Micro-enterprise saturation: Critical mass or overcrowding?*

Moritz Poll†

July 23, 2026

Latest version here.

Abstract

Should entrepreneurship support programs to extremely poor people be scaled by saturating a few villages, or maintaining low saturation and enrolling more villages at a higher cost? On the one hand, saturation may create a *critical mass* of entrepreneurs whose income generation and associated spending multipliers make each other viable. On the other hand, encouraging the start-up of many small firms in close proximity may be a recipe for *overcrowding* and artificially fierce competition in a small village economy. In a large randomized control trial in Malawi, I exogenously vary the saturation of an entrepreneurship support program to assess whether business performance, increases or decreases in the share of treated neighbors. Sixteen months after the start of the intervention, earnings take a (imprecise) 20% hit as saturation doubles. About a third of the new retail businesses have gone out of business, constituting a serious scale barrier for such programs. Earnings losses and firm exits at high saturation are of a comparable magnitude to implementer economies of scale though. The heightened competition does not lower consumer prices, but induces out-migration. Most participants start a retail business. I show that this sectoral concentration is not driven by information frictions about other entrants' intended sectoral choice. Instead, respondents highlight a lack of hard skills, while childcare compatibility plays an important role as well.

Keywords: Entrepreneurship, poverty alleviation, general equilibrium, randomized control trial, saturation design, spillovers, coordination

JEL Classification: O12, O14, O25, D7, D51, D58

*I owe deep gratitude to the participants who dedicated their time to this study; to countless local stakeholders and practitioners who shared their wisdom and advice; and of course to the staff of my implementing partner without whose tireless work this project would not have been possible. I want to thank Andrew Foster, Peter Hull, and Stelios Michalopoulos for their support and guidance. This project greatly benefited from discussions with Clare Balboni, Vittorio Bassi, Dan Björkegren, Dennis Egger, Doug Gollin, Johannes Haushofer, Reshma Hussam, Fedor Iskhakov, Joe Kaboski, Michael King, Cynthia Kinnan, Jessica Leight, Sara Lowes, Elisa Macchi, David McKenzie, Josh Merfeld, Jonathan Morduch, Paul Novosad, Matt Pecenco, Tommaso Porzio, Simon Quinn, Jonathan Robinson, Jon Roth, Bryce Steinberg, Hyun Soo Suh, Nick Tsivanidis, Davide Viviano, Maarten Voors, and Chris Woodruff. Generous funding from the Weiss Fund (BN-002), the National Science Foundation (#2315009), the Dovetail Impact Foundation, the Abdul Latif Jameel Poverty Action Lab (J-PAL GEA-25-03001 & LAI-25-03010 from UK Government International Development funding), the Livelihood Impact Fund, the Gates Foundation (BMG-23-10013-X3 via Innovations for Poverty Action), as well as the Bravo Center, Population Studies and Training Center, Watson Institute, and Nelson Center at Brown University is gratefully acknowledged. I also benefited from fellowships funded by the Watson Institute, Brown University Graduate School, the Population Studies and Training Center (under T32 NICHD grant and PSTC P2C NIH grant), and the Barrett Hazeltine Fellowship for Graduate Research in Entrepreneurship. This study was pre-registered at https://www.socialscienceregistry.org/trials/11789 and approved by Brown University IRB (#98) and Malawi's NCRSH (NCST/RTT/2/6).

†University of Chicago, Department of Economics & DIL. moritzpoll@uchicago.edu. www.moritz-poll.com

1 Introduction

The success of a new business depends on the demand and competition from the surrounding economy. This matters particularly for the design of programs that promote the creation of new businesses as a an anti-poverty strategy. In the absence of stable employment opportunities, poverty alleviation efforts in many countries rely on individual entrepreneurship by poor people themselves.[1] Policies that support entrepreneurship in developing countries are growing in popularity and size, but economic theory does not offer a clear-cut prediction of their impact at scale. On the one hand, firms may linger near subsistence if they depend on the weak demand of other similarly poor people. Providing entrepreneurship support to more people in a given area might create a *critical mass* of entrepreneurs whose income generation and associated spending multipliers make each other viable. The logic of a big-push individual-level intervention (Balboni et al., 2022) would then carry over to the aggregate local economy (Rosenstein-Rodan, 1957). On the other hand, encouraging the start-up of many small firms in close proximity may be a recipe for *overcrowding* and artificially fierce competition in a small village economy.[2] This would suggest tight upper bounds to how far these programs can be scaled – and more broadly to the approach of alleviating poverty bottom-up from the household level (Goldberg and Reed, 2023). These two scenarios of critical mass versus overcrowding motivate my study. They imply contrasting policy prescriptions for whether entrepreneurship should be fostered by saturating a few locations or scaling over a larger geographic area at low saturation.

In a large randomized experiment in Malawi, I vary the program saturation of an entrepreneurship support training to assess whether business performance increases or decreases in the share of treated neighbors. A sample of 1,872 extremely poor mothers of young children in rural Malawi across 72 household clusters are randomly assigned to receive a bundled livelihood intervention that promotes the formation of micro-enterprises. It combines monthly cash transfers, productive assets or business start-up grants, entrepreneurship training, parenting training, formation of savings groups, and one-on-one life skills coaching. I experimentally vary program saturation at the household cluster level to be either 0%, 50% or 100% of eligible households,[3] corresponding to 0%, 18.5% and 37% of the population, respectively.

I document that new entrepreneurs do not take into account the impact that their own sector choice may have on the profitability of this sector, which likely contributes to sectoral concentration. This "general equilibrium neglect" is particularly problematic for a program design that encourages the entry of a large number of new firms into a small village economy. I develop and test a simple coordination intervention encouraging households to coordinate with each other when choosing

[1] See for example Blattman and Ralston (2015); Banerjee et al. (2025).

[2] Failure rates for micro-enterprises are high, especially soon after launch (De Mel et al., 2012; Balboni et al., 2022; Misha et al., 2019) and anecdotally many beneficiary households of the type of program studied here start similar businesses to one another (Bauchet et al., 2015). Hardy and Kagy (2020) identify gender-specific barriers of entry as a reason why female entrepreneurs are likely to enter less profitable sectors.

[3] Eligibility requires that a household have a mother of a child under the age of five at enrollment and live under the extreme poverty line.

their business venture, rather than duplicating existing businesses.

I show that pure information frictions cannot explain sectoral concentration. The coordination intervention did not succeed in shifting entrepreneurs-to-be out of the retail sector. While there are weak indications of sectoral shifts in the businesses that participants *plan* to pursue, as well as the livelihoods households report very shortly after the cash grant, these differences dissipate within just a few months. Most program participants end up operating retail businesses. I provide evidence that the lack of hard skills and the rigidity of childcare norms confines these entrepreneurs in the petty trading sector.

In the face of high sectoral concentration, increased saturation is damaging for business earnings, survival, and leads to out-migration. In terms of earnings, the businesses in high and low saturation clusters initially track each other's performance closely over time, but show signs of divergence. The cumulative earnings are not significantly different from one another, but the point estimate for the high-saturation group is about 20% lower than that of the low-saturation group after sixteen months. Importantly, the earnings results are somewhat imprecise and will be followed up with more data collection. Firms in the high saturation clusters are significantly more likely to exit, with households reverting to casual labor. The share of new retail businesses drops by about one third. An even starker indication of overcrowding is that the higher saturation induces out-migration from the study area.

An implementer may yet prefer to implement at high saturation as economies of scale in program cost appear larger than the earnings losses or business failure rates. In-depth budget reviews and simulations allow me to conclude that the cost savings at high saturation are about 30%. It is important to note that these results are preliminary and pending longer-run follow-up data collection which may change the direction and precision of results reported here.

I contribute to literatures on optimal program design, experimental spillover analysis, and firm entry. The program I study was designed to be a more flexible version of a well-studied anti-poverty intervention: Ultra-poor graduation programs (UPGs), aimed at promoting entrepreneurial activity by people in extreme poverty, are the closest that development economics has come to a silver bullet for sustainably alleviating extreme poverty (Sulaiman et al., 2016; Wydick, 2020). While in-kind, conditional and unconditional cash transfers have been shown to provide effective poverty relief and invigorate micro-enterprises in the short run (Sulaiman et al., 2016; Baird et al., 2011; Blattman et al., 2014; Haushofer and Shapiro, 2016; Cahyadi et al., 2020; Brooks et al., 2022; Egger et al., 2022), in-kind transfers or strong-handed conditionalities may induce distortionary effects (Bryan et al., 2021), and the evidence on long-term effects is mixed at best and favors settings where households are above the international extreme poverty line to begin with (Gertler et al., 2012; De Mel et al., 2012; Aizer et al., 2016; Ikegami et al., 2017; Haushofer and Shapiro, 2018; Araujo et al., 2019; Millán et al., 2020; Aggarwal et al., 2022; Altındağ and O'Connell, 2023; Fiala et al., 2022). Skills training, coaching, or access to loans or savings facilities by themselves show few if any impacts (Blattman and Ralston, 2015; Berge et al., 2015; Valdivia, 2015; Bauchet et al., 2015; Chowdhury et al., 2017; Brooks et al., 2018; Sedlmayr et al., 2020; Giné and Mansuri, 2021; Leight et al., 2021, 2022). On the other hand, UPGs following the BRAC model (studied here) seem to be standing the test of time while providing results that are larger than the sum of their parts in response to the highly complex challenge that multidimensional poverty traps pose (Sulaiman et al., 2016; Banerjee et al., 2015; Bandiera et al., 2017; Banerjee et al., 2021; Balboni et al., 2022; Chowdhury et al., 2017; Sedlmayr et al., 2020; Bedoya et al., 2019). If the first Sustainable Development Goal of ending extreme poverty by 2030 is to be achieved, the development community needs to better understand how UPGs scale. My study highlights that a key barrier to scale will be the access to a broader set of sectors through the provision of diverse hard skills – a challenging barrier considering that providing a diverse set of hard skills within the same location will erode economies of scale – and dependable market-provided childcare.

Spillovers are a common concern in poverty alleviation programming. While Egger et al. (2022) find almost no inflationary pressure from cash transfers in Western Kenya, Filmer et al. (2021) argue that this may be linked to relatively low saturation in the study area while in their own study they find important price increases when cash transfer saturation is high and markets remote and Cunha et al. (2019) find differential price effects by village wealth. In the context of business training interventions, McKenzie and Puerto (2021) find no evidence of negative side effects, while Drexler et al. (2014) show that in their context, at least part of the benefits to the treated come from business stealing from the control group. Such spillovers may also extend to programs that support new entrepreneurial activity, the focus of this study. I do not find impacts of program saturation on the price of the key commodity, maize. Spillovers from the treatment group onto the control group are likewise modest. But I show that when taking an intervention to scale, it is important to consider spillovers from treated onto other treated people.

