika Phogat, John Walker

22 July 2026

Summary: This document outlines outcomes and regression specifications for estimating the short-term (0-2 year) effects of universal, unconditional, one-time lump-sum cash transfers on household socio-economic outcomes and welfare as part of the General Equilibrium Effects of Cash Transfers project in Malawi. The project is a three-level randomized controlled trial run in collaboration with the NGO *GiveDirectly.* This document describes the analyses that will be conducted to understand the short-run impacts of cash transfers on households. We specify regression equations and the primary outcomes that we intend to study. This document is part of a series of pre-analysis plans that will be filed as part of this project and we also discuss how these plans fit together. We anticipate conducting analyses beyond those pre-specified here, and this document is not meant to preclude additional analyses.

Appendix A: Baseline household census instrument
Appendix B: Baseline household survey instrument
Appendix C: Midline household phone survey instrument
Appendix D: Midline household survey instrument
Appendix F: A macro-experimental approach to studying the general equilibrium impacts of cash transfers

Executive summary

This document outlines the analysis plan for short-run impacts of cash transfers on household welfare in Chiradzulu District, Malawi, using data collected between 2025 - 2027. It is part of a randomized controlled trial of an unconditional cash transfer program conducted by the NGO *GiveDirectly* (GD). GD will provide universal unconditional cash transfers of approximately USD 550 (70 - 100% of average annual household income at baseline) to approximately 104,000 adults in Malawi's Chiradzulu district.

The treatment: All eligible participants in villages allocated to treatment receive approximately USD 550 (adjusted for inflation over time) sent directly to their mobile phones in two installments of USD 50 and USD 500 (adjusted for inflation over time). The modal household receives each transfer roughly one and three months following their villages' enrolment in the program.

Treatment allocation: This is a clustered randomized control trial. We randomize at three different geographic levels to create differential exposure across villages and market catchment areas. The three geographic levels, ranked from most to least aggregate are Traditional Authorities (TAs), Group Village Headman (GVHs), and villages. The randomization is designed as follows:

1. TAs are randomly allocated to a level of saturation, whereby a pre-determined proportion of GVHs are treated, ranging from 40-80%
2. Within treated GVHs, in both the high and low saturation TAs, 67% of villages are treated.
3. Within each village, *all* individuals eligible for treatment are treated.

Inclusion/Exclusion criteria: All individuals above the age of 18 are eligible to receive the transfer.

Primary outcomes: In line with much of the literature in development economics and *GiveWell's* model for evaluating the cost-effectiveness of charities, the primary outcome used for measuring the impact of the cash transfers on household welfare is *1.2. Household total consumption expenditure.* However, to account for the potentially multi-faceted impacts of unconditional cash transfers (documented in a vast literature to date), we additionally specify a total of 10 primary outcomes – groups into two main outcome families. These are:

<u>A - Primary Economic Outcomes:</u>
1. Total value of non-land, non-housing assets
2. Total household consumption expenditure
3. Total household productive investment (ag + non-ag.)
4. Total earned household income
5. Labor supply: Hours worked in the last 7 days

<u>B - Primary Non-Economic Outcomes:</u>
6 Food security index
7 Subjective well-being index
8 Health status index
9 Child health and education index
10 Female empowerment index

Data collection: We conduct a census in every village prior to GD introducing the treatment. From that census we sample on average 12 households in each village.

Sampled households receive the baseline survey immediately after the census. Midlines and endlines are conducted over a 12 month period, with timing of the survey cross-randomized across GVHs, and will cover representative sub-samples each month, between 1 - 24 months after transfer. At the time of writing the endline survey has not been finalized.

Data handling: We winsorize all unbounded variables (e.g. consumption, revenue, etc.) at the 99% level (if strictly positive) or at the 1% and 99% level when both positive and negative values are possible (e.g. profits). We apply winsorization at the latest possible step, i.e. when a variable is the sum of several subcomponents (e.g. consumption), we winsorize total consumption at the end instead of winsorizing each component separately before summing. We run ANCOVA specifications controlling for the baseline of the outcome variable when available in the main regression. If it is missing, we set it to the average and include a dummy.

Primary empirical specification: The main regression used to estimate the effect of the cash transfer program on our primary outcomes of interest is:

$$y_{i(v,c,k,p)} = \alpha_p + \beta Amt_{v(i)} + \sum_{r=d}^{\bar{R}} \gamma_r Amt^{r-d\ to\ r\ km}_{\neg v(i)} + \sum_{r=d}^{\bar{R}} \phi_r s^{elig,r-d\ to\ r\ km}_{\neg v(i)}$$

$$+ \omega_1 y^{BL}_{i(v,c,k,p)} + \omega_2 M^{BL}_{i(v,c,k,p)} + \varepsilon_{i(v,c,k,p)} \qquad (2a)$$

where $y_{i(v,c,k,p)}$ is an outcome for household/individual *i* in village *v* and GVH/cluster *c,* and traditional authority (TA) *k*. $Amt_{v(i)}$ is the amount of money (per adult resident in the village) transferred to households in village *v,* $Amt^{r-d\ to\ r\ km}_{\neg v(i)}$ is the amount of money (per adult resident in the buffer) transferred to households in villages other than *v* within radii bands of *r* - *d* to *r* km around the centroid of village *v*, and $s^{elig,r-d\ to\ r\ km}_{\neg v(i)}$ is the share of adults within *r* - *d* to *r* km that are eligible for transfers. We select the number of radii bands according to a pre-specified algorithm based on a split-sample approach and minimizing the Bayesian Information Criterion (BIC), as specified in Section 4.2.2. We include the baseline value of the outcome variable ($y^{BL}_{i(v,c,k,p)}$) when available to improve statistical precision, set to the mean if missing and an indicator for missingness ($M^{BL}_{i(v,c,k,p)}$). α_p are phase fixed effects (Phase I and Phase II). We weight observations by the inverse sampling probabilities at the village level to be representative of the population of eligible households.

Our primary parameter of interest is β which captures the direct effect of treatment on recipients as well as the Average Total Effect of the program including any spillovers across space (ATotE) and the Global Average Treatment effect of a universal program including both direct and spillover effects (Section 4.2.6).

Standard errors: When reporting direct and within-village effects only (β), we cluster standard errors at the village level. When reporting spillover effects or total effects involving both direct and spillover effects, we pre-specify using Conley (2008) standard errors that adjust for spatial dependence of error terms (Section 4.2.)

1. Introduction

This document outlines the analysis plan for short-run impacts of cash transfers in Chiradzulu District, Malawi, using data collected between 2025 - 2027: Baseline (on average 4 months before treatment), Midline (on average 6 months post-treatment) and Endline 1 (on average 1 year post-treatment). It is part of the General Equilibrium Effects (GE) project in Malawi, a randomized controlled trial of an unconditional cash transfer program conducted by the NGO *GiveDirectly* (GD). GD will provide universal unconditional cash transfers of approximately USD 550 (70 - 100% of average annual household income at baseline) to approximately 104,000 adults in Malawi's Chiradzulu district, representing 40% of the population of the district and 30-40% of district GDP. Transfers are adjusted for inflation quarterly.

Our study aims to understand the effect of transfers at this scale and over time, on households, local businesses, markets, prices, civic engagement, the public sector, and local economic aggregates. We will use data from full censuses of households and enterprises, representative surveys of households, enterprises, and village leaders, and market price surveys at all local markets. We will also compile additional administrative data from the district, the national government, and the national office of statistics relating to public goods, education, health, and electoral outcomes to complement our original data collected. Additionally, we are working on getting access to linked phone meta-data and mobile money transactions from households and enterprises from mobile-phone providers.

For the short-run economic impacts of the program, we are interested in three broad types of (complementary) analyses:

1. the socio-economic and welfare impacts of cash transfers on recipient households (and the spillovers to their neighbors) present in our study area at the time of the intervention, the focus of this Midline Household PAP (Egger et al. 2026a),

2. the impacts of household cash transfers (demand) and complementary supply-side interventions (information and grants to firms) on enterprises and markets (the focus of an accompanying 'Midline Supply Side PAP' that we will file jointly with this PAP),

3. the impacts of cash transfers and supply-side interventions on the local economy more broadly, including on GDP and prices, etc. (the focus of an accompanying 'local economy PAP' that we will file jointly with this PAP).

From a measurement perspective, these differ in their target populations: both include households that have remained in the study area since the intervention, but the former includes households that moved away since the intervention, while the latter includes households and firms that were newly established in the study area after the intervention. In this PAP, we focus on the impacts on households that were present at baseline.

In addition to these two plans described above, we currently plan to file (at least) three additional pre-analysis plans: i) the impacts of cash transfers on local public finance and political / electoral outcomes before and after the 2026 elections, and ii) the impacts of cash transfers on households' use of mobile phones and mobile money, using linked meta-data from mobile phone providers, iii) the impacts of cash transfers on mortality, health, and health-seeking behavior.

Relation to previous work

This project builds on the Kenya General Equilibrium Study, KGES (AEA Trial Registry 505, Egger et al. 2022, etc.). Many of the approaches herein were inspired by those taken in the KGES. We will also build on several refinements in methodology and measurement that we developed since the start of KGES (and which can be traced through the series of PAPs filed and papers written as part of the KGES over time). In particular, relative to the approach in Kenya, this project will have a) larger variation in the intensity of treatment relative to the size of the local economy due to the Malawi program being *universal*, and households being somewhat poorer at Baseline while the Kenya program was *targeted*, b) much higher-frequency data collection for both households and enterprises, particularly in the short-run, c) a more comprehensive sampling frame for local enterprises in homesteads, villages, and markets separately, d) more detailed measures of imports and exports to/from the study area, and the spatial distribution of financial flows within the study area, e) better measurement of slack / utilization throughout the cash transfer period (Walker et al. 2025), and f) more detailed measures of supply-side adjustments, including firms' optimisation decisions, capital stock, investment, price setting, mental models, etc.

The project also builds on 'High Frequency Engel and Supply Curves in General Equilibrium: Experimental Evidence from Large Universal Cash Transfers in Malawi' (AEA Trial Registry 12550). This was a GD pilot of the same individual-level universal cash transfer program conducted in Khongoni TA in Malawi. As part of that project, we collected high-frequency monthly phone surveys on expenditure, and rich market censuses and price data during the first 12 months after cash rolled out. Where appropriate, we will compare findings from Khongoni to those during the present study in Chiradzulu, particularly on short-term expenditure and price impacts. We may also combine samples for some analyses.

While the intention is that this PAP can be read and understood by itself, much of the detail and thought development can be traced in these earlier studies, and we will sometimes err on the side of brevity to avoid repetition.

Analyses to Date

To date, research assistants and the survey firm have had access to baseline data collected in each of the 10 TAs in Chiradzulu – for data quality and enumerator productivity checks and assessments. While baseline data collection has been completed and we have commenced Midline data collection, we have not yet attempted to estimate any treatment effects, except for checking for balance in tracking rates/attrition across treatment and control areas as part of regular high-frequency checks.

Note: We write this PAP before the Endline 1 survey instruments for households and enterprises were finalized. We may therefore amend this document to incorporate any future changes in survey instruments, but will do so *before* conducting any analyses of treatment effects using Midline or Endline 1 data.

2. Research Design

2.1 The GiveDirectly Intervention

In this project, the NGO *GiveDirectly* delivers one-time lump-sum cash transfers. They are unconditional, universal (all adults over the age of 18 are eligible), and delivered to individuals via mobile money. Transfers amount to approximately 550 USD per individual, will be adjusted quarterly to keep in line with national inflation, and are delivered in two tranches of 50 USD and 500 USD approximately one, and three months after enrolment.

Stage	Description
Pre Census + Baraza	GiveDirectly reaches out to the village chief to introduce the program. They conduct a visit and create a village list to estimate the population. GD then hosts a baraza (full village meeting) to announce the program (typically the day after).
Census	1st visit to households within the village to collect basic demographic information from households. This takes place the day after baraza and typically lasts a few days up to maximum a week. ● Recipients are informed of the transfer amount and transfer schedule. ● Phones are distributed to recipients who do not have them. ● SIM cards are distributed to those who do not have them or prefer to have their mobile money on a new SIM card.
National ID Drives	GD has partnered with the National Registration Bureau (NRB) to run ID drives in parallel with the enrollment process. NRB representatives bring mobile enrollment kits to register individuals with national IDs if they don't have them. They collect biodata (photos and fingerprints) as well as name, birthdate, and family information.
Registration	SIM cards are linked to national ID numbers and recipients self-enrol via USSD in the program. Recipients are provided with brochures and information to facilitate self-enrolment during SIM registration events. Staff conduct in-person support during and following registration events for individuals who struggle to self-enrol.
Pre Pay Audit	The household/recipients are audited to detect any safeguarding/fraud-related issues. This only applies to 10% of the eligible households.
Case management	Clear any cases blocking recipients from advancing to **'pay ready stage' e.g** difference between the collected account number provided by the recipients and the account number that popped up during name verification. Blockers to move to "stage pay" include: - No ID ("pending identity verification") - Unsuccessful SIM registration - Pending audit - an audit flag has not been resolved
Payment 1	First tranche of payments ($50). Typically the last week of the month.
Follow Up	Validate cash receipt and monitor resulting risks
Payment 2	Second tranche of payments: $500 (typically 2 months after payment 1). At this stage, the costs of the mobile phone are deducted for those who opt for a phone.

The transfer is universal, and large – both for individual households as well as relative to the economy overall. At baseline, households have approximately 2.2 eligible adults on average, and reported an annual consumption value (including both expenditure and home production) of 4826 PPP-adjusted USD, and 3280 PPP-adjusted USD of expenditure (at IMF's 2025 projected PPP conversion rate for Malawi). Transfers thus represent approximately 71% of annual household consumption, and 105% of households' annual cash expenditure at baseline.

Additional interventions

At the same time as rolling out their core unconditional cash transfer program at-scale in Phase I – the 4 Southern TAs in Chiradzulu – *GiveDirectly* was piloting several modifications to their core program, including:

1. Varying the transfer size: Adults in 5 randomly chosen villages received twice the amount (approx. 1100 USD), and 750 USD in 9 randomly chosen villages
2. Providing smart phones: Recipients in 6 randomly chosen villages (3 in Phase I and 2 in Phase II) received smartphones instead of the usual 'non-smart' phone offered
3. AI support tool: Recipients in 3 randomly chosen villages received both the smart-phone, and a newly developed AI tool

These additional interventions were small-scale, and confined to Phase I only. We are therefore not statistically powered to detect any differential effects of these additional interventions.[1] Our effects can therefore be implemented as the *average* effect of the core program (implemented in 98% of villages) and these additional pilot interventions. Given the small share of villages with these modifications, this average will be very similar to the effects of GiveDirectly's core program alone.

2.2 Randomization of cash transfers

We study the impacts of cash transfers in a large-scale cluster-randomized control trial in Chiradzulu district in Malawi. The district is administratively organized into 10 traditional authorities (TAs), 112 Group Village Headman units (GVHs), and 898 villages. The district has a population of approximately 460,000 (116,000 households) of which there are about 255,000 adults who are eligible to receive GiveDirectly transfers.

Randomization of cash transfers will be across 3 different geographic units (villages, GVHs, and TAs) in order to create variation in exposure to treatment not only at the village level, but also at higher geographic orders (including market catchment areas), allowing us to estimate treatment effect spillovers across larger geographies.

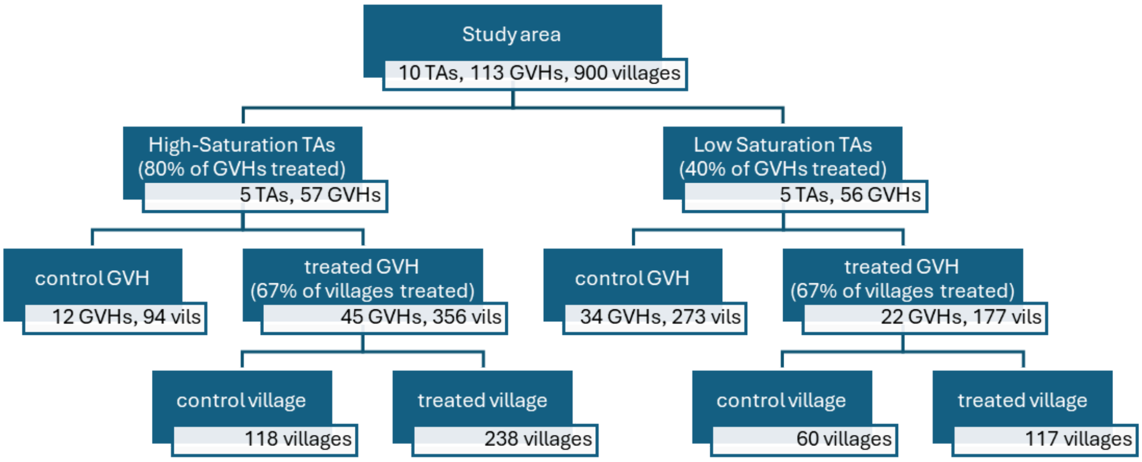
1. *TA Saturation:* Variation in the intensity of treatment across TAs generates variation in the geographic exposure to treatment of local economies. High-saturation TAs have approx. 80% treated GVHs, medium-saturation TAs 60%, low-saturation TAs 40%.

[1] We will, however, consider the different transfer amounts in spatial regression specifications involving treatment amounts, and when computing average treatment amounts across households and individuals. We return to this in Section 4. We will also show robustness of the core reduced-form average treatment effect to sub-sampling to only villages that received the USD 550 transfer.

2. *GVH Saturation:* Within treated GVHs, approximately ⅔ of villages will be treated. Retaining control villages in treated GVHs allows us to estimate within GVH spillovers. In control GVHs, no villages will be treated, acting as pure controls.

3. *Village-level treatment:* Within treated villages, treatment is *universal* and at the individual level, i.e. every adult will be eligible to receive treatment.

Figure 1 - Cash Transfer Randomization



Notes: This figure depicts the ideal randomization design. In practice, there may be deviations from this figure. For example, logistical challenges prompted us to create 3 different saturation levels at the TA-level in Phase I: Low-Saturation (40%), Medium-Saturation (60%), and High-Saturation (80%).

This design generates variation in treatment exposure at various different geographies: Since treatment is universal, cash transfers will amount to 71% of baseline consumption (a proxy for local GDP), and 105% of total cash expenditure in treated villages. This translates to approximately 45% of GDP in treated GVH, and 28% of GDP in Chiradzulu as a whole. This is substantially larger than any previous cash transfer RCT. For comparison, the Egger et al. (2022) study amounted to 15-25% of GDP in treated villages, and 4-5% of GDP across the study area Siaya County (an area roughly the same size as Chiradzulu district).

Timing of transfers

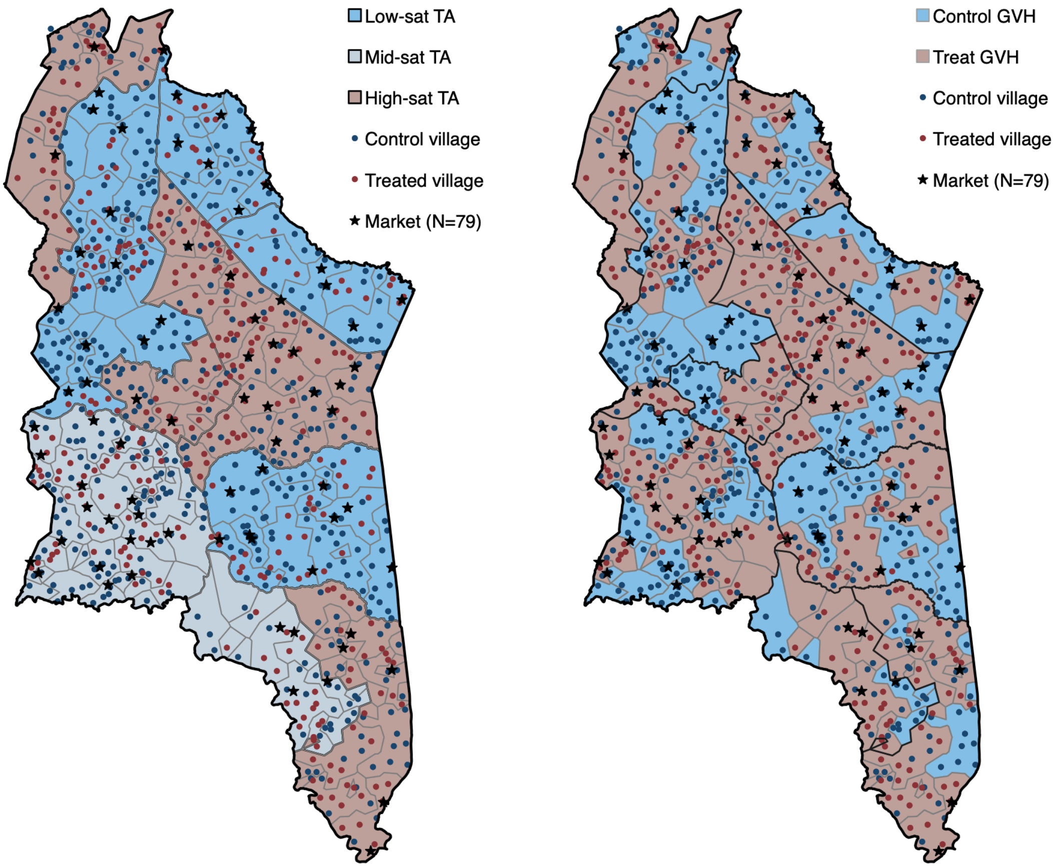
Due to elections taking place in Malawi on 16 September 2025, the roll-out of cash transfers in Chiradzulu was split into two phases (see Figure 2). In Phase I, four TAs (and approx. 400 villages) in the south of Chiradzulu were enrolled before the election (between March - August 2025). For Phase II, the remaining six Northern TAs (and approx. 500 villages) are being enrolled post-election between October 2025 and April 2026.[2]

The spatial variation in cash transfer intensity creates variation in cash transfers across different geographies. To ensure that the arrival of cash in each geography is concentrated in time, all treated villages within a GVH receive cash at the same time. Moreover, GVHs are grouped into ‘market catchment areas’ based on the ‘most important’ weekly market as stated at baseline by randomly chosen village members or by geographic distance where

[2] We will use the fact that Phase I was enrolled pre-election to study the impact of the cash transfers on local politics, public goods, and electoral outcomes. These analyses will be specified in a separate pre-analysis plan.

this data was not available.[3] These market catchment areas are then randomly ordered, and grouped into monthly cohorts, all of which will be treated at the same time (denoted by 'cohorts' in Figure 3). We will use this spatial variation in treatment intensity to estimate the spillover impacts of cash transfers on non-recipient neighbors and enterprises in these market catchment areas as well as local economic aggregates.

