Using Household Grants to Benchmark the Cost Effectiveness of a USAID Workforce Readiness Program

Craig McIntosh* and Andrew Zeitlin†

This version: April 2, 2022

Abstract

We use a randomized experiment to benchmark a workforce training program to cash transfers in Rwanda. Conducted in a sample of poor and underemployed youth, this study measures the impact of the training program relative not only to a control group, but also to the counterfactual of simply disbursing the cost of the program directly to beneficiaries in cash. The training program was successful in improving a number of core outcomes, including productive hours, assets, savings, and subjective well-being. However, cost-equivalent cash transfers move all these outcomes as well as consumption, income, and wealth. At cost-equivalent levels, cash transfers prove superior across a number of economic outcomes, while training outperforms cash only in the production of business knowledge. Above cost-equivalent levels, we see limited benefits from increasing cash transfer amounts; going from $410 to $750 generates few benefits and an apparent decrease in labor hours. There is a surprising absence of complementarity between human and physical capital interventions, with an arm receiving both interventions doing slightly worse than what we would expect from the independent impacts of each of the two, though given diminishing returns to cash this combined arm outperforms a cash transfer of approximately equal cost. Heterogeneity in impacts and spillover effects are limited, suggesting that the relative impacts of these interventions will be similar across different targeting rules and saturation levels within this population.

Keywords: Experimental Design, Cash Transfers, Employment
JEL Codes: O12, C93, J21
Study Information: This study is registered with the AEA Trial Registry as Number AEARCTR-0004388, and is covered by Rwanda National Ethics Committee IRB 114/RNEC/2017, IPA-IRB:14609, and UCSD IRB 161112. The research was paid for by USAID grant AID-0AA-A-13-00002 (SUB 00009051). We thank the Education Development Center, GiveDirectly, USAID, and DIL/CEGA for their close collaboration in executing the study, Leodomir Mfura, Marius Chabi, and Innovations for Poverty Action for their data collection work, and USAID Rwanda, DIV, and Google.org for funding. This study has been accepted by the *Journal of Development Economics* under their pre-results review process. This study is made possible by the support of the American People through the United States Agency for International Development (USAID.) The contents of this study are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government.

*University of California, San Diego, ctmcintosh@ucsd.edu
†Georgetown University, andrew.zeitlin@georgetown.edu

1 Introduction

The demographic dividend in Sub-Saharan Africa is a double-edged sword. A young population carries the prospect of many productive years ahead and a limited burden of dependency from older generations. But this opportunity can not be capitalized upon if young people are unable to find productive employment (Fox et al., 2016). In spite of gains in formal educational attainment, youth unemployment rates remain high; for example, while 40 percent of Rwanda's population is between the ages of 14–30, 65 percent of these youth are unemployed. This raises the prospect of both a lost generation of opportunity, and the political risks that accompany a large, unemployed, urban, young population (Bongaarts, 2016). Yet despite this pressing need, there is limited evidence of the relative effectiveness of alternative strategies to improve productive employment for youth.

It has long been recognized that both physical and human capital—"money and ideas"—may constrain the productivity of young entrepreneurs (Giné and Mansuri, 2014). To build human capital, skills training programs attract billions each year: Blattman and Ralston (2015) estimate that the World Bank alone spends almost a billion dollars annually on such programs.[1] Despite some signs of success in generating employment (Alfonsi et al., 2019; Diaz and Rosas, 2016), systematic reviews of the employment and income impacts of programs aimed at lifting human capital show these have been modest on average and highly variable across studies (Kluve et al., 2017; McKenzie, 2021). At the same time, the costs of relaxing capital constraints are falling due to the widespread availability of mobile money in the developing world. A large literature finds that unconditional cash transfers are invested in durables (Haushofer and Shapiro, 2016), productive assets (Blattman et al., 2018; Gertler et al., 2012), and microenterprises (De Mel et al., 2012), suggesting that cash may be a reasonable alternative in delivering economic livelihood assistance to youth. In its *Future Drivers of Growth* report, the Government of Rwanda raises the possibility that "for a significant portion of the population who will continue creating their own jobs, capital-centric programs may be more effective and cheaper to implement than simple training programs" (Government of Rwanda and World Bank Group, 2019, p. 81). But in spite of the widespread emphasis on these channels of investment, few studies have directly compared cash and human-capital investments, leaving differences in populations and resource-intensities of investment to confound comparisons between them. Rigorous comparative cost-effectiveness research across these different modalities—as well as a better understanding of their potential complementarities—is sorely needed.