The impact of firm entry on competition, mark-ups, profitability, and firm survival has a long tradition in economics. Unlike the classic study of the transition from monopoly to oligopoly through the entry of a small number of firms (Bresnahan and Reiss, 1991), I investigate why firms continue to enter an already saturated sector. The firms in question are substantially less professionalized than those in the otherwise closely related study by Bergquist and Dinerstein (2020) on intermediate traders in Kenya. My findings echo and add to conclusions drawn by Hardy and Kagy (2020) that barriers to other sectors have in part a gender dimension and are additionally a function of skill barriers.

This paper proceeds as follows. Section 2 describes the setting, research design, intervention, and empirical strategy. Results can be found in Section 3 and Section 4 concludes.

2 Experimental design

This section describes the study setting, sample selection, experimental design, intervention, and emprirical strategy. Extensive details on sampling, treatment assignment, intervention, data collection, and analysis can be found in an online Implementation Appendix.[4] Ethical protections for the vulnerable households in this study are discussed in a Structured Ethics Appendix, following Asiedu et al. (2021). Below, I focus on the details most pertinent to this paper.

2.1 Setting and census

This study takes place in Mangochi district in the South of Malawi, one of the poorest districts in the country. To determine program eligibility and convey an accurate picture of the local economy, I conducted a household census of all 28,894 households in the area, covering a population of almost 140,000 individuals (Appendix Table A1). The majority of the population are subsistence or pre-commercial farmers (Appendix Figure A1). Sectoral concentration outside of agriculture is stark as well: While the study area has 50 welders, 235 carpenters, and 297 tailors – all professions that may come to mind when envisioning the types of micro-enterprises a poor person may engage in – there are 5,631 retailers, roughly one in five households. The fact that the overwhelming majority of participants in the implementing partner's prior cohort started retail businesses[5] motivated the research questions surrounding sectoral choice and treatment effects at higher saturation. Retail has an outsized and perhaps under-appreciated role in the process of structural transformation as the first non-agricultural sector that many poor people enter and features prominently in the success of several seminal livelihood interventions (De Mel et al., 2008; Banerjee et al., 2015; Blattman et al., 2016; Bandiera et al., 2017).

2.2 Sample selection

The sampling process is described in Appendix Figure A2. Using the Malawi-validated Simple Poverty Score Card (Schreiner, 2019) and self-reported income, I identified about 40% of households as ultra-poor and eligible.[6] Using k-means geographic clustering, these roughly 11,000 households were assigned to 271 clusters. Due to how households tend to live nearby one another, clusters often roughly correspond to villages when these are small and remote. In larger towns, villages are usually comprised of several clusters or a cluster can combine households from more than one village. To determine the sample, I deployed a custom algorithm to select 72 clusters into the study in such a way as to maximize the distance between clusters and minimize the potential for spillovers with buffer clusters in between that are not part of the study.[7] The median household

[4] Readers may also refer to this appendix for a discussion of the relationship between this and a number of companion papers that focus on the livelihood aspects of this study.

[5] This phenomenon is not exclusive to the implementing partner I work with but has likewise been documented in a similar program by a large humanitarian implementer in Malawi to which I was given access.

[6] The implementing partner excluded households from participation that have no female; have no children under the age of 5 as of August 1, 2023; did not appear ultra-poor based on either their stated income or the Simple Poverty Scorecard; in which the female was unable to carry out physical labor; whose data appeared implausible to the research team or enumerator; or who were not willing to be registered or could not be found during several (mop up) visits. The latter two criteria were very rare.

[7] After cluster selection and baseline survey, but before the start of the intervention, the local government designated one of our clusters as a flood-prone area and ordered it to be vacated. As our study may have encouraged people to stay in an unsafe area, we shifted to a neighboring cluster in this one instance.

cluster is about 30 hectares in size and is home to about 70 households. The minimum distance between study clusters corresponds to the distance beyond which Egger et al. (2022) no longer find spillover effects of cash transfers.[8] In my analysis, I will therefore allow for spillovers within but not across clusters. Though this assumption may not exactly hold in practice (for example information is likely to "travel further" than cash), the assumption is that spillovers are substantially greater between neighbors within a cluster than they are across clusters. Clusters were grouped into sets of three that were assigned to a program implementer as well as being randomly assigned to one of three cohorts to be launched a few months apart from each other (early 2024, mid 2024, mid 2025, see Appendix Figure A3).

2.3 Treatment assignment

Saturation Within each implementer's area (stratum), the three clusters were assigned at random to either high treatment saturation (everybody treated), low treatment saturation (half treated, half spillover control group, randomly selected at the individual level) or pure control group (nobody treated). The saturation levels were chosen following Baird et al. (2018).[9] High-saturation clusters, where all 26 households (100% of eligible households) are treated, correspond to a saturation level among the broader population surveyed during census of about 37% in the median cluster. In low-saturation clusters, 13 randomly chosen households (50% of the eligibles) are treated, corresponding to 18.5% of all households in the median cluster. Assigning each implementer one of each type of cluster is tantamount to geographic and implementer stratification (Bruhn and McKenzie, 2009; Bai, 2022).

Economies of scale and cost benchmarking The appeal to an implementer of conducting an intervention at higher saturation is that fixed costs can be spread across more beneficiaries. My intimate familiarity with the implementing partner's budgets and cost structure allows me to quantify that the per-household cost of the program would be about 30% lower if the program were conducted with 26 households per village as compared to 13 households per village. The major cost driver is the number of implementing staff and the distances they need to travel.[10]

This means that there are two natural benchmarks for program performance as saturation increases. One is to ask, whether more or fewer businesses become successful relative to how many were supported (and how well they perform) as more people are treated. Comparing the high and low-saturation clusters provides this comparison. But in practice, some losses in program

[8] If anything, my study area's road network is substantially less developed and personal ownership of bicycles or even motorbikes is lower in Southern Malawi than in Western Kenya and the terrain is sliced up by seasonal rivers that make effective distances larger than straight-line distance.

[9] The study is designed to be well powered for spillovers among treated households. Modifying equation (6) in Baird et al. (2018) to omit the standard error of the spillover effect on the control group, I restate their Theorem 2 to conclude that the optimal high saturation is 100% and the optimal low saturation would have been below 30% and dependent on intra-cluster correlation of outcomes. However, in conversation with the implementing partner, we decided to increase the lower saturation to 50% which would represent a more policy-relevant comparison. Less saturated clusters would also make program components like savings groups difficult to implement.

[10] More detailed program and component costing information can be found in Foster et al. (2026).

effectiveness are permissible considering the important economies scale in implementation cost. In addition to assessing whether there are *any* adverse saturation effects (which will be of interest to academic economists), I therefore also determine how they relate to the cost savings (which will be of interest to policy makers and practitioners) when the goal is to maximize the number of beneficiaries with viable businesses *given* a fixed budget.

Coordination intervention and other research design elements In each of the cluster triplets, either the high- or the low-saturation treatment cluster is randomly selected for an additional coordination intervention described below. In addition, each stratum was randomly assigned to one of three cohorts to be launched a few months apart from each other (see Appendix Figures A3 and A4). The cohorts are timed such that one receives the large cash grant during the planting season, one during the harvest season, and one during the agricultural lean ("hungry") season. The optimal seasonal timing of this program is studied in Poll (2026b).

Stratification and sample balance In order to balance the sample on a number of key outcome and mechanism variables (see Table B1), I first run the treatment assignment algorithm 100,000 times (seeding each with the counter of its iteration). I then estimate baseline balance under each theoretical treatment assignment and discard assignments that lead to a statistically significant imbalance in any of the key mechanism variables at the 10% significance threshold (Morgan and Rubin, 2012). In other words, I enforce a "visually clean" balance table. Treatment is assigned separately for each of the three cohorts. I then randomly pick one of the permissible treatment assignments. As pointed out in both Bruhn and McKenzie (2009) and Morgan and Rubin (2012), classical inference that does not take this rerandomization into account would now on average (though not in every single case) be too conservative, leading to confidence intervals on the average treatment effect with overly wide coverage and too few rejections of null hypotheses. A Fisher (1935) exact test has the correct coverage, though it tests the sharp "randomization hypothesis" that the treatment had no effect on anyone rather than on average (Ritzwoller et al., 2025). Traditionally, researchers using rerandomization methods have done so somewhat *ad hoc*, rerandomizing until the balance table was satisfactory. My approach is more transparent. Given $R = 100,000$ simulated assignments and $k = 10$ pre-specified baseline balance table variables, the likelihood of discarding an assignment at the 10% significance level is $q = 1 - (1 - 0.1)^k$ and I can therefore use the remaining $R \times (1 - q) \approx 34,000$ simulated assignments for the randomization inference (Young, 2019).

2.4 Intervention

The Childhoods and Livelihoods Program The intervention is a standard BRAC-style UPG (Banerjee et al., 2015; Bandiera et al., 2017) with two key modifications. First, instead of livestock and related training, households receive a one-off labeled[11] cash grant of $300 intended for starting a business, corresponding to about $1,100 PPP.[12] Flexibility is given to participants to leverage their own informational advantage regarding their abilities and preferences, and they receive four months of weekly general-purpose business training covering financial literacy, market analysis, budgeting, recordkeeping, funding sources, and the development of business ideas. The structure of cash transfers is closely aligned with evidence on how poor households themselves prefer to receive such funds, combining a regular buffer with a large lump sum that arrives after sufficient time for deliberation (Kansikas et al., 2023). During the first four months, program participants are also enrolled in an early childhood development group training, the second notable modification of the UPG.[13] For the duration of the entire first year, households in the treatment group also receive monthly one-on-one home visits with the goal of coaching the household on preventive health behaviors and other ways of avoiding or coping with adverse shocks. Treated households can also use these visits to ask questions about training contents or their businesses. Both the treatment and control group receive a basic phone for scheduling of surveys and for receipt of cash transfers or survey incentives through mobile money. At the community level, the formation of savings and loan groups is encouraged. Local childcare centers are supported financially and with the same ECD training. Both treatment and control group are expected to benefit from this last aspect. The sequencing of program components is described in Figure 1.