Figure 2 - Map of Chiradzulu District



Note: In treated GVHs, ⅔ of villages are treated, while in control GVHs, none are. Within treated villages, all adults receive a transfer. Red dots mark the 79 market centers in the district, defined as markets that occur at least weekly, and where at least 20 sellers are present on a typical market day. For randomization, the administrative capital, Chiradzulu Boma, was absorbed into TA Kadewere/GVH Chilanga and TA Mpama/GVH Mbalame.

Exceptions

1. **TA saturation:** Phase I had 4 TAs. To maintain balance and overall saturation goals, one TA (Likoswe) was assigned to medium saturation (halfway between high- and low-saturation) non-randomly. The other two were randomly assigned to high- and low-saturation respectively.

[3] We define a market as an agglomeration of at least 15 traders who operate simultaneously. Markets were identified both from household reports of purchases and enumerator drive-throughs of the district. We collected a short census of important characteristics for each candidate market. Our final panel of markets consists of markets with at least 15 traders when visited on their main market day, during operating hours. Two markets were identified and added to the sample after the commencement of the cash transfer program.

In the process of baselining, the team discovered that one of the TAs had in fact been split into two, such that there was not a fourth TA (Maoni), which we also non-randomly assigned to medium saturation. Moreover, TA Maoni was treated all at once at the end of Phase I – thus, Maoni's timing is also not fully random to other Phase I villages.

For analyses of TA-level saturation (not primary), we will exclude medium-saturation TAs, since assignment was not random. For all other analyses, we include 'medium-saturation' as an additional control. When analysing timing relative to transfers, we will include a dummy for the cohort from TA Maoni (in addition to including dummies for each Phase I / II cohort respectively).

2. **TA head villages:** GiveDirectly determined that operationally, it was crucial that TA head villages – where TA administrators reside – were part of the treatment group. Thus, these 10 TA head villages were non-randomly assigned to the treated group, and will therefore be excluded from any analyses of treatment effects.
3. **Accidental treatment:** GD accidentally enrolled two villages. We consider this quasi-random, and include treated villages in the treatment group in primary analyses, but may present results that are robust to excluding these villages from the analysis.

2.3 Supply-Side Interventions

Recipient households will spend a significant share of their transfer at one of the 79 market centers (defined as an agglomeration of at least 15 enterprises or sellers open to customers at least weekly, but often bi-weekly or daily). The exposure of each market catchment area depends on the geographic shopping patterns of local households, i.e. where they are spending their money, supplying their labor, and selling their produce.

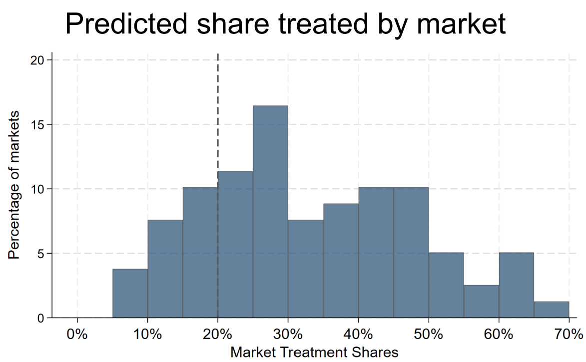
These markets are dominated by informal micro-enterprises. Based on the baseline census of all enterprises operating in all Phase I market centers within the district, about half are food sellers, 27% supply local services (e.g. barber shops, motorcycle transport), 18% are non-food vendors (e.g. charcoal, wood), and 8% are in manufacturing (e.g. tailoring, carpenters). Firms are generally small: only 15% have paid employees (50% have unpaid workers) and monthly revenues (profits) are approximately 350 USD (65 USD) on average. There is some variation across firms, but less than 1% of firms have more than 3 employees.

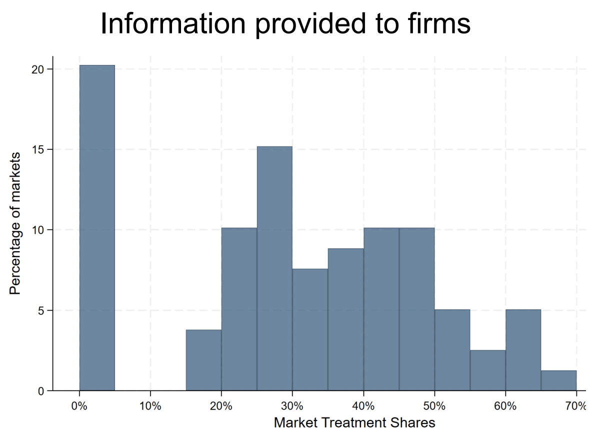
These firms are likely to experience substantial increases in demand for their products due to cash transfers (e.g. as in Egger et al. 2022). Existing work has shown that small enterprises in LMIC are often constrained in their growth, including through information and/or credit constraints. However, a yet untested idea is whether they may also be constrained by the overall economic conditions, such as local demand for goods and services as the Big Push theory suggests. In this study, we test this by cross-randomizing information provision, and grants to businesses in local markets within Chiradzulu, which are facing *randomized* increases in local demand due to GiveDirectly's cash transfers in the district. Our hypothesis is that ex-ante information about this upcoming demand shock – allowing firm owners to plan better – as well as grants to businesses may lead to a more elastic increase in local supply, higher returns to capital for enterprise owners, and, as a consequence, lower inflationary pressure and larger multipliers / welfare gains from the two interventions combined, compared to cash transfers alone.

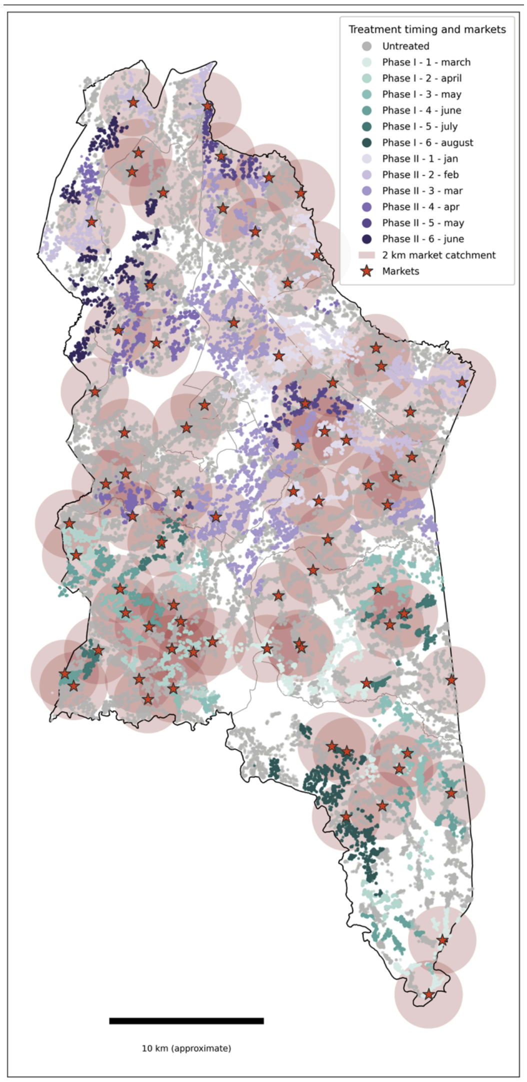
Figure 3 - Exposure of local markets to cash transfers

Panel A: Distance-weighted share of treated households

Panel B: Treatment timing







Note: These figures illustrate the extent of variation in the share of households within market catchment areas that receive transfers in Chiradzulu. As a simple heuristic, we use geographic proximity – though we will later also make use of the detailed baseline data on the geography of expenditure data for each product to assign households to market catchment areas. The three-stage randomized design creates substantial variation across the full range from 0% to 100% of households treated within 2km of a market center. No information is provided to firms if it does not meet the threshold of a 17% increase in predicted revenue or a maximal monthly peak of 10%.

We build on the ongoing cluster-randomized cash transfer rollout, and cross-randomize two distinct treatments to local business owners, targeting a sample of 9,846 enterprises operating in all market centers within the district:

- **T0 - Cash transfer intensity:** As described in the previous section, the clustered randomisation of cash transfers generates variation in each market's exposure to increased spending by cash transfer recipients. This will create variation in predicted demand increases – some markets will have very few households within their market catchment area that are treated, and thus small predicted demand shocks (which we call here 'low-demand'). Others will have a high share of their baseline customers that receive cash – the 'high demand scenario'.

- **T1 - Information (Phase I):** In this treatment arm – conducted in the Southern half of the district, and Phase I of the cash transfer rollout (Mar 2025 - October 2025) – we provide information to enterprises at markets about expected increases in demand for their products or services. We produce product-category (food, non-food, durables or non-food non-durable)-level predictions of demand effects that we provide to enterprises one month before the first major peak of "Payment Two" cash transfers going out within their catchment area. We define catchment areas either using baseline expenditure shares from household expenditure surveys, or by projecting weighted expenditure on a given product category within X km of the home to nearby markets.

  For example, we first provide the following general information about the cash transfer program to a market with a peak in January:

  - "In the coming months, an NGO will be giving out large cash transfers to households living in this area. A large number of households will receive a large, one-off payment of 750,000 Kwacha to 1.5 million Kwacha. Because people will have more money, they will be able to spend more at this market. This will offer you the opportunity to expand your business and increase your profits."
  - "In your market, we expect a large increase in spending over the three months starting from late January. After that, we expect spending will remain high over the coming year, but begin to decline again, and end up only slightly higher than normal levels."

  We then provide information about the predicted impact of these cash transfers on their business:

  - "From late January to late March next year, on average, we expect sales of [PRODUCT] to go up by {[PREDICTED % TREATMENT] x [PRELOAD SALES]} compared to what they would normally be. So, rather than [SALES] in sales you made in the last month, we expect you to earn a total of around [SALES + PREDICTED TREATMENT] in sales in February, the first full month cash transfers will go into this market."

  - "In the last 30 days, you sold [AMOUNT] of [QUANTITY] of your main product, [PRODUCT]. The increase in spending implies selling [QUANTITY SOLD IN LAST 30 DAYS x [PREDICTED % TREATMENT] more [PRODUCT] if you do not change your prices."[4]

  We form predictions using high-frequency treatment effects from our study conducted in Khongoni TA (treated in 2023/4). We first calculate the proportion of treated households within the market catchment using household counts from our census and village-level

[4] To make this more concrete to traders who may not accurately record revenues, we provide a back-of-the-envelope comparison in terms of quantity sold.

treatment information.[5] We then multiply this through by the impulse response function for a given product category to form a prediction for changes in revenues.

All markets experience some increase in predicted demand, with at least some villages treated within their catchment area. We provide information only in markets where we predict an increase in demand of at least 17% with a maximal predicted monthly peak of at least 10%. In Phase I, 28 of the 37 markets were treated under this eligibility criteria. Within treated markets, 50% of baselined traders in treated markets were selected to receive the information treatment, stratified by market, whether the enterprise operates from a structure and whether the enterprise sells only raw (unprocessed) food.

After providing this information, we talk business owners through setting a plan for responding to this information, and committing to these goals using goal-setting tools that have been tested and proven effective in other settings.

This treatment will allow us to understand how firms respond to anticipated vs. unanticipated demand shocks, how well they predict future demand, and how they reason about setting prices, and making input choices. The hypothesis is that upfront information will allow traders to make anticipatory adjustments (e.g. through stocking up, hiring workers, getting loans) that may enable them to better absorb the increase in demand, leading to larger overall gains for the local economy.

- **T2 - Business grants + information (Phase II):** This treatment – conducted in the northern half of the district and Phase II of the cash transfer rollout (Jan 2026 - Apr 2026) – provides business owners with labeled business grants, delivered via mobile money, on average approximately one month prior to when households in the business' main market catchment area receive their transfers. This treatment is provided in combination with the identical information intervention from Phase I.

  Enterprises are eligible for business grants if they sell any products or services other than raw foods, or if they operate from a fixed enterprise location (i.e. not on the floor/roaming) at any of the 42 Phase II market centers in the district. Transfers are tiered, and targeted based on firms' capital needs as indicated by our data from Phase I:

  a) non-food enterprises without a fixed structure receive 500 USD
  b) enterprises operating from a fixed structure (table, stall, building) receive 1,000 USD

  Amounts were chosen based on data from the information intervention in Phase I of the study, and calibrated to meet the average business' stated capital requirements to implement their optimal expansion plans. The amounts are equivalent to roughly one fifth of yearly average revenue, and equal to yearly profits.

  The grants are *labelled* strongly as intended for business purposes, to expand the business. To reinforce this, recipients go through a light tough goal setting and planning exercise identical to that in the information intervention (T1). This allows us to test whether information alone is sufficient to expand local supply, or whether firms are credit-constrained, such that there are substantial additional excess returns to capital.

  Business grants are randomized at the firm level, and stratified by market, and grant tier (500 vs. 1000 USD). Overall, 50% of enterprises receive grants. To estimate market-level effects, and spillovers to businesses that did not receive the grants, we

[5] We adjust for the share of households across the study boundary by assuming equal population density on either side.

generate variation in market-level intensity of supply side transfers. In 22 randomly chosen high-saturation markets, two thirds of enterprises receive grants, in the remaining 20 low-saturation markets, one third receives grants.

35 of the 42 markets are information-treated markets where the total predicted increase in demand was more than 17% and there was at least one month with a peak of at least 10% predicted revenue increase. Firms that received the grant always receive the planning intervention, whether or not they receive the information treatment.

The randomization of markets into high- and low-saturation is stratified by demand exposure intensity.[6] Moreover, this treatment allows us to estimate the returns to capital – the additional profit firms are able to generate from these grants – in markets where the size of the demand shock varies *experimentally.* This is a direct test of the big-push logic: Does increased demand increase the returns to investment?

Table 1 - Supply-side randomization

	All businesses within a market in the study area (N=9,846) Of which sampled (N=5,969)			
	Phase I (N = 2,700)		Phase II (N = 3,269)	
T0: Demand intervention (GD cash transfers)	Control (N = 1,780)	T1: Information only (N = 1,088)	Control (N = 1,319)	T2: Business grant + information (N=1,950)
‘High-demand’ / high-saturation[7]	N = 1,080	N = 1,088	N = 956	N = 1463
‘low-demand’ / low-saturation	N = 700		N = 363	N = 487

Interactions between business information/transfers and household transfers

91% of business owners eligible for supply-side interventions live within Chiradzulu district, and are potentially eligible for household cash transfers. However, existing research shows that firm owners spend a relatively small amount of the unconditional cash transfers delivered to their household on their business (on the order of 100 USD per household). This is far less than data from Phase I suggests is necessary to achieve their optimal growth trajectory for their business (see Table 2). Labeled business grants for business owners may therefore create an additional supply-side boost over and above unconditional household grants alone. Moreover, firm owners are likely to spend their business income within the local economy again, leading to second-round impacts on local economic activity.

[6] We note that we will likely be somewhat underpowered to detect firm-level spillovers. However, there is a prospect of expanding the program to a second district in the future, and thus the possibility of expanding sample size for spillover analysis.
[7] We defined a high-demand market as one that receives a high proportionate treatment impact for illustrative purposes. We will use the full range of geographic treatment intensity to estimate market-level impacts of cash transfers as specified below.

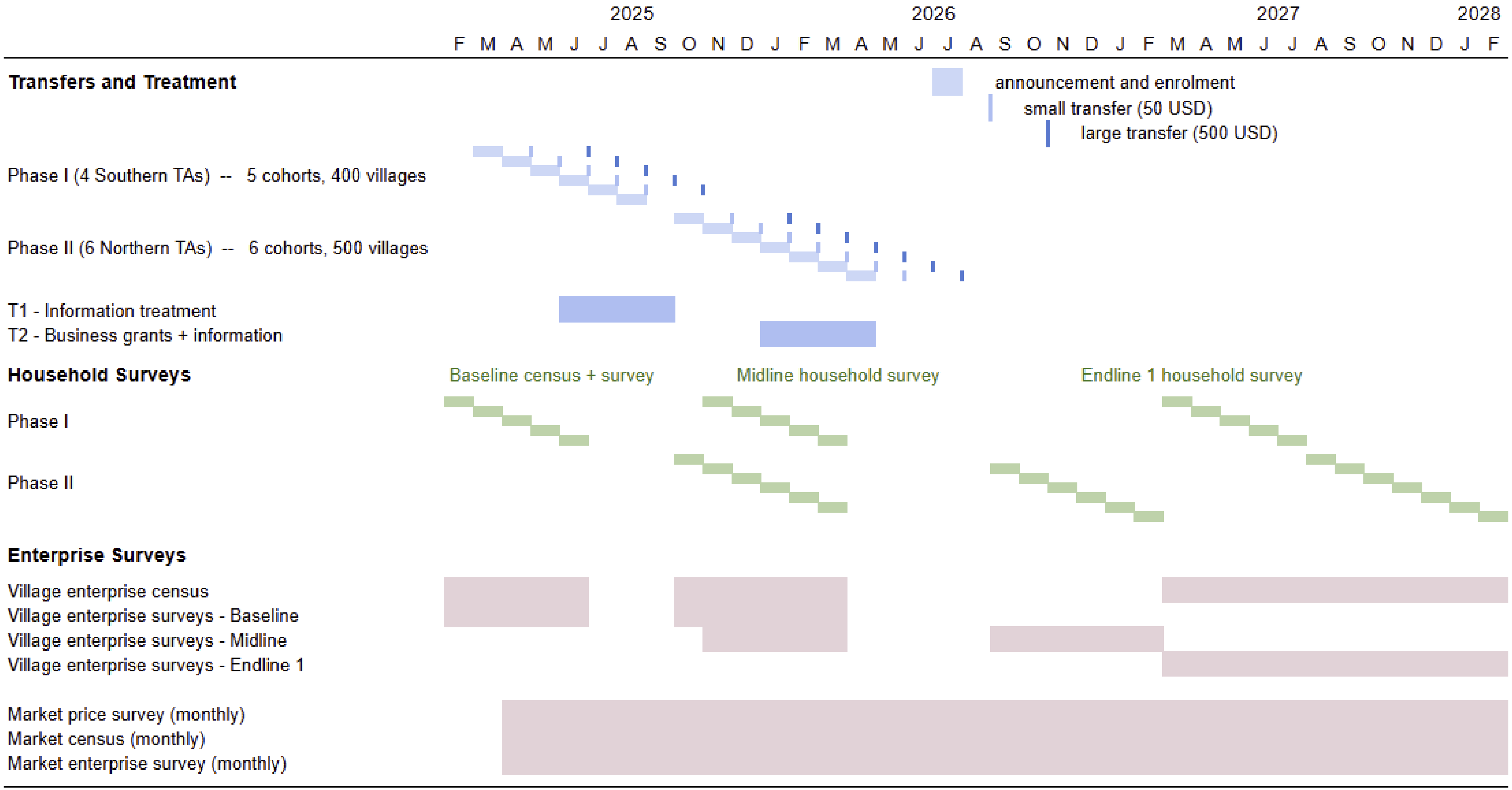
We will file a separate pre-analysis plan for the analyses i) relating to studying the impacts of information and business grants on local enterprises, and ii) relating to the impact of cash transfers on local enterprises, prices, markets, and local economic aggregates. We will also try to understand to what extent the local economic impacts of cash are changed or amplified by these supply-side interventions, and what fraction / share of the overall direct, spillover, and multiplier impacts are explained by these additional interventions relative to household transfers alone.