This study addresses this need by undertaking an exercise in *cash benchmarking*: the direct comparison of in-kind to cash-transfer programs in a single experimental setting. Together with a companion paper in a child-health application (McIntosh and Zeitlin, 2019), this study develops a methodology for making such comparisons at cost-equivalent levels, in the presence of ex-ante uncertainty about program costs. Here, we present an an individually randomized experiment that

[1] The Government of Rwanda places a high priority on training programs: Priority Area 1 of the "Economic Transformation Pillar" in its seven-year plan for the period 2017–2024 includes the key strategic intervention to "support and empower youth and women to create businesses through entrepreneurship and access to finance" (Republic of Rwanda, 2017, p. 3).

compares the tradeoffs and complementarities between human- and physical-capital interventions on youth employment outcomes.

As an applied-science exercise, cash benchmarking represents a form of comparative cost-effectiveness analysis. We compare the returns to alternative programming modalities on a pre-defined set of outcomes, and we answer this counterfactual question subject to a distributional constraint, that holds expenditure on programming per beneficiary constant across modalities. This addresses a critical policy comparison of the impact of a program relative to what would happen if its costs were distributed directly to the beneficiaries—the answer to which can both inform choices among existing programming modalities, and create pressure to improve value for money in in-kind programming (Blattman and Niehaus, 2014).

Cash benchmarking also informs a basic-science question, by lifting distinct constraints to beneficiary welfare in a comparable way. This basic-science question is in a similar spirit to Ahmed et al. (2016), who compare BRAC's ultra-poor programming to cash, or Karlan et al. (2014), who examine the comparative impact of the relaxation of credit and risk constraints in agriculture. In the context of youth livelihoods, training programs and cash grants each move alternative potential constraints to productive employment—skills and liquidity, respectively. For any given outcome, our design allows us to cast the opportunity cost of skills improvement in pecuniary terms, despite the fact that these skills cannot be bought on the market. This allows us not only to determine the benefit generated by an increment of skills improvement, but also to calculate the shadow cost of generating the same benefit by relaxing financial (rather than human capital) constraints. The inclusion of a combined arm allows us to study complementarities, asking if the returns to relaxing human capital constraints improve when financial constraints have also been relaxed.[2]

We study this question using a randomized trial with 1,848 underemployed Rwandan youth to understand how a 'standard' package of training, soft skills, and networking interventions compares not only to an experimental control group but also to an additional arm that receives household grants intended to be of equal cost to the donor—the cash benchmark. The study follows poor, underemployed youth aged 15–30 who expressed interest in participating in the training program. The core program is called *Huguka Dukore/Akazi Kanoze*, which means "Get Trained and Let's Work/Work Well Done" in Kinyarwanda (abbreviated henceforth as HD). It follows USAID's strategy on workforce readiness and skills training and was implemented by Education Development Center, Inc. (EDC). The cash transfer program was delivered by GiveDirectly (GD), a U.S.-based nonprofit that specializes in making unconditional household grants via mobile money. USAID also uses cash transfers in its programming in a number of contexts, so this study compares two different means through which it could attempt to deliver benefits.[3] These two treatments are compared to a control group, namely a set of individuals that receive neither program, and a combined arm that receives both.

[2] Fox and Kaul (2017) say that "It may be helpful to experiment more with transferable skills development initiatives. . . combined with cash transfers to youth or access to finance."

[3] USAID currently uses cash mostly in the humanitarian space, but is also involved in new efforts to explore cash as a form of development assistance in countries such as Morocco and Nigeria.

This study provides a methodology incorporating randomization of transfer amounts and ex-post regression cost adjustment that can achieve this benchmarking objective in a general way, as in our companion paper (McIntosh and Zeitlin, 2019). The cost of the "combined" arm receiving both intermediate cash and training is almost identical to the largest cash arm, giving us a very unusual ability to make two different cost-equivalent comparisons within the same study: human versus physical capital, and a combination of the two versus an very large physical capital transfer. We can then relate these two tests to perform a novel comparative form of benchmarking: is there a different horse-race result between cash and kind if we start from nothing, as opposed to adding these two interventions over the top of a moderate cash transfer base? We conclude our empirical analysis by contrasting a cost-equivalence approach with a cost effectiveness analysis across arms of unequal cost.