Size of the intervention The seed capital and consumption support combined amount to $420 (over $1,500 PPP), three times larger than the poverty trap threshold estimated in Balboni et al. (2022), equivalent to the large grants in Haushofer and Shapiro (2016) and marginally lower than those in Egger et al. (2022). To benchmark the size of the intervention, I estimate the regional GDP of the study area using my census data. I use self-reported annual income and an estimate of the value of non-marketed subsistence agriculture. While income estimates in a survey are likely to be underreported if respondents understand that the data will be used for targeting an assistance program, Appendix Table A3 shows that the poverty rate estimated from these self-reports aligns closely with that from a proxy means test validated for Malawi by Schreiner (2019). Appendix Table A4 documents the GDP estimates for the study area. Column (2) shows local production at the exchange rate at which the cash transfers in this study were converted from USD to MWK. The combined cash transfers per treatment household correspond to approximately one year's GDP per capita for the average five-person household in the study area and (because of negative selection into the sample) about 20 months of expenditure of a control group household. My study injects

[11] The grant is largely unconditional, but the NGO encourages that it be used for starting a business or supporting an existing one. Receipt requires attending the training and being able to adequately verbalize a concrete business plan, but participants can miss sessions without penalty and are coached until their plan is satisfactory and the actual implementation of said plan is not afterwards enforced.

[12] The $300 were converted at the official exchange rate to MWK 530,000 and the World Bank estimates the 2024 PPP conversion factor to be 475 MWK/USD-PPP. They are paid out in a major tranche of 75% at the six months mark and a top-up grant of 25% eight months later.

[13] A detailed analysis of the impacts of the program on young children and the trade-offs between maternal time and material resources can be found in Foster et al. (2026).

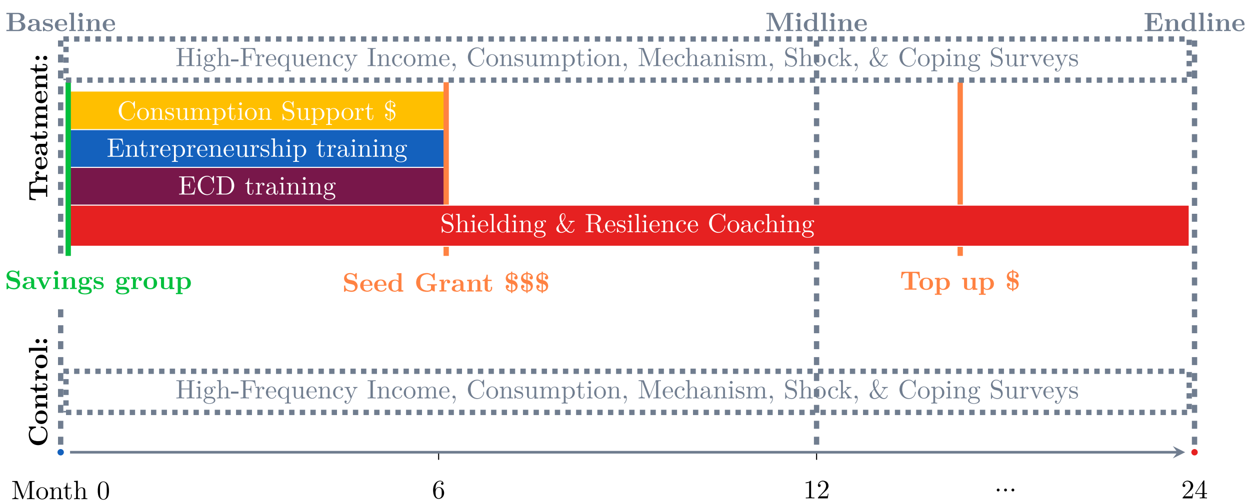

Figure 1: Program Components and Data Collection

Notes: This figure describes the program component sequencing and timing relative to data collection.

around 3.3% of one year's local marketed GDP into the study area in cash transfers alone, though this estimate falls to 1.9% when including an estimate of subsistence production in my measure of GDP.[14] This is a sizable amount that may well produce consumption spillovers that translate into increasing returns to saturation.

Coordination intervention The direction of the saturation impacts is likely to depend on the sectoral choices of the entrants. When asked why they would like to pursue a particular business idea, the majority of participants mentioned that they had seen other people earn good money this way (Figure 2). For an individual prospective entrepreneur, imitation is of course a potentially sensible strategy that steers entrants into sectors where mark-ups might be higher and that are attainable to someone "like them". But the fact that this is by far the predominant motive, points to a potential neglect of general equilibrium effects that may be detrimental when not one but a dozen or two new businesses are being launched. This makes sectoral concentration a key concern for the scalability of livelihood interventions like the UPG.

In half of the treated clusters, implementers do not guide which kinds of businesses participants start, while the other clusters receive a business selection guidance treatment. In these clusters, participants first listen to a story of three women in a similar position as them, having participated in an entrepreneurship training program and ready to receive their cash grant when they discover that they all three had the same business idea. The women in the story come to the realization that there would probably not be enough demand for all three of them to pursue the same business and they instead choose different businesses. After reading the story, the implementers ask participants to share with each other what kinds of businesses they would like to start (as a ranked preference list). Participants are then asked to consider what would happen to their profitability if they all

[14] This estimate does not include survey incentives, local procurement, or spending by study staff.

started similar businesses and encouraged (but not forced) to coordinate so that they do not all start the same businesses as one another. The goal is to first widen participants' pool of potential business ideas and second to prime them to think about the general equilibrium consequences of their business choice.

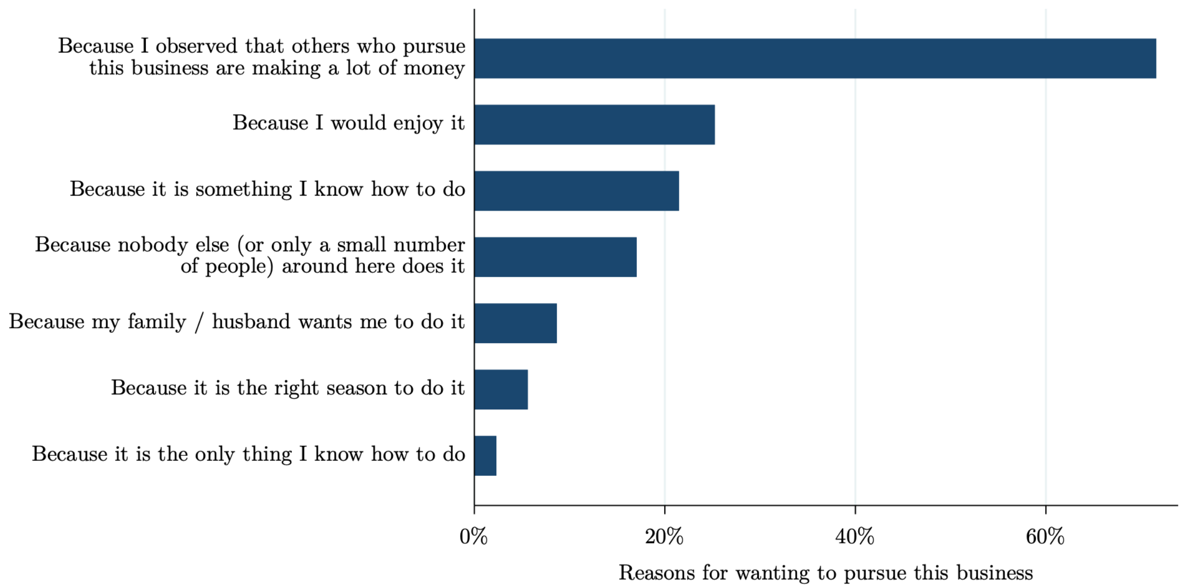

Figure 2: Evidence of general equilibrium neglect

Notes: This figure summarizes common motivations that participants state for why they would like to pursue a given business idea. The overwhelming majority expresses what I term general equilibrium neglect: That they would like to emulate the profits of others, perhaps without internalizing the effects of their own entry. I hypothesize that this will have detrimental consequences in a high-saturation implementation and it motivates devising a business coordination intervention.

2.5 Data collection

A team of external enumerators conducted three rounds of main-wave household surveys for the purpose of program evaluation: As Appendix Figures A2 indicates, each cohort started with a baseline survey to all 624 households. A midline survey followed approximately four months after the primary cash grant tranche (about 12 months after the start of the intervention), and an endline 12 months after midline and about 24 months after the start of the intervention (Appendix Figure A4). Main-wave surveys cover a wide range of topics, spanning a household roster and information on livelihoods, health, dwellings, attitudes, and child development outcomes. In addition to these main-wave surveys, all households in the study, both treatment and pure control group, receive monitoring visits from program implementers every two to three weeks.[15] At these visits, implementing staff conduct short surveys about consumption and investment choices, livelihood

[15]The spillover control group is not visited at high frequency, only for the main wave surveys.

activities, business performance, as well as shocks, coping, and preventive behavior.[16] The treatment group received their one-on-one coaching during these monitoring visits. At the end of the business trainings and before distributing the main tranche of the business grants, treatment group participants filed a business plan for the livelihood they were planning to support with the grant. These plans were digitized by program implementers and are available as outcome data as well, for treated households only, and allow us to measure right at the end of the trainings whether the coordination intervention impacted the proposed businesses.

2.6 Empirical strategy

The research design maps into two estimating equations which in turn answer my research questions: The randomly added coordination intervention allows me to test whether sectoral concentration can be fully explained by the general equilibrium neglect for which I provided evidence above. I leverage the exogenous variation between high and low saturation clusters to answer whether increased saturation leads to critical mass and spending multipliers through which the budding entrepreneurs make each other viable or to excess competition and overcrowding of small village economies.

As pre-specified (Poll, 2023),[17] I estimate the following linear regressions:[18]

$$\begin{aligned} y_i =& \beta_0 + \beta_T \text{Treatment}_i + \beta_C \text{Coordination Intervention}_i \\ & + \beta_{\text{BL}} y_i^{\text{BL}} + \beta_{\text{M}} \mathbb{1}[y_i^{\text{BL}} \text{ missing}] + \beta_{\text{FE}} S_i + \varepsilon_i \end{aligned} \tag{1}$$

$$\begin{aligned} y_i =& \beta_0 + \beta_T \text{Treatment} + \beta_H \text{High saturation} \\ & + \beta_{\text{BL}} y_i^{\text{BL}} + \beta_{\text{M}} \mathbb{1}[y_i^{\text{BL}} \text{ missing}] + \beta_{\text{FE}} S_i + \varepsilon_i \end{aligned} \tag{2}$$

Equation (1) investigates whether the addition of the coordination information intervention alters sector choices among treated households. Equation (2) analyzes the saturation impacts of the program. Standard errors are clustered at the strata (cluster triplet) level (De Chaisemartin and Ramirez-Cuellar, 2024) which is the level of randomization and I am controlling for baseline levels of all outcome variables (ANCOVA, y_i^{BL} and a missing indicator when a missing baseline value has been set to zero) as well as strata fixed effects S_i. I investigate three key outcome variables, sector choice, business survival, and earnings.