For the purposes of this PAP on household socioeconomic and welfare outcomes, we note that supply-side information and transfers are *cross-randomized* with household cash transfers. That is, treated and control households are equally likely to be business owners receiving supply-side information or grants. The same is true for *cross-randomization* at the local-economy level: high-intensity geographic areas are equally likely to receive information and business grants as low-intensity geographic areas. Under the assumption of additively separable effects (which seems reasonable in the context of cash transfers, see e.g. Kondylis & Loeser 2024), direct and spillover effects of household-level cash transfers alone – abstracting from the additional impact of supply-side interventions – can be consistently estimated without including interaction terms or terms for the intensity of supply-side treatments within the local economy. This will be our primary approach here.

Moreover, the share of households receiving any information or business grant is small – overall, there are approximately 1,000 business owners receiving additional information, and 1,650 business owners receiving labelled business grants + information. In a district of approximately 220,000 adults, and over 100,000 households, this is less than 3% of the population. And, while approximately 70M USD will be transferred to households, the total amount of business grants will be only 1.7M USD. This further justifies our primary approach for excluding supply-side intervention terms when analysing the impacts of unconditional household cash transfers on households in this PAP.[8]

[8] We will test for robustness of our findings to excluding households that received a supply-side cash transfer (T2) from both treated and control groups.

Figure 3 - Study Timeline



3. Data

This project collects data on households, enterprises, local leaders, markets, and prices. We will also compile additional administrative data from the district, the national government, and the national office of statistics relating to public goods, education, health, and electoral outcomes to complement our original data collected. And, we will have access to linked phone meta-data and mobile money transactions from households and enterprises from mobile-phone providers. In this pre-analysis Plan, we focus on data collected on households (marked green in Figure 3).

Baseline Household Census

Baseline Household censuses are conducted in each village between February 2025 - July 2025 (Phase I), and between October 2025 - March 2026 (Phase II). Censuses take place approximately 1 month (or more) prior to village barazas (meetings) held by GiveDirectly. At the time of our baseline, GiveDirectly may be known as an organisation, but individuals and villages will not have been informed about their treatment status.

The census captures household contact information, GPS, basic composition and demographics, and basic assets and economic activities of the households. We also collect a list of all self-employed enterprises that households operate, which will serve as a sampling frame for village-based enterprises in Phase I. (In Phase II, we conducted additional baseline village enterprise census for any enterprises within the village but operating outside the homestead). The protocol is as follows:

1. Enumerator teams visit each village, interview a village leader to delineate boundaries and organize guides. In line with GiveDirectly's approach, only officially gazetted villages are recognized. Enumerators then proceed to complete a full census of all households in the village, attempting each household at least twice. For households which are not present or where no knowledgeable household member was able to be surveyed, we collect enough information to identify the household during mop-ups (descriptions, GPS, neighbor information, etc.).

2. **Sampling:** Upon completion of the village census, we sample households for the baseline household survey. All permanent households (excluding holiday homes) with at least one adult household member were initially eligible for sampling. Non-consenting households, and those for which we are not able to get any tracking information (names or phone numbers) are then dropped from the eligible sample.

   In Phase 1, households were sampled shortly after completing the census in each village (typically the following day). In Phase 2, sampling took place only after census work had been completed across all villages in a particular GVH. In Phase I, we sampled 8 households per village on average initially (and increased this to 10 households after the first month). In Phase II, we will sample 12 households per village on average. The sample size in each village v is a compromise between equal sample size for each village (maximizing power at the cluster level), and population representativeness, and is determined by:

$$N_v^{sample} = N_{min} + 4 \cdot (N_v - \bar{N}_v)/\bar{N}_v$$

where N_{min} is the minimum sample (4, then 6 for Phase I, then 8 for Phase II), N_v is the number of households censused in village *v,* and $\bar{N}_v$ is the (estimated) average number of households across all villages (in each Phase).[9]

3. At the same time as baseline surveys, households where no appropriate member could be found during the census are visited again, on at least two different days to complete the full census and update information (mop-ups). Households where no knowledgeable adult could be found during the census are still eligible to be sampled for baseline surveys if they have any tracking information to maintain representativeness.

4. The target respondent for all future surveys is selected as the primary male or primary female. In households with both a primary male and primary male, we randomized the target respondent with equal probability.

Consistency with GiveDirectly treatment geography: We follow GiveDirectly's village censusing procedure and delineation as closely as possible to ensure compatibility between their treatment assignment and our classification. Our team was trained by GD staff, and accompanied their team on one of their barazas and censuses. Nonetheless, there may be discrepancies in village boundaries and definitions. We will work with GD to correct such misalignment ex-post using village boundaries and GPS collected by GD during enrolment of recipients, which we attempt to match to our own household census data for accurate determination of treatment status at the household level.

Data cleaning: After completion of the census, we generate a final census by checking the data carefully for duplicates and inconsistencies across villages. This forms our baseline universe of study households.

Baseline Household Survey

The baseline household survey is conducted immediately after the census (typically in the same week, and approximately 1 month prior to barazas/cash transfer enrolment). The sample is representative of all permanent households across all villages, and the survey collects data on contact information, household demographics (roster), birth history and maternal care, dwelling characteristics and assets, household finance and transfers, consumption and expenditure, agriculture and livestock, self-employment, employment, labor supply, and unemployment, time use, transfers and government assistance, health and nutrition, psychological wellbeing and disposition, female empowerment, and civic attitudes and participation.

We attempt to reach each sampled household (see sampling strategy above) at least 3 times on at least two different days. If we did not find a knowledgeable adult household member to survey, households were replaced from a randomly ordered replacement list.

[9] In Phase I this was estimated at 156 households per village ex ante, and was 134 ex-post. We augmented the sample for midlines ex-post using true/final village census numbers.

Midline Household Survey: December 2025 - February 2027

The midline household survey collects short-run outcomes on household consumption and spending, asset accumulation and investment, household finance and transfers, labor supply, as well as basic information on household economic activity. The focus here is on understanding the immediate short-run impacts of cash transfers and how they are spent by recipients, while Endline 1 collects a more detailed set of data.

For Phase I villages, midline data collection falls right after the 2025 election period in Malawi. For this set of households, the midline survey also contains a detailed module on civic attitudes and engagement, and political and electoral outcomes.

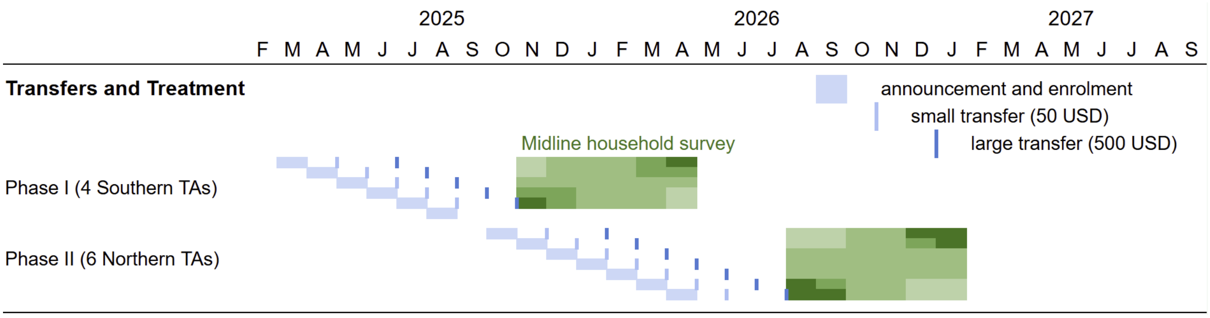
Sampling: The midline sample will consist of the baseline sample, including any households that were missed at baseline, and all replacement households reached at baseline. The total sample size will be augmented by setting $N_{min} = 8$ for all villages, to achieve an average sample size of 12. Moreover, midline sampling in Phase I additionally includes 'mop-up' households (~2.8% of the Phase I census universe) whose details had not yet been collected at the time of baseline sampling, to increase the representativeness of the sample at midline. Any duplicate cleaning as well as any ex-post additions post-baseline will also be taken into account to achieve an equal sampling probability for the cleaned universe of all households present at baseline.

Additional Enterprise Sample: In order to study potential complementarities between unconditional grants to households and grants to business owners at the household level, we will augment the household sample for Phase II Midline/Endline surveys by adding all households in which any business owners who are eligible for supply side cash transfers reside (4.7% of all Phase II households). This will allow us to look at differential impacts (among the sample of business-grant eligible households) of unconditional cash depending on whether households also received business grants (and of business grants depending on whether households also received unconditional cash).

	Household cash recipients (40%)	Household cash control (60%)	Living outside Chiradzulu
Business grant recipients (N = 1947)	678	1078	191
Business grant control (N = 1313)	459	741	113

Timing: Midline surveys will take place over approximately 12 months, to represent the seasonal fluctuations of the Malawi economy over the year. The timing of midline household surveys will be cross-randomized (at the GVH level for operational feasibility) with the timing of transfer receipt (which itself was randomized). This generates exogenous variation between when a household is surveyed, and when individuals received their large transfer (see Figure 4 for an illustration). Overall, surveys will be spaced between 0 to 10 months post transfer, with an approximately even sample size of 1,000 surveys each month to allow us to reliably estimate monthly, bi-monthly, and quarterly treatment effects of cash transfers in the short run (see dynamic impacts below).

Figure 4 - Illustrating survey timing variation



Notes: This figure illustrates the variation in timing between midline surveys and transfer receipts. Each cohort of recipients is surveyed over 6 months, with GVHs split across the different months. Darker shades of green represent a larger share of the cohort. This unequal allocation of shares across months ensures an even number of surveys for each month post transfer.

Anticipatory phone surveys: To test the impacts of the *announcement* of cash transfers, *before* the actual transfer, we conducted phone surveys with a subset of households. In Phase I: we targeted all households that reported at least one phone number at baseline (68.1% of all households), for the two cohorts that had an experimental date for their small transfers in April and May, with a total sample size of 1096. In Phase II: we targeted all households that reported at least one phone number at baseline for the cohort who received their small transfer in May, a total of 776 targeted households. Since cohorts were randomly ordered, this is a representative subset of all Phase I households. These phone surveys were conducted between when GD enrolled households, and before they received (or would have received) their small transfer. Typically, approximately 1 to 4 weeks prior to transfer receipt. Phone surveys collect a subset of the data collected at Midline, focusing on spending, loans, and labor supply, as these are the most likely short-term effects of the announcement of cash. The instrument is in Appendix C.

Endline 1 Household Survey: January 2027 - December 2027

The Endline 1 household survey will collect detailed household level data to study the impacts of cash transfers on recipient and non-recipient households in the medium run – approximately 11 - 21 months after recipients receive the large transfer. It will target the same sample as the Midline survey, and collect all data collected during the Baseline household survey, as well as – potentially – additional modules. At this point, the survey instrument is not yet finalized, and we may add additional outcomes to a later amendment to this PAP, which we will file BEFORE we conduct any analyses using Endline 1 data.

Timing: As with the Midline household survey, the Endline 1 household survey will span an entire calendar year to capture seasonality. We will follow the same GVH randomization for survey timing as for Midline, ensuring variation in the time since the large transfer between 11 and 21 months, with approximately 1,000 surveys in each month.

4. Empirical Strategy

As outlined in Section 1, the main purpose of this analysis plan is to study the short-run average treatment effects of cash transfers on recipients and their neighbors present in the study area at Baseline. This includes both direct effects on recipients as well as any potential within and across-village spillovers affecting both recipients and non-recipients.

Our analysis largely follows Egger et al. (2022), including subsequent updates developed as part of the KGES project (including a more powerful instrument for spatial regressions, and additional approaches to spatial regressions, e.g. by Faridani & Niehaus).[10] We estimate impacts on recipient households (both direct and spillovers), and on non-recipient households (those living in control villages) present in the study area at the time of transfers (Baseline).

We note that this is our current thinking to date, and that we may consider additional methodological refinements in the coming years to keep up with state of the art econometric techniques.

4.1 Reduced-form specification

Households experience both direct effects of cash as well as potential within and across-village spillovers. A relatively crude way of capturing these spillovers, but one we consider a useful benchmark is the following specification:

$$y_{i(v,c,k,p)} = \alpha_p + \beta T^{v}_{v(i)} + \gamma T^{GVH}_{c(i)} + \delta HighSat_{k(i)} + \omega_1 y^{BL}_{i(v,c,k,p)} + \omega_2 M^{BL}_{i(v,c,k,p)} + \varepsilon_{i(v,c,k,p)} \qquad (1a)$$

where $y_{i(v,c,k,p)}$ is an outcome for household/individual <u>*i*</u> in village *v* and GVH/cluster *c,* and traditional authority (TA) *k*. $T^{v}_{v(i)}$ is an indicator for village *v* being treated and, because of the universal nature of the program, *i* being treated. $T^{GVH}_{c(i)}$ is an indicator for GVH *c* being treated, i.e. ⅔ of villages within *c* being assigned to treatment. $HighSat_{k(i)}$ is a set of indicator values for different saturation levels at the TA level (low-saturation 40%, medium-saturation 60%, and high-saturation 80%). We include the baseline value of the outcome variable (y^{BL}_{ivck}) when available to improve statistical precision, set to the mean if missing and an indicator for missingness (M^{BL}_{ivck}).[11] α_p are phase fixed effects (Phase I and Phase II). We weight observations by the inverse sampling probabilities at the village level to be representative of the population of eligible households. When reporting direct and within-village effects only (β), we cluster standard errors at the village level. When reporting spillover effects (γ) or total effects involving both direct and spillover effects, we cluster standard errors at the GVH level to reflect the level of randomisation.

[10] A key difference is that cash transfers in Chiradzulu will be universal. That is, there will not be any ineligible households. We therefore do not separately estimate within-village spillovers for ineligible households, and across village spillovers for eligible and ineligible households separately.

[11] In some cases where outcomes are measured differently at baseline between Phase I and Phase II (e.g. because the survey instrument was refined, or because we did not collect this outcome in one phase), we will interact the baseline controls with indicators for Phase I and II respectively).

Our primary parameter of interest is β which captures the direct effect of treatment on recipients as well as any potential spillovers on treated households within treatment villages. Importantly, since treatment at the village level is universal, village and individual treatment are collinear, and we cannot separately identify the direct effects of cash on recipients from within-village spillovers to other households. As such, we interpret β as the average direct + within-village spillover effect of a universal cash transfer program.[12] γ measures additional across-village spillovers on both recipients and non-recipients within Group Village Headman (GVH) units, typically comprising approximately 8 - 10 villages, (but not across). Since GVH boundaries in our context are relatively unimportant for economic activity, with substantial interactions across borders, and no meaningful social or ethnic divides, we do not consider γ a primary parameter of interest. Instead, our main spillover measure will be captured in equation (2).

δ captures the impact of saturation within a traditional authority (TA), typically comprising approximately 10 GVHs and 80 - 100 villages. However, there are only 10 TAs in Chiradzulu, and we therefore are likely underpowered to detect TA-level spillovers with one district only. We therefore do not consider δ a primary parameter of interest. GD has plans to expand their program in the future, and we will consider combining Chiradzulu with future districts to estimate and conduct more powerful tests for TA-level spillovers. Nonetheless, the variation in TA saturation will create additional spatial variation in treatment exposure, which we consider in our spatial estimation approach below.

4.1.1 Testing for differential spillovers for recipients and non-recipients

In equation (1a), γ and δ measure *average* spillovers on both recipients (over-and-above the direct and within-village spillovers captured by β), and on non-recipients (those in control villages). It is possible that these across-village spillovers are different in magnitude between recipient and non-recipient households (e.g. because receiving transfers yourself, and neighbors receiving transfers are complements or substitutes). We test this by interacting spillover terms with recipient status (village-level treatment):

$$\begin{aligned} y_{i(v,c,k,p)} &= \alpha_p + \beta T^v_{v(i)} + \gamma_1 T^{GVH}_{c(i)} + \gamma_2 T^v_{v(i)} \cdot T^{GVH}_{c(i)} + \delta_1 HighSat_{k(i)} + \delta_2 T^v_{v(i)} \cdot HighSat_{k(i)} \\ &+ \omega_1 y^{BL}_{i(v,c,k,p)} + \omega_2 M^{BL}_{i(v,c,k,p)} + \varepsilon_{i(v,c,k,p)} \qquad (1b) \end{aligned}$$

γ_2 and δ_2 test for differential spillovers to recipients relative to non-recipients. If they are significantly different from zero at the 10% significance level, and economically meaningfully different, we may use equation (1b) as our primary specification for estimating impacts on recipients and non-recipients.

4.1.2 Effects of interest

The effects of interest (estimands) from this reduced-form specification are:

[12] We note another limitation. Due to the stratified randomization, conditionally on being treated, the probability of any neighbouring village within the same GVH being treated is slightly lower (even conditioning on average GVH-level treatment). Because of this, □ also captures a small decline in within-GVH spillovers, in the context where there are any. This is likely to be small relative to the direct and within-village spillovers in a setting where the average GVH has 8 - 10 villages. Our primary specification for capturing all spillover effects correctly is below.

1. **Primary: Direct Treatment Effect (recipients):** $\hat{\beta}$
   This captures the direct + within-village spillover effect of cash transfers on recipients (excluding any across-village spillovers). We consider this estimand primary.

2. **Secondary: Total Effect in Treated GVHs (recipients):** $\hat{\beta} + \hat{\gamma}$
   This captures the average total effect on recipients in treated GVHs, including direct + within village spillovers and within-GVH spillovers. (If we use equation (1b) as our primary specification, this is $\hat{\beta} + \hat{\gamma}_1 + \hat{\gamma}_2$ instead.). Note that this excludes across-GVH spillovers ($\hat{\delta}$), since we are unlikely to estimate this with precision given that there are only 10 TAs. If it turns out that δ is economically meaningful and estimated precisely enough, we may decide to include it. For some purposes, we may instead report the average effect on recipient households across the entire study area (rather than the effect in treated GVHs): $\hat{\beta} + \hat{\gamma}E(T^{GVH}_{c(i)}|T^{v}_{v(i)} = 1)$**.** We consider this estimand secondary, and rely on the spatial specification to estimate the primary total effects on recipients and non-recipients.

3. **Secondary: Total Effect in Treated GVHs (non-recipients):** γ
   This captures the average total effect on non-recipients in treated GVHs, including direct + within village spillovers and within-GVH spillovers. (If we use equation (1b) as our primary specification, this is γ_1 instead.). Note that this excludes across-GVH spillovers ($\hat{\delta}$), since we are unlikely to estimate this with precision given that there are only 10 TAs. If it turns out that δ is economically meaningful and estimated precisely enough, we may decide to include it. For some purposes, we may instead report the average effect on non-recipient households across the entire study area (rather than the effect in treated GVHs): $\hat{\beta} + \hat{\gamma}E(T^{GVH}_{c(i)}|T^{v}_{v(i)} = 0)$**.** We consider this estimand secondary, and rely on the spatial specification to estimate the primary total effects on recipients and non-recipients.