The combination of its individual level of treatment, private benefits, and investment-related outcomes make Huguka Dukore a particularly attractive candidate for benchmarking against cash transfers. It is a five-year project (2017–2021) aiming to provide 40,000 vulnerable youth with employability skills in 19 (of 30 total) districts nationwide. Targeting youth from poor households with less than secondary education, with an emphasis on women and youth with disabilities, HD offers multiple program pathways including: (i) employment preparation; (ii) individual and co-operative youth microenterprise start-up; (iii) business development for existing microenterprises. HD is based on a predecessor program, *Akazi Kanoze*, which operated in the country from 2012-2017, and which was found to be successful in an RCT led by EDC (Alcid, 2014). It is a carefully designed and intensive training program, and it is backed by rigorous research heading into this study. Moreover, programs such as Huguka Dukore seek to move a core set of outcomes related to employment, income generation, and consumption that have also been the objective of cash-transfer programs. These features mean not only that cash transfers provide a plausible alternative means to achieve HD's own objectives, but further, that there are open questions about whether the overhead costs required to deliver such a bundled program are justified by underlying market failures.

The randomized controlled trial proceeded in four steps. First, EDC's three local implementing partners within the study ran recruiting workshops drawing more than 2,000 eligible individuals who expressed interest in participating in HD. The HD-imposed eligibility criteria for their training of vulnerable youth consist of (a) ages ranging from 16–30, and (b) between 6-12 years of education. Because of the conditions placed on GiveDirectly by the Rwandan government, we further strictly limited study eligibility to (c) households registered in Ubudehe poverty status 1 or 2 (the poorest). In addition, in order to generate the highest possible compliance rates with HD, we further restricted eligibility to those who (d) expressed interest in participating in the employment and entrepreneurship readiness training. From this group we then conducted baselines, verified eligibility, and recruited an experimental study sample of 1,848 individuals who consented to the lottery and the baseline survey and were included in the randomization. These individuals come from 328 villages in non-urban areas of the three districts of Rwamagana, Muhanga and Nyamagabe. They were recruited at a first 'orientation' meeting at which the local HD implementers and the survey firm (Innovations for Poverty Action, or IPA) recorded sufficient information to enroll them and to subsequently perform baseline surveys (there were no refusals to the household survey). Then, IPA conducted a series of 13 public lotteries at the local level (a geographic division above the village called a 'sector' in Rwanda), overseen by local officials, at which individuals were assigned to four main arms and a control. The study collected baseline and midline (18 month) indicators across a range of economic, psychological, and business-related outcomes to measure comparative impacts.

We costed both programs in detail prior to, and after, the intervention period, following the "ingredients method" discussed in Levin et al. (2017). The ex-ante costing exercise was used to identify the approximate total cost of the HD intervention, as well as the estimated overhead costs to GiveDirectly of providing household grants in this context. The ex-ante costing of HD arrived at a per-beneficiary cost of $464.25. We then randomized transfer amounts at the individual level in the cash arm across four possible transfer amounts. These amounts were chosen to provide informative benefit/cost comparisons across two different margins: HD versus cash, and small versus large cash transfer amounts. Incorporating GiveDirectly's operating costs, the transfer actually received by households at equal anticipated cost to HD was $410.65. To reflect uncertainty in the ex-ante costing exercise, we randomized two bracketing cash transfer arms of $317 and $503 (cost to USAID of $394 and $590 per sample beneficiary, respectively). These large transfer amounts mean that even our smallest transfer is providing individuals with 167% of annual average per capita income. In the end, the HD program turned out to be less expensive than expected; the final ex-post costing figure of $332.27 is used to regression-adjust outcomes across transfer amounts to arrive at a comparison between HD and cash estimated at the exact cost-equivalent amount from the perspective of the donor, USAID. Because of the low costing number, an adjustment which was intended to be an interpolation over GD transfer costs ends up being an extrapolation to a value 16% lower than the smallest GD arm's cost.