The impact of the coordination intervention is captured by β_C. The sign of β_H in equation (2) is the main parameter of interest in this study, theoretically ambiguous, and merits discussion. If it is relatively close to zero for both outcomes, there would be no evidence of spillovers among

[16] Poll (2026a) investigates what these unique high-frequency data can tell us about the mechanisms at play behind the success of the UPG intervention.

[17] This study was pre-registered at https://www.socialscienceregistry.org/trials/11789.

[18] The pre-analysis plan additionally contained an interaction of the saturation level and coordination intervention. Since the latter turned out to be ineffective, I concentrate the statistical power on the simple comparisons instead, which should therefore be interpreted as weighted averages of the treatment effect of the uninteracted treatment effects (Muralidharan et al., 2023).

the treated for business performance. This would indicate room in the local economy for more firms. If the coefficient is positive for either outcome, this would speak in favor of the spending multipliers and critical mass hypothesis. If either one or both of the coefficients are negative, this would support overcrowding. If one were positive and the other negative (e.g. more surviving firms but with lower profits), then the local economy may be at capacity, but the constraint operates only along one margin.

3 Results

3.1 Balance

In Appendix Table B1, I show sample balance. As discussed above, the clean balance table follows mechanically from the randomization approach. In addition to the pre-specified set of variables, about 18% of households in the sample were engaged in retail at baseline (in line with the entire local population, cf. Appendix Figure A2) and women score roughly 20 out of 100 points on an entrepreneurship knowledge test that covers the key lessons of the business training (both balanced). Attrition at the one-year follow-up (midline) survey is 2.3% at midline and 9.6% at endline, reflecting an overall good retention.

3.2 Coordination and sectoral concentration

The coordination lesson did not lead to meaningful business diversification. Starting with the business plans that treated households submitted Column (1) in Panel A of Table 1 shows that over 80% of participants proposed to do a retail business. The coefficient on the coordination intervention is negative but small and statistically not significantly different from zero. Somewhat promisingly, the coordination lesson appears to have marginally compelled program participants to *propose* non-retail service businesses (such as restaurants or beauty salons) ahead of receiving the cash grant, as shown in Column (2). About 20% of participants propose a business that is unique within their treatment cluster (not statistically different for the coordination group) and about 30% propose an income source that is new to their village, i.e. that nobody there was recorded to be doing during the census. The coordination increases this to almost 40%.

At the midline survey, about four months after the cash grant, the *actual* sectors that households are operating in, however, seem largely unaffected by the coordination intervention. The livelihoods of all household members are assessed for both the control group and the treatment arms with and without the coordination intervention. Panel B of Table 1 shows that the overall graduation program reduced the number of households relying on casual labor and primary sector activities in favor of retail and (to a lesser extent) non-retail service sector livelihoods, as can be seen in Column (2). Among the treated households, however, the coordination intervention did not meaningfully affect actual livelihood choices (Column 3). If anything, there is a slight but significant decrease in the number of manufacturing sector livelihoods, which was not an intended consequence of the intervention. It also did not increase the number of unique or new businesses in a cluster. At endline,[19] about 16 months after the cash grant, Table 1 Panel C shows that any short-term impacts of the coordination intervention have faded.

3.3 Barriers to diversification

Contextual knowledge and focus group discussions highlight two important barriers to entering non-retail sectors: Childcare and the lack of hard skills. All program recipients are mothers of young children and childcare norms in Malawi place this burden squarely on women. Consequently, female livelihoods need to be compatible with childcare (Brown, 1970), i.e. near or in the homestead with few or no trips, not requiring constant focus, and safe enough to bring a young child along. The retail sector is attractive in this sense. Formal childcare is available in many villages and the program actively and successfully encourages its use while working (see Appendix Table B2) but most do not offer meals and only open for a few hours in the morning. Building out high-quality full-day childcare options would be a major endeavor.

A lack of hard skills is the second major reason that women do not venture further afield than petty trading. Appendix Figure A5 illustrates this: Fewer than 20% of the sample indicate proficiency in anything that is not either retail, farming, or a (marketable) home production skill (like cooking or baking). This too poses an important scale challenge to livelihood programs aimed at the very poorest: The program under study here sought to give participants maximum flexibility to start a wide range of businesses by providing a cash grant and general-purpose business skills *en lieu* of a productive asset or livestock and a training on how to use it since that would tie recipients to a particular sector. If the results of this study are any guide, providing a single hard skill as part of the program would run into the same issue of sectoral concentration. But providing a diverse array of hard skills is difficult and costly in practice as it negates the scale economies that promised to make these programs more affordable in the first place.

It appears therefore as though the sectoral concentration in retail is not simply the product of an information friction that the coordination intervention could alleviate. Even though participants appeared to display general equilibrium neglect, correcting it did not change their sector choices. Since the coordination intervention did not affect business choices meaningfully, I pool their results going forward to compare high and low saturation clusters.

[19] Endline results are preliminary and pending cohorts 2 and 3.

Table 1: Proposed and realized sectoral choice

	Panel A: Business plans			**Panel B:** Midline				**Panel C:** Endline			
	(1) Treatment mean (SD)	(2) Coordination Intervention	(3) N (BL)	(4) Control mean (SD)	(5) Treatment Effect	(6) Coordination Intervention	(7) N (BL)	(8) Control mean (SD)	(9) Treatment Effect	(10) Coordination Intervention	(11) N (BL)
Casual labor, charity	0.000	0.000	910	0.792	−0.180***	−0.006	1829	0.784	−0.323***	0.026	565
	(0.000)	(0.000)	✓	(0.406)	(0.032)	(0.030)	✓	(0.412)	(0.054)	(0.039)	✓
		[0.000]			[0.000]	[0.841]			[0.000]	[0.530]	
Agriculture, Fishing, Forestry	0.073	−0.009	910	0.628	−0.044	−0.046	1829	0.688	−0.081	0.018	565
	(0.261)	(0.016)	✓	(0.484)	(0.036)	(0.036)	✓	(0.464)	(0.043)	(0.047)	✓
		[0.549]			[0.223]	[0.203]			[0.113]	[0.704]	
Manufacturing	0.034	−0.001	910	0.140	0.023	−0.033**	1829	0.167	−0.017	−0.007	565
	(0.181)	(0.017)	✓	(0.347)	(0.019)	(0.016)	✓	(0.373)	(0.015)	(0.032)	✓
		[0.962]			[0.247]	[0.048]			[0.318]	[0.837]	
Non-retail Services	0.048	0.035*	910	0.150	0.120***	0.004	1829	0.238	0.099*	0.010	565
	(0.214)	(0.018)	✓	(0.357)	(0.019)	(0.021)	✓	(0.426)	(0.048)	(0.037)	✓
		[0.071]			[0.000]	[0.843]			[0.073]	[0.792]	
Retail	0.824	−0.015	910	0.230	0.432***	0.033	1829	0.270	0.386***	−0.053	565
	(0.381)	(0.028)	✓	(0.421)	(0.033)	(0.032)	✓	(0.444)	(0.040)	(0.050)	✓
		[0.599]			[0.000]	[0.315]			[0.000]	[0.314]	
Unique income source	0.195	0.042	910	0.239	0.073**	−0.011	1820	0.310	0.027	0.029	560
	(0.396)	(0.043)	×	(0.426)	(0.028)	(0.026)	×	(0.463)	(0.042)	(0.041)	×
		[0.365]			[0.013]	[0.666]			[0.532]	[0.497]	
New income source	0.287	0.096*	910	0.121	0.154***	0.068	1820	0.189	0.151**	0.021	560
	(0.453)	(0.052)	×	(0.326)	(0.029)	(0.042)	×	(0.392)	(0.058)	(0.062)	×
		[0.087]			[0.000]	[0.119]			[0.053]	[0.747]	

Notes: Analysis in Panel A is among treated households (who wrote a business plan) and the treatment coefficient is therefore omitted. Panels B and C include the control group and are assessed at the main-wave surveys. The last column of each panel indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.4 Saturation and business performance

Turning to the performance of the businesses, I compare high-saturation clusters to two benchmarks, as described in section 2.3. First, I compare them to low-saturation clusters to investigate whether program impacts scale linearly or display positive or negative spillovers that intensify with saturation. Second, I compare performance to a cost benchmark, since implementation at higher saturation saves the implementing partner about 30% of their per-household costs. I use three main metrics: Earnings, business survival, and prices.

Earnings Earnings take some time to react to the saturation. Figure 3 shows midline and endline earnings for the three types of clusters. At midline, four months after the major cash grant tranche, both low and high saturation cluters earn more than double what the control group earns. The earnings of the high-saturation group are on par with those of the low-saturation group and statistically significantly higher than the 30% cost savings benchmark. By endline, however high-saturation earnings treatment effects have sagged by about 20% compared to their low-saturation counterparts.[20]

Business survival and migration While earnings are dampened only somewhat by higher saturation, the impact on business survival is starker. As Table 2 shows, by the time of the endline survey (about 16 months after the cash grant), a large share of low saturation households are still operating a retail business compared to the control group (+49pp). But in the high-saturation clusters, this number is lower by about one third (-18pp). At the same time, high-saturation cluster households are significantly more likely to have reverted to the casual agricultural labor they were previously engaged in.

Perhaps the starkest indication of competitive pressure comes in the form of households' migration decisions. Table 3 demonstrates that while the treatment broadly increased short-run mobility of the households (Panels A and B) without differential impacts by saturation, the dynamics are more complex for permanent migration: At low saturation, the UPG decreases both the out-migration of individual household members and entire households. At high saturation, on the other hand, it increases it.

Prices With more firms and slightly lower earnings per firm, it seems reasonable to ask whether the business environment in the treated areas has become more competitive, entailing consumer welfare gains. I track maize prices from August 2024 onwards in the entire study area. As Malawi's foremost staple crop, maize has a pivotal role as a barometer for price levels more broadly and many of the studied businesses engage in retail of food crops and may exert downward pressure on maize prices. At the same time, it should be noted that the importance of maize makes it heavily traded, potentially limiting the scope that this intervention had. Figure 4 suggests that maize

[20] These results are preliminary and pending endline data collection in cohorts 2 and 3, which is also the reason for the relatively wide endline confidence intervals.