4.2 Spatial Specifications

Reduced form specifications capture spillovers to the extent that they are contained within administrative village and GVH boundaries. To the extent that economic and social interactions cross these boundaries, our estimates will therefore understate the full extent of spillovers. To capture spillovers that vary smoothly with distance, we model each household's exposure to the program as the total transfer intensity (money per adult) in donut-shaped distance bands around the centroid of the household's village, excluding the focal village. We estimate the following specification:

$$y_{i(v,c,k,p)} = \alpha_p + \beta\, Amt_{v(i)} + \sum_{r=d}^{\bar{R}} \gamma_r Amt^{r-d\ to\ r\ km}_{\neg v(i)} + \sum_{r=d}^{\bar{R}} \phi_r s^{elig,r-d\ to\ r\ km}_{\neg v(i)} + \omega_1 y^{BL}_{i(v,c,k,p)}$$

$$\omega_2 M^{BL}_{i(v,c,k,p)} + \varepsilon_{i(v,c,k,p)} \qquad (2a)$$

where $Amt_{v(i)}$ is the amount of money (per adult resident in the village) transferred to households in village *v*, $Amt^{r-d\ to\ r\ km}_{\neg v(i)}$ is the amount of money (per adult resident in the buffer) transferred to households in villages other than *v* within radii bands of *r* - *d* to *r* km around the centroid of village v, and $s^{elig,r-d\ to\ r\ km}_{\neg v(i)}$ is the share of adults within *r* - *d* to *r* km that are eligible for transfers. Since cash transfers are universal, and all adults are eligible, $Amt_{v(i)}$ is perfectly predicted by treatment assignment of the village, and thus exogenous. $Amt^{r-d\ to\ r\ km}_{\neg v(i)}$ is exogenous as well, conditional on $s^{elig,r-d\ to\ r\ km}_{\neg v(i)}$. This term adjusts for radii crossing study borders - no household outside of Chiradzulu is eligible for the transfer. In this universal implementation of the program across all villages in Chiradzulu, $s^{elig,r-d\ to\ r\ km}_{\neg v(i)}$ is simply the share of adults within each buffer that resides within the district.[13]

We use Conley (1998, 2008) standard errors for spatial regressions, using a positive definite kernel: $K_{ij} = 1(d_{ij} <= D) \cdot (1 - d_{ij}/10)^2$, where $1(.)$ is the indicator function and d_{ij} is the distance in km between observations *i* and *j*. D is the smallest distance r such that if $d_{ij} > r$ then *j* does not belong to a village within distance κ of *i*, where κ is the maximum radii band that we select for the regression.[14] We will also attempt to follow recent and ongoing advances in this literature if these become the norm, including e.g. Muller & Watson (2023) or DellaVigna et al. (2025).

4.2.1 Relation to the the Kenya General Equilibrium Study (KGES)

The amount transferred to each village, and within each radii band is potentially endogenous. In the KGES, it was endogenous both because of the share of villages within each band that is within the study area (those outside are ineligible to receive money). Additionally – since transfers were targeted to households with an organic roof – the amount transferred was endogenous also because of the share of households with organic roofs in each village/buffer.

Egger et al. (2022) addressed this by running Equation (2a) using bands of 2km width, and up to 20km. Their specification did not control for $s^{elig,r-2\ to\ r\ km}_{\neg v(i)}$, but used the share of *eligibles* that were treated within each buffer ($s^{eligtreat,\ r-2\ to\ r\ km}_{\neg v(i)}$) as an instrument for $Amt^{r-2\ to\ r\ km}_{\neg v(i)}$. While this overcomes the endogeneity of treatment intensity, it is not the most efficient instrument under the assumption of homogenous treatment effects. Walker et al.

[13] For the number of adults within the district, we have our own census of all villages to draw on. To estimate the adult population in a buffer outside the district, we will use other datasets. Currently, we are pursuing getting micro-data for a) the 2018 Malawi census, and b) UBR, Malawi's poverty registry, an almost universal household survey that was collected in 2019. We prefer using the 2018 census, but will use whichever dataset will be available at a finer geographic variation. (We have UBR GPS data at the household level for Chiradzulu district, and may be able to get it for adjacent districts. We do not currently know whether census data is available at the household level.). Using our preferred dataset, we will then estimate the share of adults/households/population in each buffer that resides inside the district by using the share of all adults/households/population in the buffer captured in census/UBR data that is outside the district. To get total populations of adults or households outside the district, we scale up our own inside-district-census numbers by these outside-district-shares obtained from the census/UBR.

[14] For comparability with earlier work, we will also assess the robustness of our inference to selecting a maximum radius for Conley standard errors to be 5km and 10km respectively.

(2025) – in a PAP amendment (Egger et al. 2024) – therefore refined this specification by adding $s_{\neg v(i)}^{elig,r-2\ to\ r\ km}$ controls as in Equation (2a), and instrumenting $Amt_{\neg v(i)}^{r-2\ to\ r\ km}$ using both the share of *eligible* households that were treated ($s_{\neg v(i)}^{eligtreat,\,r-2\ to\ r\ km}$), as well as its interaction with the share of households that is eligible ($s_{\neg v(i)}^{eligtreat,\,r-2\ to\ r\ km} \cdot s_{\neg v(i)}^{elig,r-2\ to\ r\ km}$). Under homogenous treatment effects, this is the efficient estimator. The first-stage of this 'interacted' instrument should perfectly predict the amount treated per household/adult, and first-stage results in Walker et al. (2025) confirm this.

We therefore opt for the OLS version without instrumenting for $Amt_{\neg v(i)}^{r-2\ to\ r\ km}$, using essentially the same specification as Miguel & Kremer (2004). This is consistent, and most efficient (under the assumption of linear homogenous treatment effects), and circumvents certain limitations or reductions in flexibility resulting from going beyond OLS.

In a more recent generalization, Borusyak & Hull (2023) advocate for re-centering each regressor by its expected amount (or controlling for it) across many iterations of the experimental design for situations such as ours where part of the variation in RHS variables is exogenous, while part is not. In our case, the exogenous variation is from the RCT, the endogenous part is from some adults being eligible (living within Chiradzulu) and others not. Since the program is universal, the expected amount received by each eligible person in each village is exactly $p * \bar{Amt}$ for all villages (a constant), and the expected amount going into the buffer around each village *v* is exactly $p * s_{\neg v(i)}^{elig,r-d\ to\ \bar{r}\ km} * \bar{Amt}$ where *p* is the average treatment probability of a village, and $\bar{Amt}$ is the average amount transferred per eligible adult. So our specification is identical to their approach, as it exactly controls for the expected value of RHS regressors.

4.2.2 Bandwidth selection

The literature on estimating standard errors in the presence of spillover effects is evolving rapidly (see e.g. Faridani and Niehaus, 2024). This section reflects our thinking at the time of publication, but given the evolving literature, it should not prevent us from updating our approach as the frontier evolves.

We determine $\bar{R}$ using the same general approach as Egger et al. (2022), namely running a series of nested regressions increasing $\bar{R}$, and minimizing the Schwartz BIC over a restricted set of models with at least one spillover term. We enforce this constraint of at least one spillover term because understanding spillovers is a key objective of our analysis. However, we make several refinements to the bandwidth selection algorithm relative to Egger et al. (2022).

Egger et al. (2022) considered 2km radii bands up to 20km, which were pre-specified. Their algorithm selected only 2km (and, very rarely, 4km radii bands), suggesting that including additional bands did not substantially improve the in-sample goodness-of-fit. They showed that the choice of optimal bandwidth was stable across splits. In Faridani & Niehaus' (2024) re-analysis using their new GATE methodology (see more details below), the optimal bandwidth is even lower, at 400m for the effects on consumption of eligible households. For context, in our Malawi setting there are [0.51, 3.07, 12.42, 62.12] villages within [500m, 1km, 2km, 5km] of one another.

Relative to the Egger et al. (2022) procedure, we will make two core adjustments. First, we will define a finer grid of possible bandwidths, including smaller bands over which to search to better match these empirical findings. Second, conditional on having adequate sample to make comparisons, we intend to make use of sample splitting, or post-selection inference adjustments to avoid bias arising due to the "Winner's Curse" (Andrews, 2024), but may report results without sample splitting as well for comparability.

In addition to the two core adjustments, we also make two technical refinements – which were helpfully pointed out to us by a re-analysis of Egger et al. (2022) by David Roodman (unpublished). First, in computing the BIC, we will do so on the OLS regressions (rather than on the IV) as OLS is our primary specification and we do not instrument for amounts. Second, we set N to be the number of clusters – villages in our case – rather than observations, as a pragmatic measure of effective sample size.

This literature is rapidly evolving with new methods to relax the trade-off between minimizing bias (50-50 sample splits) and maximizing sample for lower variance model selections and post-selection inference. The intended approach below therefore reflects our thinking at the time of publication, but may be subject to revisions:

a. We define the grid: [500m, 1km, 2km, 3km, 4km, 5km]. A minimum radius of 500m is the lowest we can reliably estimate given the density of villages in our sample. A maximum radius of 5km seems an a priori reasonable compromise for comparability with earlier work, and conservative enough in case spillovers in Malawi range further than in Kenya.

b. For each outcome family (as defined in section 5) we perform a set of nested regressions sequentially increasing the set of included bandwidths. Among these nested regressions, we select the model that minimizes BIC. In the split sample framework, this would occur in the training data, with the held out data used for inference on this model.

c. Consolidating radii bands: If the procedure selects multiple radii bands to be included, we compare the BIC between specifications including all radii bands separately, and one that has only one buffer up to $\bar{R}$. If the BIC is lower for the one-ring specification, that is our primary specification. However, we will report coefficients on all radii bands separately, and for final results show robustness to varying $\bar{R}$. We hope that this will improve overall power by limiting the number of regressors, while still enabling us to investigate the spatial decay of spillovers. We will also generally use the one-radii band specification for additional interactions, or dynamic versions of the equations below. As above, we would do this in the split sample by selecting radii in the training data with held out data used for inference on the selected model.

d. We generally estimate the bandwidth separately for each outcome as the spatial decay of spillovers may vary. However, for comparability across results, to limit researcher degrees of freedom, and to ensure accounting identities add up, we impose the same $\bar{R}$ for subcomponents of other outcomes within an outcome group – e.g. we select the radius for total consumption, and impose that same radius for all subcomponents of consumption. Similarly, we compute the $\bar{R}$ for business revenue, and then impose the same for profits, costs, etc. When considering heterogeneous spillovers, or heterogeneous treatment effects across subgroups, we also restrict the radius to be the same for comparability, so that tests for heterogeneity are conducted within the same regression specification.

As a secondary specification, we use the minimax bandwidth selection procedure proposed by Faridani & Niehaus (2024). The candidate bandwidths we will consider are: [500m, 1km, 2km, 3km, 4km, 5km]. We will use their default matrix G – at the time of writing this, G was the identity matrix, but we will update this to incorporate any future revisions to their procedure.

4.2.3 Testing for linearity

Several applications (including the extrapolation from regressions to average total effects, or global average treatment effects) rely on an assumption of linearity. We test for linearity of the effect for our main primary outcomes by replacing $Amt_{\neg v(i)}^{0-R\,km}$ with five quintile bins of this variable in equation (2a), using a specification with only a single buffer up to R:

$$y_{i(v,c,k,p)} = \alpha_p + \beta\, Amt_{v(i)} + \sum_{q=1}^{5} \gamma_q\, 1(Amt_{\neg v(i)}^{0-R\,km} \in Q_q) + \phi_r s_{\neg v(i)}^{elig,\,0-R\,km} + \ldots + \varepsilon_{i(v,c,k,p)}$$

where Q_q is the q-th quintile of $Amt_{\neg v(i)}^{0-R\,km}$ and 1(.) is the indicator function. We then test for linearity using an F-test of $\gamma_1/\bar{Q}_1 = \ldots = \gamma_5/\bar{Q}_5$ where $\bar{Q}_q$ is the mean treatment exposure for households in the q-th quintile ($= E(Amt_{\neg v(i)}^{0-R\,km} \mid 1(Amt_{\neg v(i)}^{0-R\,km} \in Q_q)$). If we reject linearity at conventional significance levels, we may test for robustness of the main results to include the binned specification (instead of the linear specification).

4.2.4 Testing for differential spillovers for recipients and non-recipients

Analogous to equations (1a) and (2a), we will test whether across-village spillovers are different in magnitude between recipient and non-recipient households by interacting spillover terms with recipient status (village-level treatment):

$$y_{i(v,c,k,p)} = \alpha_p + \beta_1 Amt_{v(i)} + \sum_{r=d}^{\bar{R}} \gamma_{1,r} Amt_{\neg v(i)}^{r-d\ to\ r\ km} + \sum_{r=d}^{\bar{R}} \gamma_{2,r} T_{v(i)}^{v} \cdot Amt_{\neg v(i)}^{r-d\ to\ r\ km}$$
$$+ \sum_{r=d}^{\bar{R}} \phi_r s_{\neg v(i)}^{elig, r-d\ to\ r\ km} + \omega_1 y_{i(v,c,k,p)}^{BL} + \omega_2 M_{i(v,c,k,p)}^{BL} + \varepsilon_{i(v,c,k,p)} \qquad (2b)$$

As above, if we find spillovers to be significantly different for recipients and non-recipients at the 10% significance level, and economically meaningfully different (at least 20% different from spillovers on non-recipients), we may use equation (2b) as our primary specification for estimating impacts on recipients and non-recipients.

4.2.5 “Gravity” specification

Our primary specification for choosing radii bands is agnostic to the economic meaning of ‘distance’. This is desirable for a primary specification covering all kinds of spillovers (economic, social, etc.) and all kinds of outcomes that may decay very differently over space – as long as geographic distance is reasonably correlated with these more 'structural' notions of distance. If we believe spillovers are primarily driven by demand and market interactions, and economic in nature (e.g. as in Egger et al. 2022), then flows of expenditure over space may be the right model of exposure. This is particularly true for economic outcomes, such as income, enterprise revenue, prices, etc. For example, a wide class of spatial models feature gravity (e.g. Donaldson and Hornbeck, 2016 or Ahlfeldt et al, 2015) where market access terms (or weighted trade flows between locations) act as sufficient statistics for spatial linkages. We will therefore consider an alternative spatial specification that weights locations based on their spending-/trade-flow weighted exposure:

$$y_{i(v,c,k,p)} = \alpha_p + \beta\, Amt_{v(i)} + \gamma \sum_{r=1}^{5} s_r Amt_{\neg v(i)}^{r-1\ to\ r\ km} + \phi \sum_{r=1}^{5} s_r s_{\neg v(i)}^{elig,r-1\ to\ r\ km} + \ldots + \varepsilon_{i(v,c,k,p)} \quad (2c)$$

where s_r is the share of expenditure of households within $r - 1$ to r km. We obtain this from the consumption module of households at baseline, which asks, for each spending category, the amount spent as well as the location at which this was purchased.[15] For primary outcomes, we use all categories of spending, but when looking at sub-outcomes (e.g. revenue of non-food enterprises) we may consider exposure shares only using sub-components of spending (e.g. spending on non-food products).

An alternative approach to using average spending shares within geographic buffers might have been to use *actual* spending shares, e.g. from each village *v* to each other village *v'* or market *m* to define treatment exposure. However, Walker et al. (2025) show that using baseline expenditure shares is preferable to the purely geographic approach only when expenditure switching is limited, i.e. when the elasticity of substitution across markets is low. Thus, using the gravity approach of (2c) has statistical advantages because a) bilateral spending shares between villages and markets rely on sparse data at baseline for only a few households, and b) being somewhat robust to such substitution as long as much of it remains within the distance-gravity patterns. In exploratory analyses, we may additionally use a specification including bilateral exposure shares from villages to villages / markets.

4.2.6 Effects of interest

The effects of interest (estimands) from spatial specifications are:

1. **Primary: Average Total Effect (AtotE) on recipients**
   This captures the average total effect of the program – at the saturation level that it was implemented – on recipient households, including direct effects, within-village spillover effects, and across-village spillover effects using the average level of saturation around recipient villages:

$$ATotE^{recipients} = E(Y_i^1(T) - Y_i^0(0) \mid i \in T)$$

   where $Y_i^1(T)$ is the potential outcome for *i* if *i* is treated (Y_i^0 if untreated), both of which may depend on T, the set of individuals treated in the actual experiment. We estimate

[15] Note: We use all spending captured both at markets and villages, and set own-village spending distance to 0.

this linearly using:

$$\widehat{ATotE}^{recipients} = \hat{\beta}\, E(Amt_{v(i)} | T^{v}_{v(i)} = 1) + \sum_{r=d}^{\bar{R}} \hat{\gamma}_r E(Amt^{r-d\ to\ r\ km}_{\neg v(i)} | T^{v}_{v(i)} = 1)$$

which is analogous to the total treatment effects reported in Egger et al. (2022). If we use equation (2b) as our primary specification, this is instead:

$$\widehat{ATotE}^{recipients} = \hat{\beta}\, E(Amt_{v(i)} | T^{v}_{v(i)} = 1) + \sum_{r=d}^{\bar{R}} (\hat{\gamma}_{1,r} + \hat{\gamma}_{2,r}) E(Amt^{r-d\ to\ r\ km}_{\neg v(i)} | T^{v}_{v(i)} = 1)$$

2. **Primary: Average Total Effect (AtotE) on non-recipients**
This captures the average total effect of the program – at the saturation level that it was implemented – on non-recipient households, including direct effects, within-village spillover effects, and across-village spillover effects using the average level of saturation around recipient villages:

$$ATotE^{non-recipients} = E(Y^{1}_{i}(T) - Y^{0}_{i}(0) | i \notin T)$$

Analogous to the effect on non-recipients, we estimate this linearly as:

$$\widehat{ATotE}^{non-recipients} = \sum_{r=d}^{\bar{R}} \hat{\gamma}_r E(Amt^{r-d\ to\ r\ km}_{\neg v(i)} | T^{v}_{v(i)} = 1)$$

which is also analogous to what was reported in Egger et al. (2022). If we use equation (2b) as our primary specification, this is $\sum_{r=d}^{\bar{R}} \hat{\gamma}_{1,r} E(Amt^{r-d\ to\ r\ km}_{\neg v(i)} | T^{v}_{v(i)} = 1)$ instead.

3. **Primary: Global Average Treatment Effect (GATE):**
This is the global average treatment effect as defined in Faridani & Niehaus (2024).[16] The average effect across the entire study-area population if the program was universally scaled to everyone in Chiradzulu district.

$$GATE = E(Y^{1}_{i}(1) - Y^{0}_{i}(0))$$

For GiveDirectly who are considering to scale their program nationwide – and for governments and actors in many other settings – this may be the policy-relevant estimand. We follow Faridani & Niehaus (2024) in estimating the GATE using the following linear estimator:

$$y_{i(v,c,k,p)} = \alpha_p + \beta T^{v}_{v_{(i)}} + \sum_{r=d}^{\bar{R}} \gamma_r streat^{r-d\ to\ r\ km}_{\neg v(i)} + \omega_1 y^{BL}_{i(v,c,k,p)} + \omega_2 M^{BL}_{i(v,c,k,p)} + \varepsilon_{i(v,c,k,p)} \quad (2d)$$

where $T^{v}_{v_{(i)}}$ is the re-centered (by the average treatment probability *p* across all villages across the study area) treatment status of village *v*. $streat^{r-d\ to\ r\ km}_{\neg v(i)}$ is the re-centered share of villages within the study area (with any residents living within the $r - d$ to r km

[16] We note that at the time of writing, Faridani & Niehaus (2024) was under revision. We will follow any later updates to their paper.

radius) that are assigned to treatment, which is equal to $n_{treat}^{r-d\ to\ r\ km} / \bar{n}^{r-d\ to\ r\ km}$ where $n_{treat}^{r-d\ to\ r\ km}$ is the number of treated villages within the $r - d$ to r km radius and $\bar{n}^{r-d\ to\ r\ km}$ is the average number of villages (in Chiradzulu) within $r - d$ to r km across all villages in the study. We recenter $streat_{\neg v(i)}^{r-d\ to\ r\ km}$ by $p * n_{treat}^{r-d\ to\ r\ km} / \bar{n}^{r-d\ to\ r\ km}$. For comparison with our primary specification, we will use the same Conley (2008) standard errors for this specification as our primary method for inference.[17]

The estimator for the GATE is then defined as:

$$\widehat{GATE}^{FN} = \hat{\beta} + \sum_{r=d}^{\bar{R}} \hat{\gamma}_r$$

As in Faridani & Niehaus (2024), we will compare this also to a GATE estimator based on the primary specification (2a/b):

$$\widehat{GATE} = \hat{\beta}\, E(Amt_{v(i)} | T_{v(i)}^{v} = 1) + \sum_{r=d}^{\bar{R}} \hat{\gamma}_r E(Amt_{v(i)} | T_{v(i)}^{v} = 1)$$

where $E(Amt_{v(i)} | T_{v(i)}^{v} = 1)$ is the average amount that was transferred to each adult participating in the program, thus simulating – by linear extrapolation – a scenario where every adult within each buffer would have gotten the average amount transferred to recipients in the actual experiment.

We note that the above estimators and – in particular, the AtotE and GATE – rely on the assumptions on spillovers outlined in Faridani & Niehaus (2024), namely a notion of the speed of spatial decay of these spillovers at increasing distance relative to the randomization design, as well as on linear extrapolation. For many outcomes, this is *a priori* reasonable. However, true 'missing intercepts' (e.g. as in Moll 2025 or Walker et al. 2025) or 'ambient effect' (as discussed in Egger et al. 2022) which affect *all* units in our study area equally, and do not decay over space, as well as severe deviations from linearity, may mean that our estimators may be biased. Where possible – i.e. where we have data from both pre- and post-implementation of cash transfers – we may investigate this possibility empirically by comparing the entire treated area against a matched 'far-away' control group (e.g. adjacent districts, or Phase I vs. as of yet untreated Phase I villages and markets). Such effects can also be recovered from more structural approaches (e.g. as in Walker et al. 2025) by relying on the functional form of structural models.