The study also features two more expensive interventions: a combined arm and a large cash transfer arm. The combined group received both the middle GD transfer amount and HD training, permitting a classic test for complementarities between human capital and financial interventions. The larger cash transfer arm was an amount chosen by the cash implementer as maximizing their own cost-effectiveness, transferring $750. These two larger arms turn out to have very similar costs of around $850, enabling a number of interesting comparisons in benefit/cost ratios as program cost rises. These more expensive arms open up a different type of comparative cost-effectiveness question: would the net benefit from cash transfers be maximized by concentrating large payments on a few individuals, or by spreading out smaller transfers to more people? And if more money is to be invested over the basic transfer, should it be in the form of additional physical capital, or is training then more effective?

Our results show that at 18 months after baseline, on average 15 months after being offered HD programming and at least 3 months after any training would have ended, Huguka Dukore has delivered real benefits. While we do not find statistically significant impacts on employment rates, we cannot rule out employment effects of HD smaller than 6 percentage points, and HD induces significant increases in non-agricultural wage labor and agriculture-based microenterprises, consistent with the training provided. The HD arm sees an increase in productive hours, a more than doubling of productive asset values, average savings increases of 60%, and improvements in subjective well-being. In addition, the HD arm performs a half a standard deviation better on a test of business knowledge built against the course curriculum, showing that the program clears the basic bar of having created real learning.

The cash transfer arm, on average 14 months after transfers took place, sees improvements across a broad range of economic and psychological outcomes. These impacts prove surprisingly invariant to the transfer amount variation present in this study, suggesting that even our lower transfers clear a barrier that generates real benefit to households. With the exception of the improvement that HD generates on business knowledge, cash improves every outcome that HD improves (generally with a greater magnitude) and in addition drives monthly income, household- and individual-level consumption, livestock value, and overall wealth, to higher levels.

Consequently, when we conduct our pre-specified comparative impact analysis, we find cost-equivalent cash to generate significantly larger benefits for income, productive and livestock assets, individual consumption, and subjective well-being, while HD is more effective at generating the human capital benefit of business knowledge. A canonical test of complementarities, using a two-by-two design, finds no evidence of complementarity between human and physical capital interventions. However, when we form the comparative benefits of cash versus training over the top of the GD Middle transfer and compare these to the comparative benefits starting from nothing, a different picture emerges. Because the marginal returns to cash over the top of moderate transfers are very limited, even in the absence of 'true' complementarities we find that training looks comparatively more attractive relative to cash starting over moderate transfers than it does from nothing. In other words, the comparative attractiveness of training increases as the amount of cash already transferred goes up. This is a direct product of the concave returns to cash in this study.

Analyses of heterogeneity and interference speak to the transportability and scalability of these results. We see little meaningful heterogeneity in the study, finding that both interventions have a relatively consistent effect across gender, baseline consumption levels, and local labor market conditions. This suggests that (at least within our relatively stringent initial eligibility criteria) there is little evidence that targeting the program differently would have generated superior results. To study spillovers, we exploit the random variation in intensity of treatment at the village level generated by the lotteries.[4] The only evidence we find for spillovers is in compliance with HD, which increases with the share of eligible individuals assigned to HD in the same village. Looking at our primary outcomes, we find no evidence that spillovers from any treatment or to any treatment group are present, suggesting that by this measure at least the study is internally valid.

These results illustrate the complexity of the comparisons created by this type of benchmarked

[4]This issue is particularly important given our household-level assignment and the recent evidence on spillovers both from job training programs (Crépon et al., 2013) and cash transfer programs (Angelucci and De Giorgi, 2009; Egger et al., 2019).

design. Considered relative to the control, HD has been successful in moving some of the key welfare indicators the program is geared towards, and has strongly improved the core metric of learning. Even in a comparative sense, it is impressive that a program that made no material transfer to its beneficiaries could generate improvements in asset values half as large, and improvements in savings two thirds as large, as a program that gave them hundreds of dollars. Nonetheless, when we compare the programs head-to-head at cost-equivalent levels, it becomes clear that at least at 18 months from baseline, cash is outperforming the training program across a set of indicators likely to be central to beneficiary economic welfare (individual consumption, income, livestock wealth, and subjective well-being). In something of a challenge to ever-more complex bundled programs, we find that each of these interventions has a distinct set of benefits that operate independently, and little is gained by providing them together.