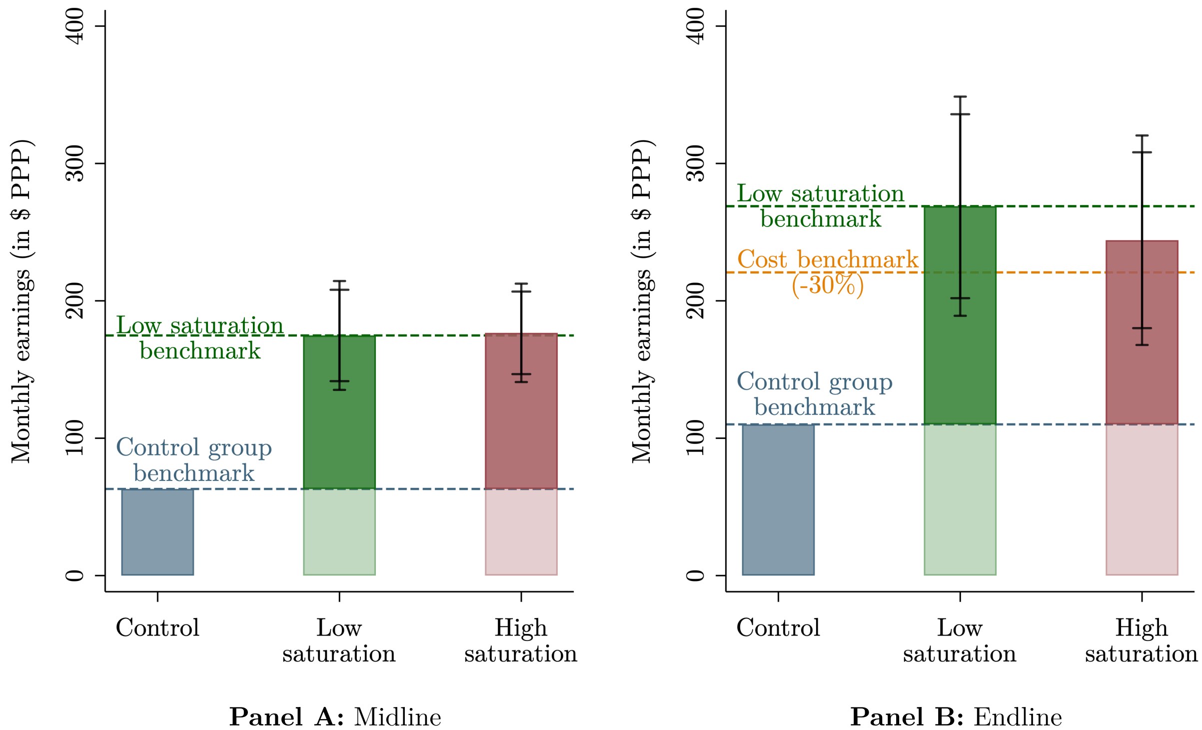

Figure 3: Earnings by saturation

Note: This graph shows monthly earnings at the time of the endline survey. The high-saturation cluster earnings are compared to three different benchmarks: The control group (blue), the low-saturation group (green), and the same benchmark discounted by the cost savings in a high-saturation program design. The spikes indicate 90 and 95% confidence intervals.

Table 2: Business survival

	(1) Control mean (SD)	(2) Treatment Effect	(3) High Saturation	(4) N (BL)
Casual labor, charity	0.784	-0.364^{***}	0.075^{**}	565
	(0.412)	(0.047)	(0.021)	✓
		[0.000]	[0.013]	
Farming, fishing, forestry	0.688	-0.084^{*}	0.015	565
	(0.464)	(0.042)	(0.052)	✓
		[0.086]	[0.779]	
Manufacturing	0.167	-0.062	0.066	565
	(0.373)	(0.044)	(0.051)	✓
		[0.199]	[0.236]	
Non-retail services	0.238	0.081^{*}	0.032	565
	(0.426)	(0.035)	(0.040)	✓
		[0.054]	[0.435]	
Retail	0.270	0.486^{***}	-0.177^{**}	565
	(0.444)	(0.048)	(0.056)	✓
		[0.000]	[0.020]	
Unique income source	0.310	0.146^{*}	-0.166^{**}	560
	(0.463)	(0.065)	(0.050)	×
		[0.057]	[0.017]	
New income source	0.189	0.178^{**}	-0.030	560
	(0.392)	(0.067)	(0.086)	×
		[0.038]	[0.735]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Mobility

	(1) Control mean (SD)	(2) Treatment Effect	(3) High Saturation	(4) N (BL)
Panel A: Day-to-day mobility				
Sell at local market	0.330	0.199***	-0.042	1829
	(0.470)	(0.039)	(0.052)	✓
		[0.000]	[0.437]	
Sell elsewhere	0.290	0.239***	0.027	1829
	(0.454)	(0.044)	(0.044)	✓
		[0.000]	[0.541]	
Minutes to market	54.3	6.017	1.324	1828
	(51.9)	(7.847)	(10)	✓
		[0.474]	[0.903]	
Panel B: Time-to-time travel				
Times left district	0.466	1.108*	-0.397	1829
	(2.065)	(0.545)	(0.587)	✓
		[0.079]	[0.544]	
Traveled to Lilongwe	0.080	0.075	0.109	1829
	(0.799)	(0.065)	(0.120)	✓
		[0.337]	[0.447]	
∼ Blantyre	0.106	-0.069	0.072*	1829
	(1.132)	(0.041)	(0.041)	✓
		[0.123]	[0.065]	
∼ Mangochi town	0.415	0.797***	0.016	1829
	(1.628)	(0.238)	(0.348)	✓
		[0.001]	[0.974]	
Panel C: Permanent migration				
Adults left HH	0.235	-0.003	0.088**	1829
	(0.581)	(0.042)	(0.040)	✓
		[0.940]	[0.037]	
HH left study area	0.050	-0.037***	0.022***	1872
	(0.219)	(0.013)	(0.007)	×
		[0.009]	[0.003]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

prices track each other through heavy seasonal fluctuations in all cluster types. Higher saturation does not seem to be a driver of more competitive retail markets for maize.

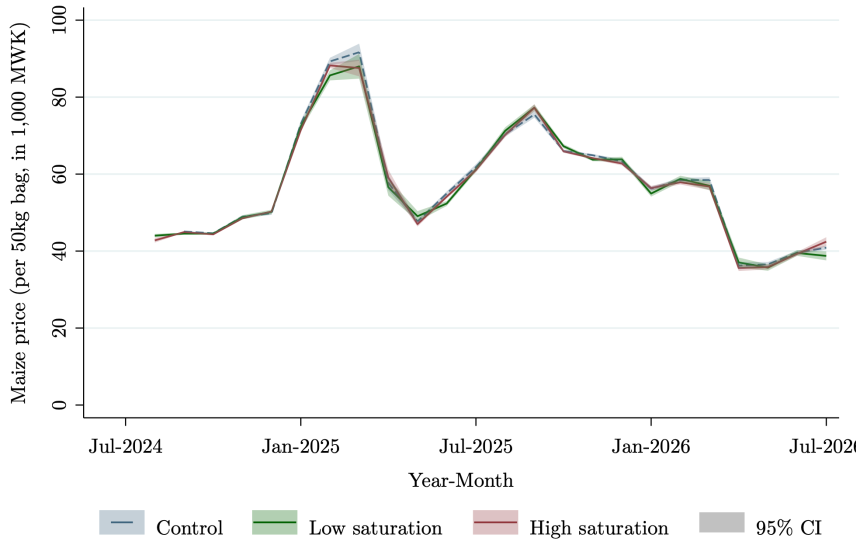

Figure 4: Maize price by saturation

Note: The price of maize, the dominant staple crop in the region, follows strong seasonal fluctuations, but does not appear affected by the high- or low-saturation treatment.

3.5 Robustness

The section discusses a number of robustness checks. Lower profits are not a mechanical byproduct of adverse selection into business as the 13 households selected into treatment among the 26 households that make up a low saturation cluster are chosen at random and Table B1 shows balance on ability and prior experience.

It does not appear as though the larger cluster size deteriorated training quality significantly. Appendix Table B3 shows that the high-saturation clusters score consistently (but not statistically significantly) lower than their low-saturation counterparts, though only by less than 1 point out of 100. I also verify whether the coordination intervention is reflected in the test scores by investigating whether coordination-related themes are mentioned by program participants. Those who received the coordination intervention were indeed more likely to mention being aware of one's competitor's weaknesses as factors that contribute to business success, though they are no more likely to say that one should find one's own niche and there is no difference in whether they highlight the presence of competitors (Appendix Table B4).

4 Discussion

A broad but perhaps under-appreciated pattern in structural transformation is that the retail sector is the first livelihood outside of agriculture for many extremely poor people. This matters for anti-poverty programs that target subsistence farmers. My paper shows that while recipients express general equilibrium neglect in how they chose their businesses, this cannot have been the only barrier to choosing livelihoods outside of the retail sector.

The high sectoral concentration has consequences for the effectiveness of the treatment. There are first signs of a divergence in earnings a few months after the cash grant, which come into sharper relief when focusing on business survival and exit. Despite the cash inflow being comparable in size to Egger et al. (2022) in Kenya who document consumption multiplier effects, in this setting and for an intervention focused on business creation, the overcrowding forces seem to dominate. A particularly stark sign of overcrowding is that the higher saturation arm induces significant out-migration. Prices do not appear to be systematically affected by saturation, suggesting that increased demand and supply either cancel each other out or the lower earnings come about by spreading a limited number of sales at fixed prices among more entrepreneurs who are already at a competitive benchmark.

The lack of hard skills and childcare compatibility emerge as likely next bottlenecks for broader sectoral diversification. Both of these are formidable barriers: Building out high-quality childcare infrastructure and changing traditional gender norms around childcare provision would require substantial investment and time. The inclusion of hard skills trainings in UPG and Cash+ interventions comes with important challenges as well: These programs are already some of the most costly per-household interventions in existence. Providing all participants in a village the same hard skill would be straightforward but likely self-defeating for the same sectoral concentration reasons I highlight here. Providing differentiated skills, in turn, would require innovation in hard skills training delivery and may not attain scale economies. Apprenticeship programs are an avenue worth exploring in future research.

Despite the evidence for overcrowding I present here, there are good reasons for implementers to pursue higher-saturation programming: The cost savings for the implementer from scale economies outweight (the point estimates on) earnings reductions and exits throughout. Out-migration, while potentially a symptom of overcrowding, is also in itself not a bad thing if households use the cash injection to migrate somewhere where they can be more productive.