We may use the estimated coefficients from Equations (1a/b and 2a/b/c) to calculate additional estimators of interest, and simulate different saturations of programs. We may also explore heterogeneous effects by individual or local area characteristics as specified below.

[17] Depending on feasibility and an upcoming revision of their paper, we may modify the Faridani & Niehaus (2024) standard errors to accommodate our slightly modified own- vs. other village treatment specification, adjusted for our GVH-level clustered randomisation as a robustness check for comparison.

Moreover, structural approaches to modeling spillovers may yield different estimands or sufficient statistics beyond those proposed above (e.g. similar to 4.2.5).

4.2.7 Structural Estimates

We may attempt to combine the empirical estimates outlined in this PAP, and the data collected, to estimate, test, and discipline structural models. In a collaboration with STEG, we co-fund 6 teams of macro-development economists who will write such models *ex-ante*, using baseline data only, and then test these models against our Midline and Endline 1 data ex-post (see Appendix F for more details). These 'Structural Pre-Analysis Plans' will be registered under a separate AEA Trial Registry 18604.

4.3 Covariate adjustment

In each of our specifications, we will present results from augmented models, in which we use data-driven methods to select additional covariates from the available set of (flexible functions of) pre-treatment covariates. The purpose of this adjustment is to improve precision and power by flexibly partialling out predictable variation in outcomes, while preserving valid inference on the treatment and spillover parameters by avoiding ad hoc, outcome-driven covariate selection.

We do this primarily following Chernozhukov et al. (2018), where we will employ Doubly-Debiased Machine Learning (DDML) to conduct covariate adjustment using flexible machine-learning estimators for the nuisance components (e.g., conditional outcome and exposure models), while constructing an orthogonal (Neyman-orthogonal) score for the target treatment/spillover parameters. We implement cross-fitting (sample splitting) so that nuisance estimates are formed out-of-fold, mitigating regularization/overfitting bias and enabling standard asymptotic inference under high-dimensional and nonparametric nuisance estimation.

An emerging literature emphasizes that spatial dependence can affect both inference and the resampling/splitting steps used by ML-based estimators (e.g. Cao & Leung, 2025). In addition to reporting Conley standard errors, we will explore modern methods to adapt DDML cross-fitting to account for spatial dependence. This might include the use of block cross-fitting.

4.4 Fisher exact test of treatment effects

In addition to the large-sample approach outlined in Sections 4.1 and 4.2, we perform Monte Carlo approximations of exact tests of the treatment effect (Fisher 1935) as in Egger et al. (2022). Randomization inference allows us to test the Fisherian sharp null hypothesis that the treatment effect on each individual *i* is exactly 0. To remain analogous with our main specification, we use the conventional Wald statistics from Equations (1a/b) and (2a/b/c) corresponding to the estimands of interest as our test statistics (as in Young (2022)). We calculate exact p-values for each estimand under the null hypothesis using a Fisher permutation test, taking 2,000 permutations of treatment assignment and roll-out as it was implemented in the actual intervention, i.e.:

1) We randomly assign TAs to treatment intensity status (keeping the two phases separate, and considering implementation constraints as in the real implementation)

2) We randomly assign GVHs to treatment status
3) We randomly select treatment villages within treated GVH
4) We randomly assign treatment timing within each phase (I/II) to different monthly batches, keeping the overall speed of enrolments by GD across months as given.

We then calculate the exact p-value as the share of Wald Statistics for each main estimand that exceeds the one we find in our sample in absolute value.

4.5 Heterogeneous impacts

4.5.1 Unconditional cash vs. business grants

We test for the differential impact of unconditional cash transfers to all households and business grants to enterprise owners within the household by focusing on the sample of business-grant eligible households (all households within which a business-grant eligible business owner resides). Within that sub-sample, we interact recipient status of unconditional cash and business grants, across all specifications above. For instance, in the reduced-form specification, we estimate:

$$\begin{aligned} y_{i(v,c,k,p)} &= \alpha_p + \beta_1 T^{v}_{v(i)} + \beta_2 T^{supply-side-grant}_{i} + \beta_3 T^{v}_{v(i)} * T^{supply-side-grant}_{i} \\ &+ \gamma T^{GVH}_{c(i)} + \delta HighSat_{k(i)} + \omega_1 y^{BL}_{i(v,c,k,p)} + \omega_2 M^{BL}_{i(v,c,k,p)} + \varepsilon_{i(v,c,k,p)} \qquad (1c) \end{aligned}$$

where $T^{supply-side-grant}_{i}$ denotes receipt of business grants to an eligible business owner within the household, and $T^{v}_{v(i)}$ denotes the household living in a treated village. β_1 represents the impact of household grants alone, β_2 the impact of supply-side transfers alone, and $\beta_1 + \beta_2 + \beta_3$ the impact of both. We will estimate these for overall household impacts, but also specifically look at differential impacts within the household on the business owner themselves, relative to other household members.

There may also be differential spillovers from household vs. business grants. The saturation of business grants varies randomly across Phase II markets, and creates systematic exogenous variation in business grants across geographies. However, business grants represent a very small share of the total amount of money transferred (1,650 recipients vs. 104,000 recipients for household cash) and we may not be powered to detect broader household-level spillovers separately from overall spillovers of cash. In the local economy PAP, we discuss estimating market-level spillovers to other enterprises. We may also explore estimating such differential spillovers to household level outcomes in secondary analyses.

4.5.2 Heterogeneity of impacts of unconditional cash and/or business grants

We conduct several analyses to study heterogeneous treatment effects. We do so in two different ways. First, we pre-specify dimensions of heterogeneity that we hypothesize *a priori* are likely to be important modifiers or are hereto understudied. Second, we implement machine learning–based methods that use the full vector of pre-treatment characteristics to characterize heterogeneity in a principled, algorithmic manner, while maintaining out-of-sample discipline.

We are interested in heterogeneity:

a) by characteristics of the ‘receiving’ household, for both direct and spillover effects (e.g. which households/individuals benefit more/less from the overall program)
b) by characteristics of the ‘emitting’ households of spillovers (e.g. which households/individuals create the largest spillovers to other households), and
c) by characteristics of the overall economy within which these direct and spillover effects and interactions take place.

Since people’s residence is endogenous (and thus their characteristics are correlated with those of their neighbors), it is not possible to cleanly separate these channels. In this PAP, we focus on a). b) amounts to estimating heterogeneity in peer-effects. This is challenging in our context. ‘Peer groups’ here are all individuals who interact in a sense which could generate spillovers (e.g. all participants in the same market). Since these groups may be large (and we did not explicitly vary the targeting characteristics of the program across regions), the law of large numbers implies the variation in their characteristics will be small, and we may be underpowered to estimate the heterogeneity of spillovers by these characteristics (similar to a point made in Angrist 2014).[18] However, our findings on a) together with some structure may already provide some insights into b) and c), e.g. as the households spending their transfer most locally might have the largest positive impact on local firms, etc.[19]

We generally use Baseline survey data to define dimensions of heterogeneity (this avoids a ‘bad control’ problem). However, we note that the Midline and Endline 1 sample size is planned to be somewhat larger than the Baseline sample. Where possible, we therefore use Baseline census data to define dimensions of heterogeneity (for those not surveyed at Baseline). For variables not available in the Baseline census, we may use census data to predict Baseline values – e.g. we will use the subset of assets we asked about in the census to predict Baseline consumption for households not surveyed at Baseline. In pooled analyses combining survey- and census-based measures, we allow for separate intercepts by data source (survey and census respectively) due to different measurement, and define baseline dummies stratified by data source.

4.5.1 A priori dimensions of heterogeneity

Our *a priori* pre-specified dimensions of heterogeneity are:

For individual-level outcomes:

Demographics

1. An indicator for being the household head
2. An indicator for female

[18] To illustrate: To test whether richer or poorer households generate larger or smaller spillovers to their neighbors, we need variation – among neighbors within a radii band – in the share of rich vs. poor households that were treated. As the radii get larger, that share will approach the average share treated asymptotically.

[19] We will test one specific theory of a type of person that may generate particularly large spillovers – namely credit-constrained business owners. In our complementary supply-side interventions, we test whether targeting such business owners may generate larger spillovers in terms of increasing the aggregate supply response, and dampening inflationary pressure.

3. An indicator for being 18-25 years of age at baseline
4. An indicator for being married/partnered at baseline
5. An indicator for having children at baseline
6. An indicator for being below-median age-adjusted years of education at baseline (we calculate the median education within each 5-year age-bin)

Socio-Economics

7. An indicator for below-median HH consumption at baseline
8. An indicator for below-median HH assets at baseline
9. An indicator for working primarily in agriculture at baseline
10. An indicator for wage employment being a significant source of income at baseline
11. An indicator for running a non-agricultural enterprise at baseline
12. An indicator for below-median labor supply at baseline
13. An indicator for screening positive for depression at baseline on the PHQ-8
14. An indicator for reporting above-median levels of intra-household conflict at baseline
15. Intended prioritization of cash expenditure on consumption, household durables, investment (agriculture and/or business), human capital investment
16. An indicator for having above median locus of control
17. An indicator for having an above median value of an inverse covariance-weighted average ("Anderson index") across 10-year income and educational aspirations

For household-level outcomes:

Demographics:

1. An indicator for female-headed households
2. An indicator for HH head below 25 years of age at baseline
3. An indicator for HH head being married/partnered at baseline
4. An indicator for HH having children at baseline
5. An indicator for having children below 5 years at baseline
6. An indicator for the HH head being below-median age-adjusted years of education at Baseline(we calculate the median education within each 5-year age-bin)

Socio-Economics

7. An indicator for below-median consumption at baseline
8. An indicator for below-median assets at baseline
9. An indicator for below-median labor supply at baseline
10. An indicator for wage employment being a significant source of income at baseline
11. An indicator for running a non-agricultural enterprise at baseline
12. An indicator for screening positive for depression at baseline on the PHQ-2 (score exceeding 2)
13. An indicator for reporting above-median levels of intra-household conflict at baseline
14. Intended prioritization of cash expenditure on consumption, household durables, investment (agriculture and/or business), human capital investment
15. An indicator for having above median locus of control
16. An indicator for having an above median value of an inverse covariance-weighted average ("Anderson index") across 10-year income and educational aspirations

Heterogeneity based on economy-wide characteristics:[20]

1. Market access: indicator for above vs. below median market access, as defined by gravity-weighted population density as in Donaldson & Hornbeck (2016)
2. Remoteness: Travel time to major road
3. Market size: indicator for above-median village GDP per capita

We test for heterogeneity based on these *a priori* dimensions by interacting all regressors in our primary specifications with indicators for these characteristics (equivalent to running these separately for each subgroup in the case of categorical measures of heterogeneity). We then report our primary estimands of interest (except the GATE, which by its nature is an average across the entire population) for both groups, and use a Wald test for equality. We will report sharpened q-values (as above) to correct for multiple hypothesis tests across each group of dimensions of heterogeneity.

$$y_{i(v,c,k,p)} = \alpha_p + Z_i \cdot \alpha_p + \beta_1 T^{v}_{v(i)} + \beta_2 Z_i \cdot T^{v}_{v(i)} + \gamma_1 T^{GVH}_{c(i)} + \gamma_2 Z_i \cdot T^{GVH}_{c(i)} + \delta_1 HighSat_{k(i)} + \delta_2 Z_i \cdot HighSat_{k(i)} + \ldots + \varepsilon_{i(v,c,k,p)} \quad (1d)$$

$$y_{i(v,c,k,p)} = \alpha_p + \beta_1 Amt_{v(i)} + \beta_2 Z_i \cdot Amt_{v(i)} + \gamma_1 Amt^{0\ to\ \bar{R}\ km}_{\neg v(i)} + \gamma_2 Z_i \cdot Amt^{0\ to\ \bar{R}\ km}_{\neg v(i)} + \ldots + \varepsilon_{i(v,c,k,p)} \quad (2d)$$

where Z_i is a dimension of heterogeneity. Note that for spatial specifications, we opt for a one-buffer specification from distance *0* to $\bar{R}$ – fixed at the same maximum radius selected for the same outcome in the primary specification *(2a/b)* – rather than our multi-ring specification. This is to increase the power with which we may detect heterogeneous spillover effects. We may additionally explore a multi-ring specification to test for differential spatial decay rates of spillovers for households with different characteristics. If we find that the spatial decay for households with different characteristics is economically meaningfully and statistically significantly different, we may report the multi-ring specification instead.

In the 'local economy PAP', we will additionally study the heterogeneity of effects by characteristics of the economy, including market access (as in Egger et al. (2022)), spatial density, market structure, baseline capacity utilization (as in Walker et al. (2025)) etc.

4.5.2 Timing of transfer receipt

The timing of the roll-out was randomized, so we can look at treatment effects by the timing of when households received cash. We are particularly interested in how transfer timing relates to the agricultural calendar. The harvest season in Chiradzulu begins in April and lasts for several months. By September, the lean season begins. We therefore test for the heterogeneity of treatment effects by:

1. Phase I vs. Phase II

[20] We consider these tests of separate interest, and will conduct multiple hypothesis corrections for this group of heterogeneity dimensions separately from the other groups.

2. An indicator for receiving the large in the lean season (September to February)
3. A more flexible specification with monthly or quarterly dummies for timing of transfer receipt.

An important caveat is that the rollout of transfers over approximately a full calendar year means that indicators for phase, lean season, and other changes over the year are correlated. For instance, national elections took place in Malawi on 16 September 2025 and may affect those treated later / in Phase II differently.

Moreover, whether recipients received their transfer *early* relative to other households in the district is correlated with agricultural season and treatment timing. Those receiving transfers later may have different impacts due to e.g. different knowledge sets (they can learn from the experience of earlier treated neighbors), or due to differences in spillovers (e.g. because earlier-treated households may have already set up businesses, leading to more competitive pressure for later treated households). While we cannot disentangle this from seasonality across the district as a whole, we can look at this locally. Random roll-out timing generates some variation in whether neighbors were treated *before* or *after* recipients received their own transfer. We test for this as follows:

$$y_{i(v,c,k,p,m)} = \alpha_m + \beta_1 Amt_{v(i)} + \beta_2 Amt_{v(i)} * Amt^{0\ to\ \bar{R}\ km}_{before\ i,\neg v(i)} + \beta_3 Amt_{v(i)} * Amt^{0\ to\ \bar{R}\ km}_{after\ i,\neg v(i)} + \gamma^{before} Amt^{0\ to\ \bar{R}\ km}_{before\ i,\ \neg v(i)} + \gamma^{after} Amt^{0\ to\ \bar{R}\ km}_{before\ i,\ \neg v(i)} + \ ...\ + \varepsilon_{i(v,c,k,p,m)}$$

where α_m are month of treatment fixed effects for each individual/household (based on the experimental treatment timing assignment month), and $Amt^{0\ to\ \bar{R}\ km}_{before\ i,\ \neg v(i)}$ and $Amt^{0\ to\ \bar{R}\ km}_{after\ i,\ \neg v(i)}$ are pre-adult amounts transferred within each buffer *before* and *after* month *m.* β_2 vs. β_3 is a test for whether spillovers on recipients are different when neighbors are treated before vs. after they received their transfer before or after them. Analogously, it tests for whether the total effect – including the direct treatment effect – on recipients depends on the treatment timing of their neighbors. As above, we consider the one-buffer specification primary *a priori,* but may report the multi-buffer specification if spatial decay of spillovers from those treated before or after are economically meaningfully and statistically significantly different.

4.5.3 ML-based approaches to estimating heterogeneous treatment effects

We will also use principled, machine learning-based approaches to estimate heterogeneity in treatment effects using the full vector of pre-treatment variables in Endline 1 data. We outline our planned approach below, but note that methods are developing rapidly, and we will adopt new approaches if they become standard or offer other desirable properties in our setting.

First, leveraging randomized assignment, we will implement the Generic Machine Learning (GenericML) framework (Chernozhukov et al., 2019) to conduct inference the following features of the CATE: (i) the best linear predictor (BLP), (ii) Group Average Treatment Effects (GATES) across five groups sorted by predicted impact, (iii) CLAN comparisons of baseline characteristics between the most and least impacted groups, and (iv) summary statistics describing the most/least predicted-impact units. We will select the appropriate learners in the GeneriML framework by tuning and comparing performance using cross-fitting with an appropriate out-of-sample loss criterion. Given the relatively small sample (by machine learning standards), we anticipate that boosted forests will work well in this setting. We will then use repeated cross-fitting and aggregation as in GenericML.

Second, we will apply the generalized random forest framework of Athey et al. (2019) with a known propensity score in the RCT setting. This allows for direct estimation of the CATE, rather than inference on its features. We will again use this to explore predictors of relatively large treatment effects, and to explore tradeoffs between measures of poverty and predicted conditional average treatment effects, as in Haushofer et al. (2025). This allows us to identify interesting groupings of baseline characteristics that may be predictive of other (non-targeted) outcomes.

In addition to estimating heterogeneity of treatment effects based on baseline household characteristics, we note here two particular directions of enquiry that are of interest: First, we may explore the 'joint' distribution of outcomes, such as whether there are trade-offs between the different uses of cash (e.g. whether households that saved more are those that have longer-term consumption increases (see section 4.8 below), or whether households with the largest effect on consumption have larger effects on subjective well-being).

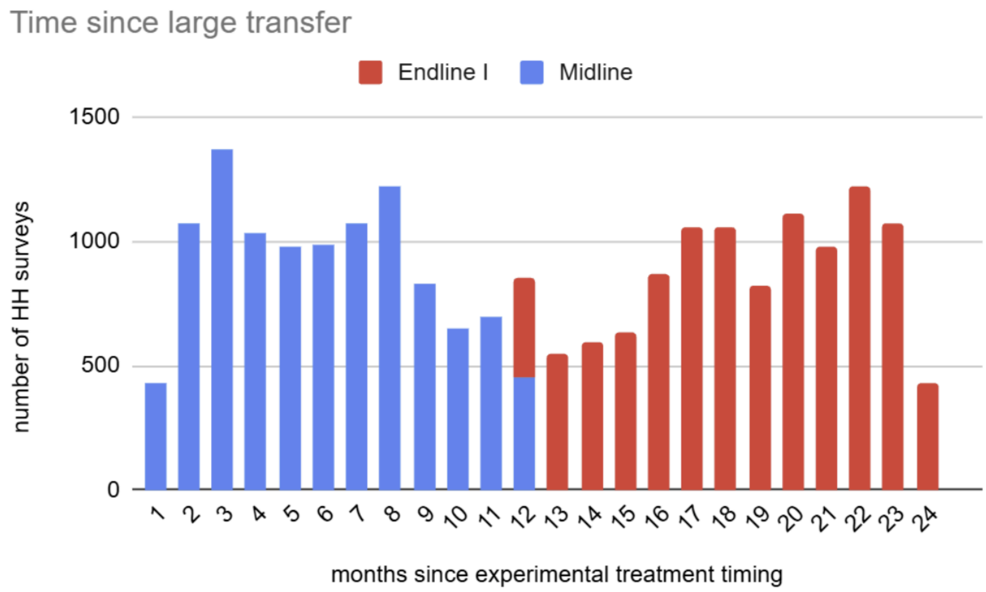
4.5.4 Exploratory analyses on transfer size

The fact that a) transfers are targeted at the individual level, b) that there were 14 randomly selected villages which receive a larger transfer amount (9 got 750 USD, 5 got 1,100 USD), and c) that a small subset of households will receive additional labelled business grants indicates there will be significant variation in transfer size. A regression discontinuity design around having household members just around the age cut off of 18, would lead to a significant discrete increase in *household* level transfers,but is likely to be underpowered. In more exploratory research, we will investigate these heterogeneities by transfer size at the household level, and test whether a linear model (where effects scale with transfer size) is appropriate or can be statistically rejected. To be optimally powered, we may combine both approaches and variation from a), b), and c).

4.6 Dynamic impacts

A key object of interest are dynamic impacts of cash transfers. The onset of cash transfers in each village, as well as survey timing across villages were randomized (see above), yielding random variation in the time expired between cash flowing into a household's village (and its surroundings) and the time of outcome measurement. Control villages likewise received an experimental treatment timing ($t_0(v)$) in which they would have been treated, had they been selected for treatment.

The timing of transfers is clustered at the GVH level, such that treated villages will receive transfers approximately at the same time as their neighboring villages. However, villages outside the GVH will receive transfers up to five months before or after one's own village in Phase I, and up to 6 months in Phase II due to random rollout. This will provide additional variation in the time in which nearby villages are treated, which will identify dynamic effects of spillovers across households (even conditional on one's own treatment timing).



The high-frequency nature and coverage of household surveys from as early as one month post transfers is a key strength of our setting, allowing us to trace out the impacts of cash at high temporal resolution. In our intended schedule, Midline 1 spans 1 - 12 months, Endline 1 spans 12 - 24 months after recipients receive their large transfers. Importantly – each control village is also assigned a randomized 'experimental' start date when they would have received transfers, so that this time is well-defined. Overall, we will survey at least 500 households in each month up to two years after large transfers went out.

First, we will estimate pooled regressions across both Midline and Endline 1, as well as conduct tests for whether the estimated effects at Endline 1 are different from those at Midline. Second, we will investigate dynamic treatment effects more flexibly. We modify equations *(1a/b)* and/or *(2a/b)* to estimate impacts not on average across all months and cohorts, but dynamically:

$$y_{i(v,c,k),t} = \psi_{kt} + \sum_{l \in L} \beta_l T^{v}_{v(i),t-l} + \sum_{l \in L} \gamma_l T^{GVH}_{c(i),t-l} + \sum_{l \in L} \delta_l HighSat_{k(i),\, t-l} + \ \dots \ + \varepsilon_{i(v,c,k),t} \qquad (3a)$$

$T^{v}_{v(i),t-l}$ is a dummy for outcome $y_{ivck,t}$ for individual *i* in a treatment village *v* observed at time *t* being observed *l* periods after their experimental treatment timing (and analogously for $T^{GVH}_{c(i),t-l}$, and $HighSat_{k(i),\, t-l}$).[21] ψ_{kt} are cohort-by-period-fixed effects, which control for underlying levels of inflation as well as other time-specific shocks or ambient effects that affect all cohorts similarly over time, and ensure that we only compare outcomes within the same treatment cohort between treatment and control villages.[22]

[21] Note that since all villages within a GVH are treated at the same time, the treatment indicators for the village- and GVH levels will be collinear for a given cohort in a given month.
[22] Since survey timing was clustered at the GVH level (for logistical reasons), and since some GVHs are large, we may not have variation GVH-level treatment in all cohort-by-period cells for short periodicities (e.g. months) that may cover only one GVH in a cohort. Because of this, we will run specifications both with cohort-by-month fixed effects, and cohort- and month- fixed effects separately to test for robustness. If they look quantitatively similar (suggesting limited bias from negative weightings), but the latter is much better powered, we may opt for the latter as primary.

Analogously, we extend the primary spatial specification for the dynamic case:

$$y_{i(v,c,k),t} = \psi_{kt} + \sum_{l\in L} \beta_l Amt_{v(i),\, t-l} + \sum_{l\in L} \gamma_l Amt^{0\ to\ \bar{R}\ km}_{\neg v(i),\, t-l} + \phi_r s^{elig,\, 0\ to\ \bar{R}\ km}_{\neg v(i)} + \omega_1 y^{BL}_{ivck} + \omega_2 M^{BL}_{ivck} + \varepsilon_{i(v,c,k),t} \qquad (3b)$$

where $Amt_{v(i),\, t-l}$ is the amount transferred into village *v* (where *i* is resident) in period *t - l* and $Amt^{0\ to\ \bar{R}\ km}_{\neg v(i),\, t-l}$ is the amount transferred into villages outside *v* within the *0* to $\bar{R}$ km buffer. As for the heterogeneity regressions, we opt for a one-ring spatial specification as our *a priori* primary specification. But we may investigate whether the pattern of spatial decay varies dynamically (e.g. rippling out further over time) by running a multi-ring specification instead. We may prefer to report the multi-buffer specification if spatial decay of spillovers varies dynamically in an economically meaningful and statistically significant way. As above, we may also allow for differential spillovers for recipients and non-recipients if specifications *(1b/2b)* are selected as primary instead of *(1a/2a).*

Periodicity: L are sets of time periods before measurement. There is a trade-off between estimation precision and resolution of the impulse response function. We anticipate leading with a quarterly frequency as our default, but may explore other periodicities, including monthly and bi-yearly.

Note that we treat amounts sent as a 'small' transfer and those sent as a 'large' transfer symmetrically, assuming that the impact of these scales linearly (e.g. as shown is the case across many cash transfer programs in Kondylis & Loeser 2024). In more exploratory analyses, we may thus test whether the small and large transfer impacts after *l* periods are different, by including separate 'amount' terms in Equation *(3b).*

The new DID literature: Several papers (summarized by Roth et al. (2023)) recently pointed out that estimates two-way fixed effects estimators (TWFE) can be biased when treatment timing is heterogeneous as is the case in our study. The concern would be that, if a) effects are heterogeneous across early- and late-treated individuals, and b) our panel is imbalanced (i.e. there are more observations from late-treated cohorts at lower lags and more early-treated cohorts at later lags[23]), then our sampling strategy implies that early-treated cohorts will end up being under-represented among the 'control group' for early observations (with $Amt_{v(i),\, t-l} = 0$ for low *l*) since on average we observe them later. However, the fact that a) everything is randomized, b) we have a sizable pure control group (the 'never treated'), c) we run event-study regressions with fully saturated treatment effects should limit this bias (Borusyak et al. 2024). Simulations show that cohort-by-time fixed effects resolve this issue, analogous to approaches proposed in Callaway, Goodman-Bacon, and Sant'Anna (2024), de Chaisemartin & d'Haultfoeuille (2023), or de Chaisemartin et al. (2024). Another 'clean' way to address the issue is to run our primary static specifications *(1a/b)* and *(2a/b)* separately for each lag, restricting the sample to only observations with experimental

[23] Figure 4 shows how we cross-randomize survey timing relative to treatment timing. This is explicitly attempting to avoid correlation between the time of measurement and the time of treatment. Nonetheless, to maximize power and sample size for each 'time-since-treatment' bin, there will be a slight imbalance in cohorts – with late treated cohorts being somewhat oversampled in early months (and vice versa with early-treated cohorts) to ensure enough sample size at the tail ends.

treatment timing exactly *l* periods ago. For *(2a)*:

$$y_{i(v,c,k),t} = \alpha_t + \beta_l \left(\sum_{l=0}^{\infty} Amt_{v(i),t-l} \right) + \gamma_l \left(\sum_{l=0}^{\infty} Amt^{0\ to\ \bar{R}\ km}_{\neg v(i),\ t-l} \right)$$

$$+ \phi_r s^{elig,\ 0\ to\ \bar{R}\ km}_{\neg v(i)} + \omega_1 y^{BL}_{ivck} + \omega_2 M^{BL}_{ivck} + \varepsilon_{i(v,c,k),t} \qquad (3c)$$

where we restrict the sample to all observations observed at time t such that their experimental treatment timing was at $t_0(v) = t - l$. A complication with this approach is that not all transfers to a village may go out in the same month (since some households' transfers are delayed for operational reasons). Moreover, while treatment timing is clustered within GVH and market catchment area, some villages will be treated before and after village v. We take this into account to some degree by only summing over *past* treatment.

Each approach (3b or 3c) has their relative strengths and weaknesses. Moreover, this literature is rapidly evolving and has not yet settled on the best estimator for settings such as ours. Our choice of the estimator will be guided by the latest developments in the literature at the time of analysis. For now, we will test for robustness of our findings by estimating both *(3b)* and *(3c)* in the spatial case – which should yield similar results, and compare them to a more standard two-way fixed effects estimator (*(3a),* but with period- and cohort-fixed-effects separately) – and analogously for the reduced-form estimates. We consider *(3b)* primary, but may switch to *(3c)* if results are either very different, or yield better powered estimates.

Primary Estimands: As for our main static specifications, the primary dynamic treatment effect estimators of interest (which we also call 'impulse response functions'), remain:

1. the direct + within-village spillover impact of the program on recipients: $\hat{\beta}_l$ from Equation *(3a)* provides the impact *l* periods after the large transfer, and the cumulative impulse response function up to L periods after the large transfer is:

   $$\widehat{IRF}^{reduced-form}(L) = \sum_{l=0}^{L} \hat{\beta}_l T^{v}_{v(i),\ t-l}$$

   Note that this does not include any potential anticipatory effects of cash transfers after the announcement. And, it also does not include any effects of transfers between the small transfer and the large transfer (since no surveys take place between the small and large transfers, and *l* is defined relative to the receipt of the large transfer). For some impacts, we may use additional phone surveys and data from Khongoni to estimate these impacts (see section 4.7.1 below).

2. $\widehat{ATotE}^{recipients}$: $\hat{\beta}_l E(Amt_{v(i)} | T_{v(i)} = 1) + \hat{\gamma}_l E(Amt^{0\ to\ \bar{R}\ km}_{\neg v(i)} | T_{v(i)} = 1)$ from Equation *(3b)* provides the impact *l* periods after *l* periods, and the cumulative impulse response function up to L periods is:

   $$\widehat{IRF}^{recipients}(L) = \sum_{l \in L} \hat{\beta}_l E(Amt_{v(i)} | T_{v(i)} = 1) + \sum_{l \in L} \hat{\gamma}_l E(Amt^{0\ to\ \bar{R}\ km}_{\neg v(i)} | T_{v(i)} = 1)$$

where $Amt_{v(i)}$ and $Amt_{\neg v(i)}^{0\ to\ \bar{R}\ km}$ are the *total* amounts (large and small transfers) transferred into villages and buffers around recipients. This adds up all cumulative period-wise effects of a program that transferred the average amounts all at once up to *L* periods. (Or, alternatively, adds up the effect up to *L* periods for all the money transferred, regardless of when it was transferred.)[24]

3. $\widehat{ATotE}^{non-recipients}$: $\hat{\gamma}_l$ from Equation *(3b),* and $\widehat{IRF}^{non-recipients}(L)$ calculated analogously to $\widehat{IRF}^{recipients}(L)$.

4. $\widehat{GATE}$: $(\hat{\beta}_l + \hat{\gamma}_l)E(Amt_{v(i)}|T_{v(i)} = 1)$ from Equation *(3b)* provides the impact *l* periods after *l* periods, and the cumulative impulse response function up to L periods is:

$$\widehat{GATE}\ (L) = \sum_{l \in L} (\hat{\beta}_l + \hat{\gamma}_l)E(Amt_{v(i)}|T_{v(i)} = 1)$$

where $E(Amt_{v(i)}|T_{v(i)} = 1)$ is the average amount transferred to recipients.

4.6.1 Anticipatory effects of cash transfers

The dynamic estimators above only account for impacts of cash transfers *after* they were received. However, GD informs recipients approximately one to two months prior to the small transfer about the upcoming transfers. During this enrollment stage, recipients may already begin to act differently, e.g. by borrowing against future transfers, in anticipation of the transfers. We conducted phone surveys with a representative subset of Baseline households between announcement and receipt of the first small transfer. For the subset of data that we collect via phone – consumption, loans, labor supply primarily – we will additionally estimate anticipatory effects of cash, and add a period $l = -1$ to our IRFs, lasting for the duration between announcement and receipt of the small transfer (typically one to two months).[25]

4.6.2 Estimating the Marginal Propensity to Spend

A key dynamic object of interest on the household side – which will also enter into the construction of local economic aggregates and the transfer multiplier – is the marginal propensity to spend. We largely follow Egger et al. (2022) in our approach, and split spending into:

a) *Stocks vs. flows:* We collect spending in two different modules in our surveys. First, a consumption/spending module, and an assets module. Assets are collected as stocks,

[24] In Egger et al. (2022), we simulated a cash transfer program that had a transfer roll-out schedule identical to the one implemented, i.e. with multiple transfers. This meant that for any impact up to L months, small transfers would have accumulated effects up to *L* periods, large transfers that were delivered *D* periods after the small transfers would have accumulated effects up to *L - D* periods only, but eventually would have additional effects later on. We opt here for a simpler and more parsimonious characterisation of the cumulative impact up to *L* periods for all the money transferred.
[25] We note that the phone survey sample is representative only of the households owning phones. If we find substantially different average treatment effects by phone ownership in the in-person surveys, we may scale up the treatment effects in the phone survey by the ratio of the average effect between phone owners and non-owners.

while consumption/spending is collected as a flow variable over the last week/month/year. Stocks represent accumulated spending on flows (net of depreciation). When estimating the marginal propensity to spend on flow variables up to L periods after the large transfer, we integrate up the spending impacts from equation *(3b)*: $\hat{IRF}^{recipients}(L)$. For stocks, we instead add up the accumulated stocks up to Midline, and Endline 1 respectively, and assume that asset accumulation occurred linearly up to each survey round (i.e. households spent the same additional amount on accumulating assets each period). That is, for stocks, the accumulated spending impact relies on equation *(2a)*:

$$\hat{IRF}^{stock}(L) = \frac{min(L, L_{ML}^{max})}{\bar{L}_{ML}} \cdot \hat{AtotE}_{ML}^{recipients} + \frac{min(L - L_{ML}^{max}, 0)}{\bar{L}_{EL} - \bar{L}_{ML}} \cdot (\hat{AtotE}_{EL}^{recipients} - \hat{AtotE}_{ML}^{recipients})$$

where L is the number of periods and $\bar{L}_{ML/EL}$ is the average number of periods after transfers at which Midline/Endline data was collected, and L_{ML}^{max} is the maximum number of periods after transfers at which Midline data was collected. The aim of this is to spread out the average impact measured at midline across months during which midline was collected, and any *additional* effect measured at Endline over the months after Midline data collection ended, and endline data collection started.

b) *Spending vs. consumption:* We separate total consumption into cash spending and, for food items only, consumption from own-production (which respondents are asked to value at market prices).

c) *Consumption vs. investment vs. savings:* We separate spending into consumption, investment, and savings. We count as 'investment' any household spending on

d) *Local spending vs. imports:* Our surveys ask, for each product category, the location of purchase. We will separate spending into local spending (including own-consumption of food) and imports from outside Chiradzulu to estimate the marginal propensity to spend locally (MPSL). This will be an important quantity for estimating impacts on local economic activity.

To obtain the share of these consumption responses relative to transfers, we will divide the total impacts calculated above by the average transfer amount received by recipients. Additionally, we will divide impacts on spending the transfer amount + the accumulated impact on household income, i.e. by dividing the LHS by each household's total income reported at time *t* (primary outcome 1.3) and multiplying the coefficient by the average income at time *t* across all households.

4.7 Forecasting long-term treatment effects using short-term outcomes

An emerging literature on statistical surrogates (e.g., Athey et al., 2025) provides methods for identifying long-run treatment effects using intermediate outcomes (surrogates), where the primary outcome is independent of the treatment, conditional on these surrogates. In this context, our interest is in predicting the long-run effects of cash on welfare, as proxied for by consumption. In the context of cash transfers, where interventions have diverse, far-reaching effects including market-level adjustments, the choice of an adequate, minimum set of intermediate outcomes to form the surrogate index for a given outcome is a challenging problem. One core question, however, is whether short-term outcomes are predictive *over and above* baseline covariates in predicting heterogeneous treatment effects in the long-run. Answering this question will have important implications for poverty targeting.

In principle, our large sample would permit a within-study sample split, using one subsample to estimate the surrogate index and the other to estimate treatment effects. Our preferred approach, however, is to construct the surrogate index using external experimental datasets from the cash transfer literature. This strategy avoids loss of power from within-sample splitting and leverages the substantial existing evidence base. In either case, unless more appropriate methods are developed, we will use cross-fitted doubly-debiased machine learning methods following the recommended settings in Athey et al. (2025) and the papers it references. Given the small-to-moderate sample size, we will likely rely on AutoML, with LASSO, elastic nets, random forest and gradient boosting as candidate estimators and hyperparameters selected via cross-validation.

Constructing the surrogate index from *across* studies relies on the assumption that the conditional relationship between intermediate outcomes and long-run outcomes is stable across settings. We will therefore restrict attention to studies with comparable populations, outcome definitions, and time horizons.

4.8 Balance and attrition

To test for balance we run, for each outcome, our main specifications (1a/b) and (2a/b) using y_{ivct}^{BL} on the RHS (and dropping baseline controls). For reduced form specifications, we report β and γ from specification (1a/b) – clustering standard errors at the village level for β and at the GVH level for γ. For spatial specifications, we report $ATotE^{recipients}$ and $ATotE^{non-recipients}$. We report individual Wald test statistics for each outcome and check if the share that rejects the null at 5/10% significance is close to 5/10%. We also report FDR-adjusted q-values across all primary outcomes and within each family. And lastly, we report the p-value from a Wald test in a SUREG framework that tests for each estimand being jointly equal to zero across all primary outcomes (and outcomes within each group). We do this for each sample – those surveyed at Baseline, at Midline, and at Endline 1.

Since Baseline survey instruments were refined and updated between Phase I and Phase II, we may additionally conduct separate balance tests for outcomes where measurement has changed substantially in addition to the pooled (and more powered) balance tests (which already include a phase-fixed-effect.

We already include baseline controls in our primary specifications, and DDML controls in secondary analyses, and therefore do not plan to modify any of the analyses due to unlikely unlucky imbalances beyond that – though we may consider adding controls for significantly imbalanced baseline variables, or make the specification with DDML controls primary.

We analyse attrition by estimating our primary specifications (1a/b) and 2(a/b) using an indicator for being surveyed at Baseline, at Midline, and at Endline 1. For Midline and Endline, we also restrict the sample to those surveyed at Baseline. If we find worrying levels of differential attrition, we will investigate whether attrition is differential on observables by regressing Baseline values of primary outcomes jointly on indicators for being surveyed at Midline / Endline 1 (for those surveyed at Baseline). In this case, 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.

5. Primary outcomes

5.1. General principles

Outlier treatment: We will generally winsorize all unbounded variables (e.g. consumption, revenue, etc.) at the 99% level (if strictly positive) or at the 1% and 99% level when both positive and negative values are possible (e.g. with profits). We apply winsorization at the latest possible step, i.e. when a variable is the sum of several subcomponents (e.g. with consumption), we winsorize total consumption at the end instead of winsorizing each component separately before summing. (Note: This will mean that winsorized subcomponents do not necessarily sum exactly to the winsorized total.) We will test sensitivity to the cutoff points for winsorization for our primary outcomes in a robustness test.

Monetary values: We will generally report monetary values in USD PPP for international comparison. When doing so, we will use the same USD PPP rate (from the World Bank) across all outcomes. Where appropriate, we may deflate monetary values back to baseline values using our own price index collected in the closest month and closest market to each household (see ‘Local Economy PAP’) before converting to USD PPP using the Baseline rate, in order to express values in terms of baseline USD PPP.

Standardized indices: We follow Anderson (2008) when constructing standardized indices. Z-scores are standardized using the mean in control GVHs, and components are weighted by the inverse variance-covariance matrix. When specific outcomes are listed as index components, we also intend to analyze these outcomes and will report coefficients on all subcomponents (possibly in an appendix).

5.2 Multiple Inference adjustment

We follow Haushofer et al. (2017a) for our treatment of multiple inference adjustment. To summarize briefly: The naive p-values on our primary outcomes are correct for readers with an a priori interest in the hypothesis that cash transfers affect a particular outcome among them. To control for multiple inference, we calculate sharpened q-values (i) within each group (economic & non-economic) of our primary outcomes (as defined in section 5.3) and (ii) within each outcome “family” (as defined in section 6) 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 all 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 estimand of interest. We note that norms around multiple testing are still evolving in economics, and through the above methods seek to follow current best practices.

The variables detailed in this section are our primary outcomes of interest; more details on their construction follow in the next section. We consider these outcomes important a priori, but we may investigate evolving machine learning approaches for hypothesis generation to identify outcomes that were affected by treatment from a wider list, such as that proposed by Ludwig et al. (2023).

A - Primary Economic Outcomes

1. Total value of non-land, non-housing assets
2. Total consumption expenditure
3. Total earned household income
4. Total household business revenue
5. Labor supply: Hours worked in the last 7 days

B - Primary Non-Economic Outcomes

6 Food security index
7 Subjective well-being index
8 Health status index
9 Child health and education index
10 Female empowerment index

6. 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 5.2 for each "family" of outcomes, denoted by a separate section, for main outcomes within each family.[26]

Outcomes marked by * were not collected at Baseline Phase 1. Outcomes marked by † were not collected at Midline. Outcomes marked by ‡ indicate variables where the definition has slightly changed between Baseline and Midline; in these cases, we may consider creating consistent measures for robustness. § indicates variables that were only collected at Midline.

1. Assets

Summary measure – Total value of non-land assets‡: Sum of livestock value (1.1), productive assets (1.2), household durables (1.3), and net financial assets (1.4).

1.1. **Total value of livestock**: Sum of value of cattle, goats, sheep, chicken, other birds, pigs and fish and other animals 6.1.5 + 6.1.15a

1.2. **Total value of (potentially) productive and enterprise-related assets**: Sum of value of Grain/stone mill, handheld tools*, plows and complementary tools*, pastoral/livestock tools*, water pump, carts (hand or animal drawn), wheelbarrows and agricultural tools* (variables 4.1.5 (Asset codes 22, 110-115 and 117), and total value of enterprise assets from enterprises owned by the household (variable 7.51.a+7.52.a+ 7.53.a+7.55.a)*). We also include any other assets listed among household durables (1.3) that households report also using for productive/business purposes. Note: This is a relatively inclusive definition of productive assets, and we

[26] All outcomes of the form X.X.X are considered secondary; main outcomes take the form X.X with no other designation as secondary.

may create a secondary split where we include only assets that are exclusively for productive or business use. To avoid duplication, we will count each asset only once, i.e. either as a (potentially) productive, or household item, but not both.