References

Aggarwal, Shilpa, Jenny Aker, Dahyeon Jeong, Naresh Kumar, David Sungho Park, Jonathan Robinson, and Alan Spearot, "The dynamic effects of cash transfers: Evidence from Rural Liberia and Malawi," 2022.

Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney, "The long-run impact of cash transfers to poor families," *American Economic Review*, 2016, *106* (4), 935–71.

Altındağ, Onur and Stephen D O'Connell, "The short-lived effects of unconditional cash transfers to refugees," *Journal of Development Economics*, 2023, *160*, 102942.

Araujo, Maria Caridad, Mariano Bosch, and Norbert Schady, "Can Cash Transfers Help Households Escape an Inter-Generational Poverty Trap?," *The Economics of Poverty Traps*, 2019, pp. 357–382.

Asiedu, Edward, Dean Karlan, Monica Lambon-Quayefio, and Christopher Udry, "A call for structured ethics appendices in social science papers," *Proceedings of the National Academy of Sciences*, 2021, *118* (29), e2024570118.

Bai, Yuehao, "Optimality of matched-pair designs in randomized controlled trials," *American Economic Review*, 2022, *112* (12), 3911–3940.

Baird, Sarah, Craig McIntosh, and Berk Özler, "Cash or condition? Evidence from a cash transfer experiment," *The Quarterly Journal of Economics*, 2011, *126* (4), 1709–1753.

_ , J Aislinn Bohren, Craig McIntosh, and Berk Özler, "Optimal design of experiments in the presence of interference," *Review of Economics and Statistics*, 2018, *100* (5), 844–860.

Balboni, Clare, Oriana Bandiera, Robin Burgess, Maitreesh Ghatak, and Anton Heil, "Why do people stay poor?," *The Quarterly Journal of Economics*, 2022, *137* (2), 785–844.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman, "Labor markets and poverty in village economies," *The Quarterly Journal of Economics*, 2017, *132* (2), 811–870.

Banerjee, Abhijit, Emily Breza, Esther Duflo, and Cynthia Kinnan, "Can microfinance unlock a poverty trap for some entrepreneurs?," Technical Report, National Bureau of Economic Research 2025.

_ , Esther Duflo, and Garima Sharma, "Long-term effects of the targeting the ultra poor program," *American Economic Review: Insights*, 2021, *3* (4), 471–86.

_ , _ , Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry, "A multifaceted program causes lasting progress for the very poor: Evidence from six countries," *Science*, 2015, *348* (6236), 1260799.

Bauchet, Jonathan, Jonathan Morduch, and Shamika Ravi, "Failure vs. displacement: Why an innovative anti-poverty program showed no net impact in South India," *Journal of Development Economics*, 2015, *116*, 1–16.

Bedoya, Guadalupe, Aidan Coville, Johannes Haushofer, Mohammad Isaqzadeh, and Jeremy P Shapiro, "No household left behind: Afghanistan targeting the ultra poor impact evaluation," Technical Report, National Bureau of Economic Research 2019.

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, and Bertil Tungodden, "Human and financial capital for microenterprise development: Evidence from a field and lab experiment," *Management Science*, 2015, *61* (4), 707–722.

Bergquist, Lauren Falcao and Michael Dinerstein, "Competition and entry in agricultural markets: Experimental evidence from Kenya," *American Economic Review*, 2020, *110* (12), 3705–3747.

Blattman, Christopher and Laura Ralston, "Generating employment in poor and fragile states: Evidence from labor market and entrepreneurship programs," *Available at SSRN 2622220*, 2015.

_ , Eric P Green, Julian Jamison, M Christian Lehmann, and Jeannie Annan, "The returns to microenterprise support among the ultrapoor: A field experiment in postwar Uganda," *American Economic Journal: Applied Economics*, 2016, *8* (2), 35–64.

_ , Nathan Fiala, and Sebastian Martinez, "Generating skilled self-employment in developing countries: Experimental evidence from Uganda," *The Quarterly Journal of Economics*, 2014, *129* (2), 697–752.

Bresnahan, Timothy F and Peter C Reiss, "Entry and competition in concentrated markets," *Journal of political economy*, 1991, *99* (5), 977–1009.

Brooks, Wyatt, Kevin Donovan, and Terence R Johnson, "Mentors or teachers? Microenterprise training in Kenya," *American Economic Journal: Applied Economics*, 2018, *10* (4), 196–221.

_ , _ , _ , and Jackline Oluoch-Aridi, "Cash transfers as a response to COVID-19: Experimental evidence from Kenya," *Journal of Development Economics*, 2022, *158*, 102929.

Brown, Judith K., "A Note on the Division of Labor by Sex," *American Anthropologist*, 1970, *72* (5), 1073–1078.

Bruhn, Miriam and David McKenzie, "In pursuit of balance: Randomization in practice in development field experiments," *American Economic Journal: Applied Economics*, 2009, *1* (4), 200–232.

Bryan, Gharad, Shyamal Chowdhury, Ahmed Mushfiq Mobarak, Melanie Morten, and Joeri Smits, "Encouragement and distortionary effects of conditional cash transfers," 2021.

Cahyadi, Nur, Rema Hanna, Benjamin A Olken, Rizal Adi Prima, Elan Satriawan, and Ekki Syamsulhakim, "Cumulative impacts of conditional cash transfer programs: Experimental evidence from Indonesia," *American Economic Journal: Economic Policy*, 2020, *12* (4), 88–110.

Chaisemartin, Clément De and Jaime Ramirez-Cuellar, "At what level should one cluster standard errors in paired and small-strata experiments?," *American Economic Journal: Applied Economics*, 2024, *16* (1), 193–212.

Chowdhury, Reajul, Elliott Collins, Ethan Ligon, and Munshi Sulaiman, "Valuing Assets Provided to Low-Income Households in South Sudan," 2017.

Cunha, Jesse M, Giacomo De Giorgi, and Seema Jayachandran, “The price effects of cash versus in-kind transfers,” *The Review of Economic Studies*, 2019, *86* (1), 240–281.

Drexler, Alejandro, Greg Fischer, and Antoinette Schoar, “Keeping it simple: Financial literacy and rules of thumb,” *American Economic Journal: Applied Economics*, 2014, *6* (2), 1–31.

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker, “General equilibrium effects of cash transfers: Experimental evidence from Kenya,” *Econometrica*, 2022, *90* (6), 2603–2643.

Fiala, Nathan, Julian Rose, Filder Aryemo, and Jörg Peters, “The (very) long-run impacts of cash grants during a crisis,” 2022, (961).

Filmer, Deon, Jed Friedman, Eeshani Kandpal, and Junko Onishi, “Cash transfers, food prices, and nutrition impacts on ineligible children,” *The Review of Economics and Statistics*, 2021, pp. 1–45.

Fisher, Ronald Aylmer, *The design of experiments*, Edinburgh: Oliver and Boyd, 1935.

Foster, Andrew, Moritz Poll, and Hyun Soo Suh, “Female Micro-Entrepreneurship and Child Development,” Technical Report 2026.

Gertler, Paul J, Sebastian W Martinez, and Marta Rubio-Codina, “Investing cash transfers to raise long-term living standards,” *American Economic Journal: Applied Economics*, 2012, *4* (1), 164–92.

Giné, Xavier and Ghazala Mansuri, “Money or management? A field experiment on constraints to entrepreneurship in rural Pakistan,” *Economic Development and Cultural Change*, 2021, *70* (1), 41–86.

Goldberg, Pinelopi Koujianou and Tristan Reed, “Presidential Address: Demand-Side Constraints in Development. The Role of Market Size, Trade, and (In)Equality,” *Econometrica*, 2023, *91* (6), 1915–1950.

Hardy, Morgan and Gisella Kagy, “It’s getting crowded in here: experimental evidence of demand constraints in the gender profit gap,” *The Economic Journal*, 2020, *130* (631), 2272–2290.

Haushofer, Johannes and Jeremy Shapiro, “The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya,” *The Quarterly Journal of Economics*, 2016, *131* (4), 1973–2042.

_ and _ , “The long-term impact of unconditional cash transfers: Experimental evidence from Kenya,” *Busara Center for Behavioral Economics, Nairobi, Kenya*, 2018.

Ikegami, Munenobu, Michael R Carter, Christopher B Barrett, and Sarah Janzen, “Poverty traps and the social protection paradox,” *The economics of poverty traps*, 2017, pp. 223–256.

Kansikas, Carolina, Anandi Mani, and Paul Niehaus, “Customized cash transfers: Financial lives and cash-flow preferences in rural Kenya,” 2023.

Leight, Jessica, Daniel O Gilligan, Michael Mulford, Alemayehu Seyoum Taffesse, and Heleene Tambet, "Aspiring to more? New evidence on the effect of a light-touch aspirations intervention in rural Ethiopia," Technical Report 2021.

_ , Josué Awonon, Abdoulaye Pedehombga, Rasmané Ganaba, and Aulo Gelli, "How light is too light touch: The effect of a short training-based intervention on household poultry production in Burkina Faso," *Journal of Development Economics*, 2022, *155*, 102776.

McKenzie, David and Susana Puerto, "Growing markets through business training for female entrepreneurs: A market-level randomized experiment in Kenya," *American Economic Journal: Applied Economics*, 2021, *13* (2), 297–332.

Mel, Suresh De, David McKenzie, and Christopher Woodruff, "Returns to capital in microenterprises: Evidence from a field experiment," *The Quarterly Journal of Economics*, 2008, *123* (4), 1329–1372.

_ , _ , and _ , "One-time transfers of cash or capital have long-lasting effects on microenterprises in Sri Lanka," *Science*, 2012, *335* (6071), 962–966.

Millán, Teresa Molina, Karen Macours, John A Maluccio, and Luis Tejerina, "Experimental long-term effects of early-childhood and school-age exposure to a conditional cash transfer program," *Journal of Development Economics*, 2020, *143*, 102385.

Misha, Farzana A, Wameq A Raza, Jinnat Ara, and Ellen Van de Poel, "How far does a big push really push? Long-term effects of an asset transfer program on employment trajectories," *Economic Development and Cultural Change*, 2019, *68* (1), 41–62.

Morgan, Kari Lock and Donald B Rubin, "Rerandomization to improve covariate balance in experiments," *The Annals of Statistics*, 2012, *40* (2), 1263–1282.

Muralidharan, Karthik, Mauricio Romero, and Kaspar Wüthrich, "Factorial designs, model selection, and (incorrect) inference in randomized experiments," *Review of Economics and Statistics*, 2023, pp. 1–44.

Poll, Moritz, "Pre-Analysis Plan: Micro-enterprise saturation and poverty graduation at high frequency," Technical Report 2023.

_ , "Poverty alleviation at high frequency," Technical Report 2026.

_ , "Seasons of change: Experimental evidence on the optimal timing for poverty alleviation," Technical Report 2026.

Ritzwoller, David M, Joseph P Romano, and Azeem M Shaikh, "Randomization inference: Theory and applications," *Journal of Political Economy: Microeconomics*, 2025.

Rosenstein-Rodan, Paul N., "Notes on the theory of the 'big push'," in "Economic Development for Latin America: Proceedings of a conference held by the International Economic Association" Springer 1957, pp. 57–81.

Schreiner, Mark, "Simple Poverty Scorecard® Tool Malawi," 2019.

Sedlmayr, Richard, Anuj Shah, and Munshi Sulaiman, "Cash-plus: Poverty impacts of alternative transfer-based approaches," *Journal of Development Economics*, 2020, *144*, 102418.

Sulaiman, Munshi, Nathanael Goldberg, Dean Karlan, and Aude de Montesquiou, "Eliminating extreme poverty: Comparing the cost-effectiveness of livelihood, cash transfer, and graduation approaches," 2016.

Valdivia, Martin, "Business training plus for female entrepreneurship? Short and medium-term experimental evidence from Peru," *Journal of Development Economics*, 2015, *113*, 33–51.

Wydick, Bruce, "When Are Cash Transfers Transformative?," 2020.

Young, Alwyn, "Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results," *The Quarterly Journal of Economics*, 2019, *134* (2), 557–598.

Appendix

A Additional descriptive figures and tables

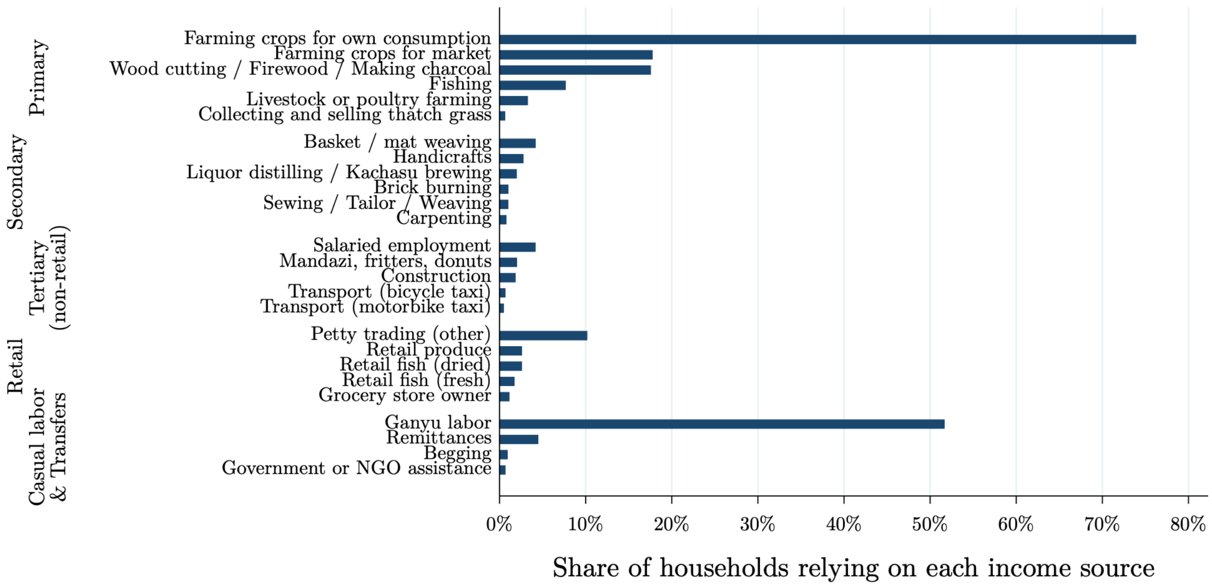

Figure A1: Most common income sources in the study area

Note: This figure describes the income sources of the population as stated during the area-wide census. People are predominantly farmers or casual "‘ganyu" laborers. Outside of the primary sector, retail is another common income source.

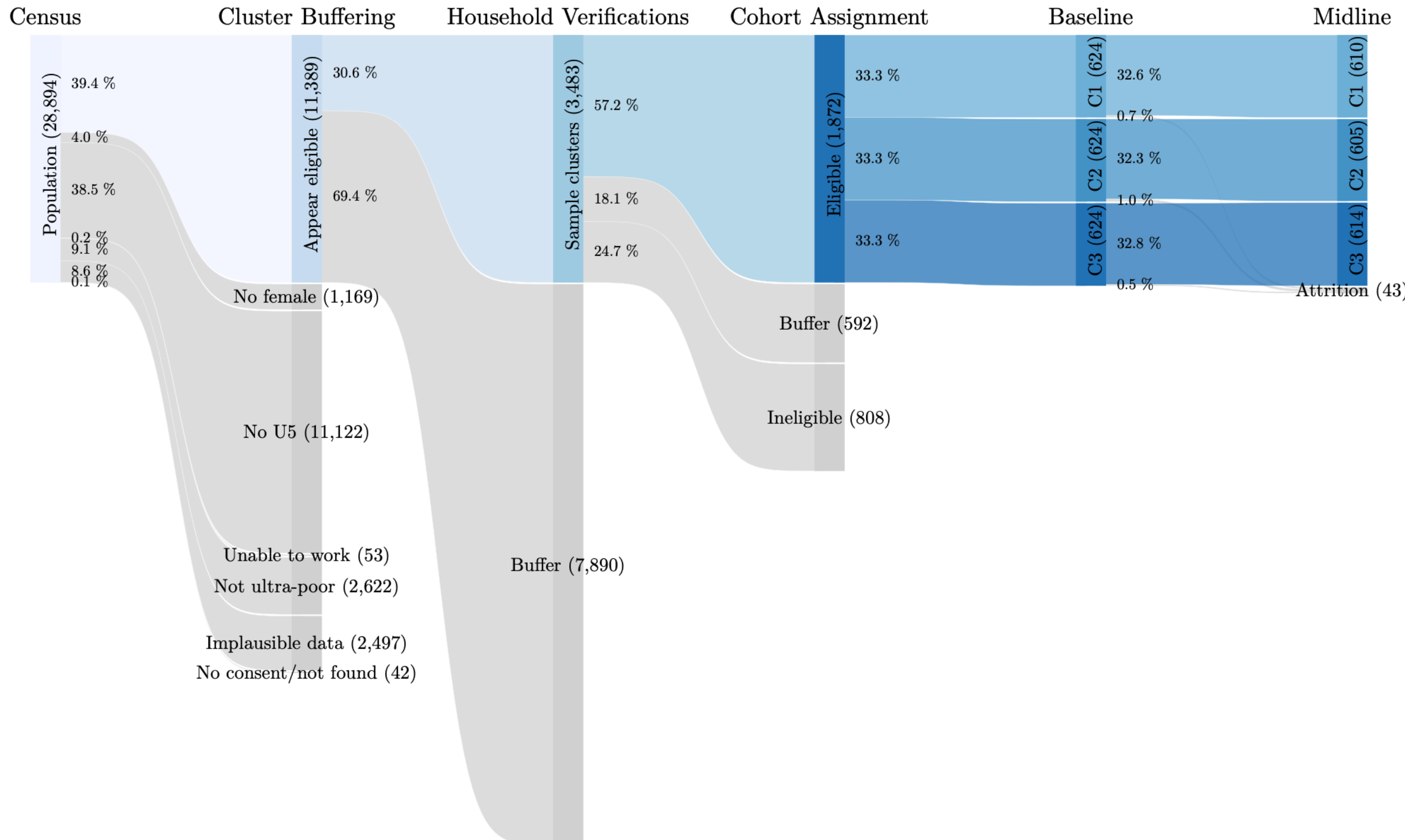

Figure A2: **Sampling funnel**

Note: This figure describes how the sample is selected from the census population. Starting from the census population of about 29,000 households, about 60% were deemed ineligible, mostly for lack of a child under age five. Among apparently eligible households, 271 clusters of 40-50 households were formed by proximity using k-means clustering. A second custom algorithm then selected 72 of these clusters in such a way as to maximize the distance between the nearest two clusters, in order to minimize spillover concerns (cf. Appendix Figure A3). The other clusters serve as buffer zones. Among the selected clusters, household verifications were conducted that resulted in some households being deemed ineligible under the same criteria as in the census, while selecting the 26 eligible households closest to the cluster centroid and excluding the remainder as additional buffer households on the periphery of the clusters. The 72 clusters were grouped into triplets by proximity and assigned an implementing staff. They were also assigned at random to three cohorts to undergo the intervention at different times in the agricultural season. The full sample was surveyed at baseline and the graph shows attrition in the subsequent survey rounds.

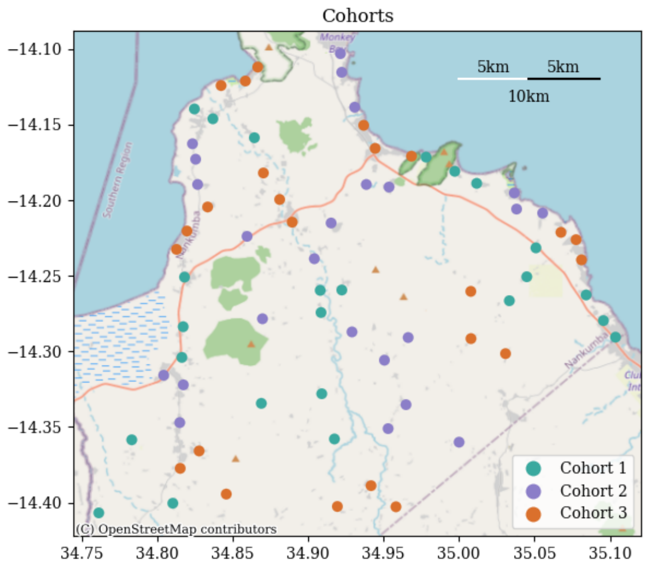

Figure A3: **Clusters** by cohort

Note: This map shows the 72 study clusters of 26 households each that make up the study sample, color-coded by the cohort they are assigned to.