1.3. **Total value of household durables‡**: Sum of the reported values of household durable assets owned by the household. The measure includes bed frames, mattresses, blankets, wardrobes/chests/trunks, tables, chairs/stools, sofas/armchairs, cabinets/cupboards, stoves, kitchen equipment/utensils, bowls/plates/cultery*, irons, refrigerators, lanterns, lamps/torches, clocks/watches, televisions, radios/cassette/CD/music players/speakers, smart phones, non-smart mobile phones, computers/tablets, generators, car batteries, solar panels/solar lanterns, bicycles, motorized vehicles, sewing machines, water tanks, mortar/pestles, mats/mphasa, clothes and shoes, plastic/wooden buckets and crates, iron sheets, bricks/stones, other building materials, and suitcases. The measure excludes gold and silver jewelry and productive or enterprise-specific assets such as grain/stone mills, hand tools, plows etc. We attribute 'other' non-mentioned assets based on their most suitable category into household or productive assets. To avoid duplication, we exclude abovementioned assets from household durables if the respondent reports that they are also used in an enterprise owned by the household, and include them among 'productive' assets instead.

1.4. **Net financial assets:** Net financial assets (assets 1.4.1 - liabilities 1.4.2)

  1.4.1. **Total financial assets:** Sum of financial balances from bank, mobile money, cash, personal savings, group savings (variables 4.2.1-4) + outstanding loans given to anyone outside the household in the last 12 months that respondent expects to get back (variable 4.3.18. if 4.3.19 is yes + 4.3.20.a.ii)

  1.4.2. **Total financial liabilities:** Sum of outstanding loans taken from commercial banks/lenders/MFI, moneylenders and Community Savings groups* (Variables - 4.2.6a-8a + 4.2.8a*) + loans from anyone else outside the household in the last 12 months that the respondent expects to pay back (4.3.8b if 4.3.9 is yes + 4.3.10.a.ii)

  1.4.3. **Credit constraints*:** The maximum amount of money the household could raise in an emergency (4.3.21)* (We will also look at an indicator for being able to easily raise 100,000 MWK - variable 4.3.21.)

  1.4.4. **Interest rate*:** Average interest rate on any currently outstanding loan. Constructed from variables 4.2.6.d-f*, 4.2.7.d-f* or 4.2.8.d-f*

1.5. **Total acreage of land owned:** Variable 4.1.1

  1.5.1. **Value of land owned by the HH:** Total amount of land owned in acres * price per acre, where price = rental price per m2[27] / interest rate, where the rental price is an average in village-level reported annualized rental costs (1.5.2), and the interest rate is an average village-level interest rate (1.4.4).

  1.5.2. **Land rental price per acre**: Constructed from variable 6.2.6.j

1.6. **Estimated housing rental value of owned dwelling (annualized)**, respondents own report (Variable 4.0.2.b)

[27] Our primary specification uses households' own valuation of their land. In a robustness check, we will use village or regional average prices of land to value agricultural land.

1.6.1. **Housing value:** Rental value annualized (1.6) / interest rate (1.4.4).

1.6.2. **House has non-thatched roof**: Indicator from variable 4.0.5.a for roof made of materials other than grass or leaves. Incomplete roofs are also coded as zero.

1.6.3. **House has electricity**: Indicator from variable 4.0.7 for having electricity from any source.

1.6.4. **Household has a latrine (inside vs outside the house)**: Indicator from variable 4.0.6 for toilet

2. Consumption

Summary measure – Total consumption in last 12 months: Sum of annualized total food consumption (outcome 1.2.1) and annualized non-food expenditures (outcome 1.2.2)

2.1. **Total food consumption in the last 12 months***‡: For 15 food categories, expenditure in the last 7 days (outcome 2.1.1) plus consumption value from own production (outcome 2.1.2) and gifts received (outcome 2.1.3) in the last 7 days, annualized

2.1.1. **Expenditure on food items:** variables 5.1.1

2.1.2. **Food consumption from own production:** variables 5.1.2

2.1.3. **Food consumption from gifts:** variable 5.1.3

2.2. **Total non-food consumption/expenditure in the last 12 months:** variables 5.2.1.a-aa + 5.2.3.b

2.3. **Total expenditure (excl. home production & gifts) in the last 12 months:** variables 5.1.1 + 5.2.1.a-aa + 5.2.3.b

2.4. **Total investment expenditure:** Sum of housing expenditure (2.4.1), durables expenditure (2.4.2), education expenditure (2.4.3), and health expenditure (2.4.4.), plus investment in agriculture, livestock, and non-agricultural businesses (3.6), all annualized.

2.4.1. **Total housing expenditure in the last 12 mo**.*: Sum of home repair/maintenance, and improving/expanding home (variable 5.2.1.p-q).

2.4.2. **Total expenditure on durables in the last 12 mo**.*: Variable 5.2.1.w

2.4.3. **Total education expenditure in the last 12 mo**.: Variables 5.2.1.r.

2.4.4. **Total medical expenditure in the last 12 mo**.*: Variable 5.2.1.v.

2.5. **Total expenditure on temptation goods in the last 12 mo**.*: Sum of alcoholic drinks and tobacco products in the last 7 days (variable 5.1.1.n) plus lottery tickets/gambling in the last month (variable 5.2.1.d), both annualized.

2.6. **Total social expenditure in the last 12 mo**.*: Sum of recreation/entertainment expenses in the last month (variable 5.2.1.g) times 12 plus religious expenses, charitable donations, weddings and funerals (variable 5.2.1.s-u), and dowry/bride price in the last 12 months (variable 5.2.1.x).

2.7. **The flow-value of durables:**We commit to exploring impacts on estimates of the flow value of durables, constructed as described in Haushofer et al. (2017a), and using methods summarized in Amendola & Vecchi (2014). We collect prices and current values for different vintages across all assets, allowing us to derive asset-specific depreciation rates. We will follow best practices in the evolving literature to use these to construct flow values of durables. We think this is a quantitatively important dimension for the welfare impacts of large lump-sum cash transfers and may include it in our primary measure of consumption if operationally/computationally feasible.

Spending in line with plans: We will also look at whether households spent money (or, for treated households, their transfer) in line with baseline intended spending plans, and whether this concordance with plans. Since these measures are difficult to define for treated and control group alike, and partially ex-post endogenous, we consider this analysis important but exploratory.

3. Income and profits

Summary measure – Total pre-tax earned household income in the last 12 mo.: Sum of total profits from agriculture and livestock in the last 12 months (3.1) plus total profits from non-ag. business in the last 12 months (3.2) and the total value of wages, salaries and in-kind transfers earned in the last 12 months (3.3) plus total cash value of government or NGO transfers (3.4) + net private remittances (3.5) + net rental income (3.6).

3.1. **Total profits from agriculture and livestock in the last 12 mo**.: 3.1.1 + .3.1.2

3.1.1. **Profits from agriculture:** Value of agricultural output (for each crop, the quantity produced valued at the village average price across all households of that quantity, outcome 1.4.2.1) net of loss of harvest from theft/vandalism/natural disasters* (variable 6.2.20), salaries paid to agricultural workers outside the household (variable 6.2.30), agricultural inputs (variables 6.2.23* minus 6.2.25* + 6.2.31a minus 6.2.31b + 6.2.32a minus 6.2.32b + 6.2.35.a-e), net rental cost of land for agricultural activities (Variables 6.2.6i multiplied by 6.2.6j minus 6.2.6k multiplied by 6.2.6l).

We will also calculate profits using the secondary measure of agricultural output described in 1.4.2.1.)

3.1.2. **Profits from livestock[28]:** Revenue from livestock (1.4.1.3.), net of salaries paid to pastoral workers outside the household (variable 6.1.19), pastoral inputs (variables 6.1.22.a-d), rental cost of pastoral land (variable 6.1.21.a multiplied by 6.1.21.b)

3.1.3. **Net revenue from livestock sales and purchases:** Variables (6.1.7 plus 6.1.15b) minus (6.1.10 plus 6.1.15c). Also added to livestock profits in an alternative measure of livestock profits.

3.2. **Total profits from non-ag. business in the last 12 mo**.: Annualized sum of self-reported profits for all businesses owned (variable 7.27b multiplied by 7.29 if joint ownership and annualized by multiplying by the number of months it was open 7.20).[29]

3.3. **Total wages, salaries and in-kind transfers earned last 12 mo**.: Sum of cash salary last month (variable 8.24) plus total value of benefits last month (variable 8.25.a-f),

---

[28] Livestock sales and purchases can serve both as a business for generating income (e.g. for households engaging in livestock fattening or trading as a livelihood strategy) as well as as a storage of value / asset. Annual profits from livestock would include the expected appreciation/sales value of a household's stock over the course of the year, yet this is tricky to measure. Our primary measure of income therefore excludes net livestock sales, but we look at this as a secondary outcome.

[29] Our primary measure will use self-reported profits. We may explore alternative constructions based on revenues minus costs for robustness.

annualized to a yearly measure by multiplying by the number of months worked in the last 12 months for all workers in the household.

3.3.1. **Hourly wage rate for those employed/working for wages**: Cash salary last month (variable 8.24) times (7/30) to get a measure of weekly earnings, divided by total hours worked in the last week (variable 8.23), 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 9), we look at selection into working for wages as an outcome.

3.4. **Government/NGO transfers:** Sum of outcomes 3.4.1-2.

3.4.1. **Government transfers :** government transfers from Social Cash Transfer Programme (SCTP), climate smart enhanced public works program*, fertilizer/pesticide subsidies in the form of coupons from the Affordable Input Program (asked in Section 11 in Phase I and Section 6 in Phase II). Variables 11.1.a plus 11.1.d plus 6.2.25, 6.2.31b, 6.2.32b and 6.2.33b if Phase II or 11.4.a if Phase I.

3.4.2. **Cash/in-kind transfers from Gov/NGOs/Church:** cash and in-kind food and non-food aid from various sources like government, NGOs, churches/mosques, etc. Sum of variables 11.2.c 11.2.e, 11.3.a.ii, 11.4.b and 11.7*.

3.4.3. **Taxes paid (incl. fees, bribes, etc.):** Variables 5.2.1.y-z

3.4.4. **Contribution to local public goods[30]:** Sum of variables 5.2.1.aa and 15.1.6*

3.5. **Net value of private remittances and goods sent in the last 12 mo.:** Remittances received (3.5.2) - remittances sent (3.5.2).[31]

3.5.1. **Remittances and goods sent:** Sum of variables 4.3.8.b if 4.3.9 is no + 4.3.10.a.i (depending on number of transfers received) across all transfer relationships. Not including loans.

3.5.2. **Remittances and goods received:** Sum of variables 4.3.18.b if 4.3.19 is no + 4.3.20.a.i (depending on number of transfers sent) across all transfer relationships. Not including loans.

3.6. **Net land rental income in the last 12 mo.**: For agricultural activities (Variables 6.2.6i multiplied by 6.2.6j minus 6.2.6k multiplied by 6.2.6l) and for pastoral activities (variables 6.1.21.a multiplied by 6.1.21.b) minus house rental payments (variable 4.0.2.b).

We anticipate further exploring the spatial patterns of income within and across study area villages and markets as well as transactions with agents outside the study area (factor income from 'abroad'). Bringing in enterprise data, we also aim to attribute income to factor income paid by locally operating enterprises. We will provide more details in the local economy PAP.

[30] We will file a more detailed PAP on civic engagement and political outcomes where we look into this more closely, including volunteer labor for public goods construction, participation in community groups, vote buying, handouts, etc.

[31] Remittances sent could, in some sense, be construed either as a tax or as a consumption expenditure (e.g. if sent to children living outside the household). We may therefore also report total income before subtracting remittances sent, if this is quantitatively important.

4. Business revenue and operations

Summary measure – Total household business revenue in the last 12 months: Sum of total revenue from agriculture and livestock in the last 12 months (outcome 4.1) plus total self-employment revenue in the last 12 months for all businesses owned (outcome 4.2).

4.1. **Total revenue from agriculture and livestock in the last 12 mo.**: Sum of value of agricultural production (4.1.1.) and pastoral revenue (4.1.7.)

4.1.1. **Value of agricultural production[32]:** Value of agricultural output (for each crop, the quantity produced valued at the village average market price). Constructed from variable 6.2.7 summed across all the crops grown (codes (1-50)).

Secondary measure: Sum of a) income from crop sales (variable 6.2.11), b) value of crop consumed from own production (variable 6.2.8), c) value of crop currently in storage* (variable 6.2.18), again valued at village average market price.

4.1.2. **Revenue from agriculture in the past 12 mo:** Value of all agricultural output sold (excl. own consumption, gifts, etc.). Primary measure uses units of crops sold multiplied (variable 6.2.10) by market prices, where we use average price of the crop from our price surveys from the closest market, where available and average reported unit values across all households reporting selling the crop within the village (if available) or GVH. Secondary measure uses crop-level self-reported revenue from selling crops (variable 6.2.11)

4.1.3. **Reported agricultural output prices:** Variable 6.2.12 for Crop codes (1-50), revenue-weighted index across all crops, standardized by control village mean for each crop.

4.1.4. **Total acreage of land used:** Variable 6.2.6.a. Note that in Baseline Phase 2 we asked about specific plot/garden used for agriculture in the last 12 months and in Baseline Phase 1, we asked about total size of the land used for agricultural activities in the last 12 months

4.1.5. **Maize yield (kg):** Variable 6.2.7 for Maize.

4.1.6. **Agricultural productivity‡:** Calculated as value of crop output/harvest per acre/m2.

4.1.7. **Revenue from livestock in the past 12 mo:** Sum of livestock sales and sale (6.1.7 plus 6.1.15b) and the value of livestock animals/poultry/fish/animals products sold and consumed from own production (Variables 6.1.13 plus 6.1.14). We will also look at both separately.

4.1.7.1. **Livestock sales:** Sum of variables 6.1.7 and 6.2.15.b

4.1.7.2. **Sale of livestock products:** Value of livestock animals/poultry/fish/animals products sold and consumed from own production (Variables 6.1.13 plus 6.1.14)

[32] Our primary measure uses the household's reported quantity harvested multiplied by market prices, where we use average price of the crop from our price surveys from the closest market, where available and average reported unit values across all households reporting selling the crop within the village (if available) or GVH. As a secondary measure, we also use the household's own reported harvest value.

4.2. **Total revenue from non-ag. business in the last 12 mo**.: Sum of variable 7.24.b multiplied by number of months business was open (7.20) for all businesses owned.

4.2.1. **Number of non-ag. business operated by any household members in the last 12 months**: Variable 7.3.

4.3. **Total number of employees employed in household businesses:** 4.3.1 + 4.3.2.

4.3.1. **Total number of workers employed for agricultural work‡.:** Sum across the number of workers employed across each agricultural (variable 6.2.28) and pastoral activity (variable 6.1.18).

4.3.2. **Total number of employees of non-ag businesses**: Variable 7.38

4.4. **Total household productive capital stock:** Sum of capital stock in agriculture (4.4.1), livestock (4.4.2), and non-ag enterprises (4.4.3)

4.4.1. **Total value of capital stock for agriculture*****:** Sum of value of grain/stone mill, handheld tools*, plows and complementary tools*, pastoral/livestock tools*, water pump, carts (hand or animal drawn), wheelbarrows and agricultural tools* (variables 4.1.5), and the value of agricultural land used by the household (outcome 1.1.5.1). (Note: if land values turn out to be too noisy, we may look at non-land assets separately.)

4.4.2. **Value of livestock:** Total value of all livestock owned by the household. Constructed by summing the household's self-reported total herd value for each animal type owned (variable 6.1.5) across all animal types.

4.4.3. **Total value of capital stock for non-ag enterprises excluding land*****:** Sum of the value of all the buildings[33], furniture, vehicles, machinery and tools owned (variable 7.51.a + 7.52.a + 7.53.a + 7.55.a)*, stock, and supplies owned by the enterprise (variables 7.43.b)

4.4.4. **Total outstanding business loans*:** Sum of value of outstanding loans (moneylenders, commercial banks, savings associations like ROSCA, friends and family) taken out specifically for business use. Constructed from variables 7.47.a, 7.48.a, 7.49.a and 7.50.a

4.5. **Total household operating costs in the last 12 mo.:** Operating costs in agriculture and livestock (4.5.1) and non-agricultural businesses (4.5.2).

4.5.1. **Total operating costs in agriculture and livestock (annualized)‡**.: Sum of labor costs (4.5.1.1), land rental costs (4.5.1.2) and non-labor non-land input costs (4.5.1.3).

4.5.1.1. **Labor costs in ag & livestock (annualized):** Total wages paid to workers outside the household (variable 6.2.30 plus variable 6.1.19).

4.5.1.2. **Land rental costs‡:** rental cost of agricultural land (variable 6.2.6.k multiplied by 6.2.6.l) and rental cost of pastoral land (variable 6.1.21.a multiplied by 6.1.21.b)

4.5.1.3. **Non-labor, non-land inputs:** Sum of expenditure on agricultural and pastoral inputs in the last 12 months: (variables 6.2.23* minus 6.2.25* + 6.2.31a minus 6.2.31b + 6.2.32a minus 6.2.32b + 6.2.35.a-e) plus (variables 6.1.22.a-d)

---

[33] We use the same valuation method based on imputed rent as specified in the assets section.

4.5.2. **Total operating costs in non-ag. business in the last 12 mo**.‡: Sum intermediate input costs (4.5.2.1), labor costs (4.5.2.2), capital costs (4.5.2.3),

4.5.2.1. **Intermediate input costs:** Annualized expenditure on stock/inventory/material/supplies **(**variables 7.44)

4.5.2.2. **Labor Costs:** Monthly Wages paid to workers outside of the household (variable 7.40) multiplied by the months the business was operational (variable 7.20)

4.5.2.3. **Capital rental costs*.:** Sum across businesses of monthly rental[34] costs for buildings/premises, furniture, vehicles, and machines/tools/electronics, plus monthly payments on business loans from commercial lenders/MFIs, moneylenders, ROSCAs/savings groups, and friends/family (variables 7.47.b + 7.48.b + 7.49.b + 7.50.b)* plus (variables 7.51.b + 7.52.b + 7.53.b + 7.55.b))*, multiplied by months operational (variable 7.20).

4.5.2.4. **Other costs (taxes, fees, etc):** Annualized costs of security, storage, advertising, banking (variable 7.58*) plus monthly fees or bribes (variable 7.57*)

4.6. **Total enterprise investment in the last 12 mo excluding land.:** 4.6.1 + 4.6.2.

4.6.1. **Total investment in agriculture & livestock in the last 12 mo‡:** Sum of expenditure on Tools & Machinery, Buildings or structures, and Irrigation / water, including ditches, drainage, etc. for agriculture in the last 12 months (variables 6.2.36.a-c) and sum of expenditure on tools and machinery for pastoral activities in the last 12 months (variable 6.1.22.a)

4.6.2. **Total investment in non-ag enterprises in the last 12 mo.***‡: Sum of all enterprises' total expenditure/investment in the past 30 days across all assets, times number of months the businesses were operational (variables 7.46 multiplied by variable 7.20)

4.7. **Agricultural practices index‡:*** Weighted, standardized index of indicators for whether the household adopted the following agricultural practices: irrigation, row planting, ploughing, composting, crop rotation, leaving land fallow, organic fertilizer, chemical fertilizer, pesticides/herbicides, and improved seeds. Higher values indicate adoption of more productive agricultural practices. The index is constructed among households engaged in agricultural crop production. This is constructed using variables 6.2.31-6.2.32 and 6.2.34.a-h.

These outcomes are intended to measure business activities of households surveyed. In the local economy PAP, we will additionally bring in more detailed data from our enterprise survey that can be combined and integrated with these data, and will have a richer analysis of the supply/production side of the economy.

5. Labor supply and time use

Summary measure – Total household hours worked in the last 7 days (5.2): Sum of respondent's hours worked in agriculture (variable 6.2.26), pastoralist activities (6.1.17),

[34] Note, in separate enterprise surveys, we asked for more detailed information on capital repair and maintenance costs. These were not asked in household surveys (which received a smaller set of questions). We will include analyses on repair and maintenance in our local economy PAP.

self-employment (variable 7.36) and employment in the last 7 days (variable 8.23), summed across all household members.

5.1. **Respondent total hours worked in the last 7 days**: 5.1.1-3.

 5.1.1. **Respondent hours worked in ag & livestock:** Respondent's hours worked in agriculture (variable 6.2.26) + pastoralist activities (6.1.17)

 5.1.2. **Respondent hours worked in self-employment:** Respondent's hours worked in self-employment in the last 7 days (variable 7.36).

 5.1.3. **Respondent hours worked in employment:** Respondent's hours worked in employment in the last 7 days (variable 8.23)

 5.1.4. **Respondent currently employed:** Indicator for respondent is currently employed/working for pay (variable 8.1)

 5.1.5. **Respondent currently self-employed:** Indicator for respondent main decision-maker for self-employed enterprise (variable 7.2)

 5.1.6. **Respondent hours spent actively searching for jobs, applying for jobs, or in interviews in the last 7 days***: Variable 8.30

 5.1.7. **Respondent's hours spent on household chores in the last 7 days**: Variable 9.1a-e.

5.2. **Total household hours worked in the last 7 days**: Sum of all household members' hours worked in agriculture (variable 6.2.26), livestock (6.1.17), self-employment (variable 7.36) and employment in the last 7 days (variable 8.23).

 5.2.1. **HH hours worked in ag & livestock:** Hours worked in agriculture (variable 6.2.26) + pastoralist activities (6.1.17), summed across all HH members.

 5.2.2. **HH hours worked in self-employment:** Hours worked in self-employment in the last 7 days (variable 7.36)), summed across all HH members.

 5.2.3. **HH hours worked in employment:** Hours worked in employment in the last 7 days (variable 8.23)), summed across all HH members.

 5.2.4. **Respondent currently employed:** Indicator for any household member being currently employed/working for pay (variable 8.1)

 5.2.5. **Respondent currently self-employed:** Indicator for any household member being the main decision-maker for self-employed enterprise (variable 7.2)

 5.2.6. **Total household hours spent actively searching for jobs, applying for jobs, or in interviews in the last 7 days***: Variable 8.30

We will further explore results at the individual level, including all adult household members and running regressions at that level.

6. Food security

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

6.1. **No. of days adults in household skipped meals or cut the amount of meals in the last 7 days**: Variables 12.3.a

6.2. **No. of days children in household skipped meals or cut the amount of meals in the last 7 days**: Variables 12.3.b

6.3. **No. of days adults in household gone entire days without food in the last 7 days**: Variable 12.4.a

6.4. **No. of days children in household gone entire days without food in the last 7 days**: Variable 12.4.b

6.5. **No. of days adults in household gone to bed hungry in the last 7 days**: Variable 12.2.a

6.6. **No. of days children in household gone to bed hungry in the last 7 days**: Variable 12.2.b

7. Subjective well-being and mental health

Summary measure – Subjective well-being index‡: Weighted, standardized average of depression, anxiety, perceived financial stress*, self-efficacy/locus of control, and overall life satisfaction, appropriately signed so that positive values represent better subjective well-being.

7.1. **Depression**: Scale score calculated from PHQ-2 questionnaire in section 13.1

  7.1.1. Indicator for PHQ-2 exceeds a score of 2

  7.1.2. Indicator for PHQ-8 exceeds a score of 9 (moderate or severe depression).

7.2. **Anxiety**: Scale score calculated from GAD-2 questionnaire in section 13.1

  7.2.1. Indicator for GAD-2 exceeds a score of 2

7.3. **Perceived Financial stress***‡: Scale score calculated from - 13.4.1-.4. This was not collected in Phase 1 baseline.

7.4. **Self-efficacy/locus of control:** Scale score calculated from the Section 13.3. Higher values indicate greater perceived self-efficacy/control.

7.5. **Wellbeing**: Scale score calculated from Variable 13.1.1 which asks about the respondent's overall life satisfaction. This was not collected at Baseline.

8. Health

Summary measure – Health status index‡: Weighted, standardized average of self-reported health (outcome 1.8.1), index of symptoms (outcome 1.8.2), and experienced a major health problem (outcome 1.8.4), appropriately signed so that positive values indicate better health outcomes for the respondent.

8.1. **Self-reported level of general health*‡**: If the respondent reports health is "very good" in variable 12.5, coded as 4, otherwise from 12.6 "somewhat good" is coded as 3, "poor" is coded as 2, and "very poor" is coded as 1.

8.2. **Index of recent symptoms†**: Weighted, standardized average of variable 12.7.A-Y of indicators for health conditions.

8.3. **Days of work, school missed by the respondent due to poor health in the last 30 days†**: Variable 12.14.

8.4. **Average number of days of work, housework or school missed across all household members due to poor health in the last 30 days**: Average number of days missed across household members aged six or older. Constructed using roster variable 3.0.11

8.5. **Experiences major health problems that affect work or life†**: Variable 12.10.

8.6. **Total out-of-pocket medical expenditure in the last 30 days:** Sum of expenditure on hospital/clinic medical care and modern medicines for the respondent. Constructed from variables 12.8 and 12.9

8.7. **Indicator for using modern treatment following a major health problem†**: If the respondent visiting a medical professional, taking prescribed medicine, or taking over-the-counter medicine. Constructing using variable 12.2(options 1-3)

8.8. **Functional difficulty/disability index:** Average of indicators for whether household members experience difficulty seeing, hearing, walking, remembering/concentrating, completing self-care, or communicating. Appropriately signed so higher values indicate fewer functional difficulties.

8.8.1. **Proportion of household members with any functional difficulty:** Number of household members selected in at least one functional-difficulty domain divided by household size. Constructed using variables 12.13 to 12.18.

9. Child health and education index

Summary measure – Child Education and Health index: Weighted, standardized average of total education expenditure (outcome 9.1) and proportion of school-aged children in school (outcome 9.2), appropriately signed so that higher values represent better education outcomes.

9A - Education

Summary measure – Education index: Weighted, standardized average of total education expenditure (outcome 9.1) and proportion of school-aged children in school (outcome 9.2), appropriately signed so that higher values represent better education outcomes.

9.1. **Total education expenditure in the last 12 months**: Variable 5.2.1.r

9.2. **Proportion of school-aged children in school**: Number of school-aged children (6-16) in the household roster attending school in current/most recent school term (variable 3.0.10), divided by total number of school-aged children on the household roster (3.0.4).

9.3. **No. of days attended school in the last five days school was in session**: Variable 3.0.14 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), converted to a household-level average across school-aged children.

We will also look at several secondary measures:

9.4. **Attendance by schooling type (primary, secondary and tertiary)**: Defined as the share of household members in each relevant school-age group attending the corresponding schooling type: primary-age children (6-13 yrs) attending primary school, secondary-age children (14-18) attending secondary school, and tertiary-age youth (19-24) attending post-secondary/tertiary education. Constructed using variable 3.0.10 (current enrollment) and variable 3.0.4 (age of the household member)

9.5. **Days of school missed due to being sent home because of school fees, most recent school term***: Variable 3.0.15 for household members enrolled in the current school year. Set to zero if the child is sent home zero times. Converted to a household-level average across school-aged children.

9B - Child Health

Summary measure - Early Child and maternal health care index*†**:** We construct a weighted, standardized index of recent child survival, antenatal care, facility delivery, postnatal care, and reported pregnancy/birth complications, signed so that higher values indicate better child/maternal health outcomes. We construct child/birth-level outcomes for eligible births and aggregate to the household level by averaging across eligible births in the household.[35] (Note: the following outcomes are only measured for births after 2023).

9.6. **Under-1 mortality***†: Variables 3.1.g.ii-iv

9.7. **Indicator for receiving at least four (4) antenatal care visits***†: Receipt of at least 4 antenatal (ANC) visits is the recommendation of the World Health Organization's (WHO) Focused Antenatal Care Model (FANC), which was designed to promote quality antenatal care in resource-constrained settings where the traditional ANC model (comprising 7-16 visits) would be a challenge to implement (Mchenga et al. 2019). Constructed from Variable 3.1.g.viii.1

9.8. **Indicator for receiving postnatal care***†**:** Variable 3.1.g.x

9.9. **Indicator for delivery at a health facility vs home***†**:** Variable 3.1.g.ix

9.10. **Indicator for birth complications during delivery*†:** Variable 3.1.g.ix.8

9.11. **Out of pocket expenditure on antenatal, delivery and postnatal visits*†:** Sum of variables 3.1.g.viii.8, Variable 3.1.g.ix.11 and Variable 3.1.g.x.7

Our aim is to capture variation driven possibly by cash transfers. As cash transfers rolled out over a period of multiple years and may have created both direct and spillover effects that operated at different time and spatial horizons it is not obvious which cutoff birth date to use for child outcomes. In our primary analyses, we will focus only on children born after the experimental start date (the randomly assigned date at which transfers to a village start). In secondary analyses, we include all children born since 2025 (when the first transfers in the study area went out).

For consistency with other outcomes in this pre-analysis plan, our main measure averages across children to generate a household-level measure that can be analyzed with the same regression equations as outlined above. In addition to averaging across children, we will also conduct a child-level analysis, which allows us to include additional controls, such as indicators for gender, child age bracket and/or year of birth.

In exploratory analysis, we may also investigate whether effects depend on child gender, age, etc. (at the time of transfer) and the time spent within the household since then. Beyond average effects, we are also interested in distributions of outcomes (particularly extreme negative outcomes) and anticipate exploring this using e.g. quantile regressions, or the minimum outcomes across children in each household.

10. Female empowerment

A number of the outcomes in this section rely on questions from section 14 of the household survey and are only administered to respondents where we could ensure full privacy and who are married or cohabiting with their partners. Some of these questions like partner

[35] We may file a separate pre-analysis plan focused specifically on fertility, maternal care, early child health, and child mortality at a later date.

expectations and violence are only administered to married/cohabibiting women, so our summary measure (and a number of the outcomes below) will be restricted to this sample.

Summary measure – Female empowerment index‡: Weighted, standardized average of indices capturing financial autonomy, intra-household decision making,domestic discord, economic control and intimate partner violence, appropriately signed so that positive values reflect more female empowerment/less domestic violence.

10.1. **Financial autonomy index‡:** Weighted, standardized average of indicators for whether the respondent has her own bank account, own mobile phone, and own mobile money account constructed using variables 14.1

10.2. **Bargaining power/relative economic position‡:** Measure of respondent's earnings relative to her partner; reverse-coded so higher values indicating the respondent earns the same or more than her partner. Constructed using variable 14.4

10.3. **Intra-household decision making index ‡**: Average of indicators for whether the respondent has meaningful say, either alone or jointly with her partner, in decisions about buying/selling household durables, children's schooling, children's clothes/shoes, child discipline, and household savings. Higher values indicate greater respondent participation in household decision-making. Constructed using variable 14.5

10.4. **Economic control index ‡:** Weighted, standardized average of indicators for partner requiring permission for large purchases, taking respondent's earnings/savings, refusing money for household expenses, or requiring respondent to give up/refuse work; reverse-coded so higher values indicate less control/coercion by the respondent's partner. Constructed using variable 14.7

10.5. **Domestic discord index ‡**: Weighted, standardized average of frequency of arguments over major household items/assets, child behavior/discipline, relatives, and alcohol; reverse coded so higher values indicate less conflict.Constructed using variable 14.8

10.6. **Controlling behavior index ‡:** Weighted, standardized average of indicators/frequencies for partner jealousy when respondent talks to other men, accusations of unfaithfulness, limits on contact with family, and insistence on knowing respondent's location; reverse-coded so higher values indicate less controlling behavior. Constructed using variable 14.9.

10.7. **Violence index‡**: Weighted standardized average of frequencies of physical, emotional, and sexual violence; reverse coded so higher values indicate less incidence of each type of violence. Constructed using variable 14.10

  10.7.1. **Frequency of physical violence in the last 12mo.‡**: Sum of indicator variables for 'sometimes' or 'often' experiencing physical violence (variable 14.10).

  10.7.2. **Frequency of emotional violence in the last 12 mo**.‡: Sum of indicator variables for 'sometimes' or 'often' experiencing emotional violence (variables 14.10).

  10.7.3. **Spouse raped or performed non-consensual sexual acts on the respondent‡**: Indicator variable from variable 14.10.

10.7.4. **Frequency of sexual violence in the last 12 mo**.‡: Sum of indicator variables for ‘sometimes’ or ‘often’ experiencing sexual violence (variable 14.10)

11. Migration

11.1. **Net change in number of household members since Baseline**: Number of household members currently in household (other than respondent) minus number of household members as of Baseline date (from variables 3.0.3).

We will also look at emigration and immigration separately, as well as at temporary / return migration, and we will look at migration by destination, including by within vs. outside the study area, rural vs. urban destinations, and treated vs. control villages. We are particularly interested also in migration of younger recipients aged 18 - 24.

11.2. **Indicator for emigration:** Indicator for whether the entire household moved away.

We will also look at migration destinations (e.g. by within vs. outside the study area, rural vs. urban destinations, and treated vs. control villages).

Note: In longer-term follow-ups, we may conduct additional censuses or village-level surveys with village elders asking about any new households that moved into the village, information which we will bring in to look at migration at the extensive margin.

12. Marriage and fertility

We look at two outcomes to study changes in marital patterns over time (we do not seek to interpret either of these as necessarily positive or negative):

12.1. **Newly married since Baseline**: Constructed by comparing changes in household rosters between baselines and midlines.

12.2. **Separated/divorced since Baseline**§: Indicator from variable 2.9 for having separated or divorced since Baseline, conditional on ever being married.

We also explore, primarily for new marriages since Baseline (but we can also look at this across the full sample) constructing an index of marital quality, which we construct as a mean effects index based on i) age gap (male - female), ii) education level, iii) ownership of mobile phone, television, radio. We are particularly interested in marriage outcomes by gender, and we may lead with presenting results disaggregated by gender. In Wave 2, we will additionally include questions about arranged marriages as well as payments of bride price for new marriages which we will also explore.

12.3. **Number of pregnancies since Baseline***†: Number of pregnancies since Baseline (variable 3.1.g.vii) for respondents under age 50 at Baseline. (Note: for male respondents, this is the number of pregnancies of all their partners.)

We are interested in fertility patterns by gender and may lead with disaggregated results. Heterogeneity by baseline age may also be especially important for this set of outcomes.

13. Crime and safety

Summary measure - Crime and Safety Index: Standardized index across crime and safety outcomes 1.1.1-7 (signed such that positive is good).

13.1. **Indicator for any crime in the last 12 months*‡**: Indicator for any theft, robbery or assault inside or outside the homestead. Constructed from variable 4.1.12b

13.2. **Indicator for theft/robbery in the last 12 months‡:*** Constructed from variable 4.1.12b

13.3. **Indicator for fraud/scam in the last 12 months‡:*** Constructed from variable 4.1.12b

13.4. **Indicator for assault/violent crime in the last 12 months‡:*** Constructed from variable 4.1.12b

13.5. **Total livestock losses from theft, vandalism, or destruction‡**:Constructed from variable 6.1.29

13.6. **Total crop/agricultural losses from theft, vandalism, or destruction‡**::Constructed from variable 6.2.19

13.7. **Total cost of losses to theft, robbery or scams in the last 12 months:** Sum of variables 4.1.12c.i

References

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., & Wolf, N. (2015). The economics of density: Evidence from the Berlin Wall. *Econometrica, 83*(6), 2127–2189. https://doi.org/10.3982/ECTA10876

Amendola, N., & Vecchi, G. (2014). *Durable goods and poverty measurement* (Policy Research Working Paper No. 7105). World Bank.

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, 103*(484), 1481–1495.

Andrews, I., Kitagawa, T., & McCloskey, A. (2024). Inference on winners. *The Quarterly Journal of Economics, 139*(1), 305–358. https://doi.org/10.1093/qje/qjad043

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics, 30*, 98–108. https://doi.org/10.1016/j.labeco.2014.05.008

Athey, S., Chetty, R., Imbens, G. W., & Kang, H. (2019). The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. *NBER Working Paper No. 26463*. https://doi.org/10.3386/w26463

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics, 47*(2), 1148–1178. https://doi.org/10.1214/18-AOS1709

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika, 93*(3), 491–507. https://doi.org/10.1093/biomet/93.3.491

Borusyak, K., & Hull, P. (2023). Nonrandom exposure to exogenous shocks. *Econometrica, 91*(6), 2155–2185. https://doi.org/10.3982/ECTA19367

Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *The Review of Economic Studies, 91*(6), 3253–3285. https://doi.org/10.1093/restud/rdae007

Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. C. (2024). Difference-in-differences with a continuous treatment. *NBER Working Paper No. 32117*. https://doi.org/10.3386/w32117

Cao, J., & Leung, M. P. (2025). Neighborhood stability in double/debiased machine learning with dependent data. *arXiv preprint arXiv:2511.10995*.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal, 21*(1), C1–C68. https://doi.org/10.1111/ectj.12097

Chernozhukov, V., Demirer, M., Duflo, E., & Fernández-Val, I. (2019). Generic machine learning inference on heterogeneous treatment effects in randomized experiments. *arXiv preprint arXiv:1712.04802*.

Conley, T. G. (2008). Spatial econometrics. In S. N. Durlauf & L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics* (2nd ed., Vol. 7, pp. 741–747). Palgrave Macmillan.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics, 92*(1), 1–45. https://doi.org/10.1016/S0304-4076(98)00084-0

de Chaisemartin, C., & D'Haultfœuille, X. (2023). Difference-in-differences estimators of intertemporal treatment effects. *NBER Working Paper No. 29873*. https://doi.org/10.3386/w29873

de Chaisemartin, C., D'Haultfœuille, X., & Vazquez-Bare, G. (2024). Difference-in-difference estimators with continuous treatments and no stayers. *AEA Papers and Proceedings, 114*, 610–613. https://doi.org/10.1257/pandp.20241049

DellaVigna, S., Imbens, G., Kim, W., & Ritzwoller, D. M. (2025). *Using multiple outcomes to adjust standard errors for spatial correlation* (NBER Working Paper No. 33716). National Bureau of Economic Research. https://doi.org/10.3386/w33716

Donaldson, D., & Hornbeck, R. (2016). Railroads and American economic growth: A "market access" approach. *The Quarterly Journal of Economics, 131*(2), 799–858.

Egger, D., Haushofer, J., Miguel, E., Niehaus, P., & Walker, M. (2022). General equilibrium effects of cash transfers: Experimental evidence from Kenya. *Econometrica, 90*(6), 2603–2643. https://doi.org/10.3982/ECTA17945

Egger, D., Haushofer, J., Miguel, E., & Walker, M. (2024). "Amendment to: GE effects of cash transfers: Pre-analysis plan for Endline 3 child mortality analysis", August. AEA RCT Registry. https://www.socialscienceregistry.org/trials/505

Faridani, S., & Niehaus, P. (2024). Linear estimation of global average treatment effects. *NBER Working Paper No. 33319*. Revised January 2026. https://doi.org/10.3386/w33319

Fisher, R. A. (1935). *The design of experiments*. Oliver & Boyd.

Haushofer, J., et al. (2017). "Pre-analysis plan for household welfare analysis," July. AEA Trial Registry. https://www.socialscienceregistry.org/trials/505.

Haushofer, J., Niehaus, P., Paramo, C., Miguel, E., & Walker, M. (2025). Targeting impact versus deprivation. *American Economic Review, 115*(6), 1936–1974. https://doi.org/10.1257/aer.20221650

Kondylis, F., & Loeser, J. (2024). Cash transfer size and persistence. Working paper. https://johnloeser.github.io/assets/kl.pdf

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies, 76*(3), 1071–1102. https://doi.org/10.1111/j.1467-937X.2009.00536.x

Ludwig, J., & Mullainathan, S. (2023). Machine learning as a tool for hypothesis generation. *NBER Working Paper No. 31017*. https://doi.org/10.3386/w31017

Mchenga, M., Burger, R., & von Fintel, D. (2019). Examining the impact of WHO's Focused Antenatal Care policy on early access, underutilisation and quality of antenatal care services in Malawi: A retrospective study. *BMC Health Services Research*, 19(1), 295. https://doi.org/10.1186/s12913-019-4130-1

Miguel, E., & Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica, 72*(1), 159–217.

Moll, B. (2025). *The missing intercept problem when going from micro to macro* [Slides]. London School of Economics. http://benjaminmoll.com/micro_to_macro/

Müller, U. K., & Watson, M. W. (2023). Spatial correlation robust inference in linear regression and panel models. *Journal of Business & Economic Statistics, 41*(4), 1050–1064. https://doi.org/10.1080/07350015.2022.2127737

Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics, 235*(2), 2218–2244. https://doi.org/10.1016/j.jeconom.2023.03.008

Walker, M. W., Shankar, N., Miguel, E., Egger, D., & Killeen, G. (2025). Can Cash Transfers Save Lives? Evidence from a Large-Scale Experiment in Kenya. NBER Working Paper 34152 (2025), https://doi.org/10.3386/w34152.

Young, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics, 134*(2), 557–598. https://doi.org/10.1093/qje/qjy029