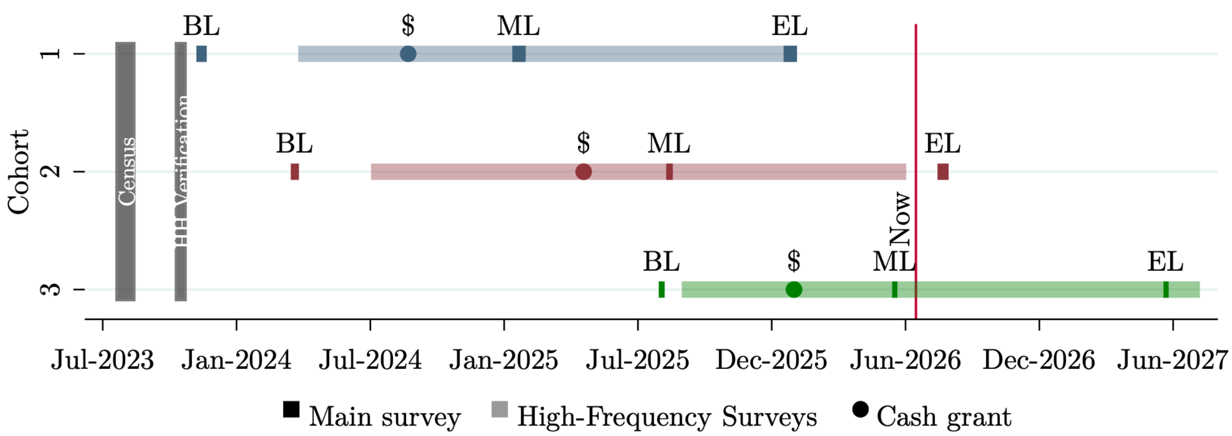

Figure A4: Study timeline

Note: This figure describes the study timeline. A census of the study area was conducted in August 2023. Household eligibility verifications followed in October 2023. The baseline survey for cohort 1 was conducted in December 2023 and high-frequency surveys in this cohort started in April 2024. The large one-off cash grant was paid out in August 2024. The midline survey was collected in January 2025, followed a year later by the endline survey.

Table A1: Study population

	(1) Count
HOUSEHOLDS	
Number of surveys conducted	28,853
Female primary adult present	27,684
U5 present	16,882
No children present	3,254
INDIVIDUALS	
Population	138,548
Adults	65,162
Children	73,386
U5s	20,998
Number of villages	162
Number of CBOs	25
Number of ADCs	5

Census of all households in Mangochi district's subdistrict ("Traditional Authority") Nankumba. The urban center Monkey Bay and the tourist destination Cape Maclear were omitted from this census as well as an area in the South-West of Nankumba where another NGO is implementing a livelihood program. Area Development Committees (ADCs) are administrative subdivisions of a subdistrict, while Community-Based Organizations (CBOs) are associations of several villages organizing development initiatives or providing local public goods together.

Table A2: Livelihoods

	(1) Percent
AGRICULTURE	
Crop farming	76.6
Subsistence crop farming	73.5
Marketed crop farming	17.8
Share of yield marketed	29.2
Share of yield marketed (conditional)	69.8
Livestock	3.3
PIECE WORK	
Ganyu	51.7
Exclusively subsistence + ganyu	27.6
BUSINESS	
Business	58.0
Employees > 0	4.1
Local non-ag employment potential	3.4
Observations	29084

Table A3: Poverty

	(1) Percent
Skip meal (past two weeks)	76.8
Female illiterate	37.7
Thatch roof	64.5
No furniture	80.3
No phone	51.9
No bedding	15.7
<2.15 PPP 2017 USD (self report)	92.7
<2.15 PPP 2017 USD (PMT)	85.1
# children / adult (if $>$ 0)	1.5
# income sources	1.5

Table A4: Estimate of local GDP

	(1) MWK	(2) USD	(3) USD BM	(4) USD PPP
GDP (millions)	12,707	12.07	7.06	35.56
Subsistence production (millions)	8,614	8.18	4.79	24.11
GDP/capita	91,717	87.10	50.95	256.69
Subsistence production/capita	62,171	59.04	34.54	174.00
GDP/adult	195,009	185.19	108.34	545.79
Subsistence production/adult	132,189	125.54	73.44	369.97

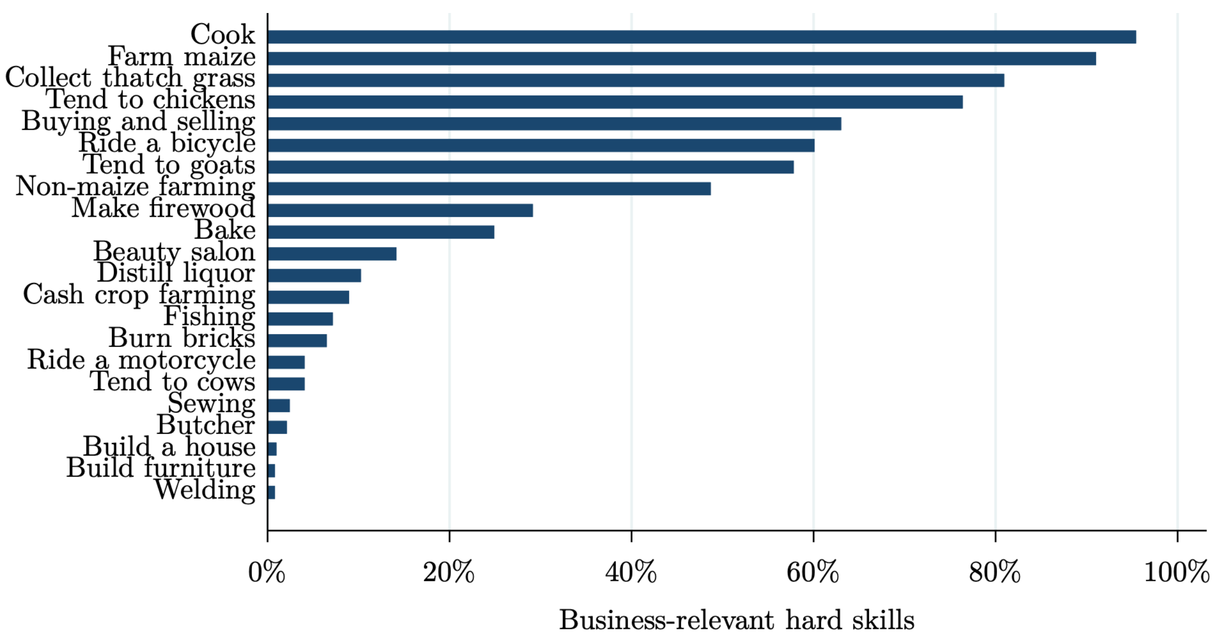

Figure A5: **Hard skills** outside of the retail sector and household chores are not common in this population.

B Additional results

Table B1: Balance

	(1) Control Mean (SD)	(2) Treatment (SE)
HOUSEHOLD AND PARTICIPANT CHARACTERISTICS		
Household size	5.432	-0.118
	(1.827)	(0.097)
Age	31.6	-0.564
	(8.443)	(0.424)
Years of education	5.657	-0.122
	(2.880)	(0.191)
Literate	0.692	0.005
	(0.462)	(0.027)
LIVELIHOOD AND ASSETS		
Hours worked (7 days)	22	-0.364
	(17.9)	(0.783)
HH income (12 months, MWK)	475,629	11,315
	(480,571)	(47,515)
Total savings, assets, livestock, minus debt (MWK)	256,874	-17,471
	(365,607)	(24,353)
Largest purchase (12 months, MWK)	38,607	-2,816
	(121,528)	(6,344)
POVERTY AND FOOD SECURITY		
Thatch / tarp / no roof	0.746	0.021
	(0.436)	(0.023)
Food insecurity score (0-8)	7.532	0.049
	(1.341)	(0.064)
HEALTH AND NON-HEALTH SHOCKS		
Health episodes (12 months)	15.6	-0.578
	(12.8)	(0.920)
Non-health shocks (12 months)	20.6	-0.377
	(18.1)	(0.853)
NOT PRE-SPECIFIED		
Engaged in retail at baseline	0.181	-0.007
	(0.385)	(0.024)
Entrepreneurship knowledge score (0-100)	19.2	0.171
	(7.035)	(0.426)
Observations		1,872
F		0.58
$p(F = 0)$		0.851

This table describes the baseline balance. The main panel reports the control group means and coefficients of the treatment indicator in a regression predicting each baseline outcome variable. At the bottom of the table, an F statistic is reported for a regression in which all outcomes jointly predict the treatment. Only the pre-specified variables are included in the F-test. Standard errors are clustered at the household cluster level and reported in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B2: Childcare arrangements while working

	(1) Control mean (SD)	(2) Treatment Effect (SE)	(3) N (BL)
Left the baby home alone	0.026	-0.018	1528
	(0.124)	(0.016)	×
		[0.123]	
∼ with another child younger than 10 years old	0.023	-0.017^{***}	1528
	(0.077)	(0.005)	×
		[0.000]	
∼ with another child that is 10 years or older	0.085	-0.035^{***}	1528
	(0.156)	(0.011)	×
		[0.001]	
∼ with the father	0.027	-0.009	1528
	(0.095)	(0.015)	×
		[0.829]	
∼ with another adult family member	0.300	-0.074^{***}	1528
	(0.293)	(0.023)	×
		[0.001]	
∼ with an adult neighbor	0.071	-0.019^{**}	1528
	(0.136)	(0.007)	×
		[0.007]	
Took the baby with you	0.268	-0.065^{**}	1528
	(0.305)	(0.031)	×
		[0.037]	
Brought baby to the CBCC or another childcare facility	0.199	0.238^{***}	1528
	(0.286)	(0.044)	×
		[0.000]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B3: Post-training evaluation results

	(1) Treatment mean (SD)	(2) High Saturation	(3) N (BL)
Business knowledge	0.456	−0.005	923
	(0.160)	(0.012)	✓
		[0.693]	
ECD knowledge	0.521	−0.009	923
	(0.135)	(0.010)	✓
		[0.373]	
Preventive health knowledge	0.604	−0.010	923
	(0.208)	(0.015)	✓
		[0.529]	

Notes: Analysis in this table is among treated households (who wrote a business plan) and the treatment coefficient is therefore omitted. The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B4: "Name factors that contribute to business success."

	(1) Treatment mean (SD)	(2) Coordination Intervention	(3) N (BL)
Competitor weaknesses	0.342	0.108***	923
	(0.475)	(0.037)	✓
		[0.013]	
Business niche	0.623	0.021	923
	(0.485)	(0.030)	✓
		[0.491]	
Aware of competitors	0.488	−0.010	923
	(0.500)	(0.036)	✓
		[0.769]	

Notes: Analysis in this table is among treated households (who wrote a business plan) and the treatment coefficient is therefore omitted. The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.