The benchmarking approach taken in this study complements the numerous studies that have shown meaningful impacts of cash transfers on important life outcomes in the short term, such as child nutrition (Aguero et al., 2006; Seidenfeld et al., 2014), schooling (Skoufias et al., 2001), mental health (Baird et al., 2013; Samuels and Stavropoulou, 2016), teen pregnancy and HIV (Baird et al., 2011), microenterprise outcomes (De Mel et al., 2012), consumer durables (Haushofer and Shapiro, 2016), and productive assets (Gertler et al., 2012). The evidence on the long-term impacts of cash transfers is more mixed (Blattman et al., 2018), but some studies have found substantial impacts (Aizer et al., 2016; Balboni et al., 2019; Barham et al., 2014; Fernald et al., 2009; Hoynes et al., 2016).[5] Most existing cash benchmarking research is based on the comparison of cash aid to food aid (Ahmed et al., 2016; Cunha et al., forthcoming; Hidrobo et al., 2014; Hoddinott et al., 2014; Leroy et al., 2010; Schwab et al., 2013). These studies have typically found that food aid leads to a larger change in total calories while cash aid leads to an improvement in the diversity of foods consumed. Efforts to benchmark more complex, multi-dimensional programs to cash include BRAC's Targeting the Ultra-Poor program (Chowdhury et al., 2016), microfranchising (Brudevold-Newman et al., 2017), and graduation programs (Sedlmayr et al., 2020). In practice, ex ante uncertainty about operational costs has meant that such studies often exhibit substantial differences in the value of cash and in-kind programming delivered, or benchmark only a subset of the in-kind bundle against cash. Our approach allows us to make exact cost-equivalent comparisons in estimating relative impacts across modalities.

In the remainder of the paper, we provide details of the experimental design, survey structure, and costing exercise in Section 2, and then present the results of the study in Section 3. Section 4 draws together lessons for how implementers might consider value for money in this context. Section 5 concludes.

[5] For examples of studies that find dissipating long-term benefits, see Baird et al. (2019), Araujo et al. (2017), and Brudevold-Newman et al. (2017). Evidence from systematic reviews of cash transfers on schooling (Molina-Millan et al., 2016) and child health (Manley et al., 2013; Pega et al., 2014) has also been uneven, with substantial heterogeneity in findings across studies.

2 Experimental Design

2.1 Interventions

Huguka Dukore: Employment and entrepreneurship readiness training

Huguka Dukore is a five-year activity providing 40,000 vulnerable youth with increased opportunities for wage and self-employment through a suite of interventions that includes market relevant work readiness training, employability skills training, work based learning, internship opportunities, links to employment, and entrepreneurship training at the youth level. The program builds on lessons learned from EDC's prior work in this area through the predecessor Akazi Kanoze Youth Livelihoods Project.

The HD program consists of a number of separate modules which are taken serially over the course of a year. The first of these is *Work Ready Now!*, which focuses on a combination of traditional business skills—such as basic accounting—and "soft" skills hypothesized to be both valuable and transferable across jobs and employment sectors (see Campos et al., 2017, for related evidence). This *Work Ready Now!* curriculum consists of eight sub-modules: Personal Development, Interpersonal Communication, Work Habits and Conduct, Leadership, Health and Safety at Work, Worker and Employer Rights and Responsibilities, Financial Fitness, and Exploring Entrepreneurship. It emphasizes soft skills and goal-setting, includes material on negotiation and group decision-making, covers workplace health issues including gender equality and HIV, and concludes with a discussion of the planning, marketing, and savings involved required to launch self-employment. This time-intensive module is taken by all students as the lead-in to the HD training, and consists of 10 five-day weeks of full-day training.

From here, students choose the additional modules and the employment sector of work in which they receive additional training, and the curriculum splits according to the nature of employment opportunities in local markets. In more urban areas, students would then move on to a Technical and Vocational Training module, Transition to Work programming, and Work Based Learning Services. While this is an option available to HD participants in our study, because our study areas are almost exclusively rural, HD instead encourages students to focus on self-employment. In this track, the next module of HD would be the 'Be Your Own Boss' training, which is an entrepreneurship curriculum that is tailored to the specific interests and opportunities in each cohort of students, and which lasts another 10 weeks. The curriculum for this component includes forming a business idea, identifying a practical business opportunity, outlining the details of business operations and financing, and establishing a formal business plan. Whereas the HD program had originally been designed to have these wage- and self-employment tracks be mutually exclusive, in practice participants were given the option of attending both if they wished. 80% of those assigned to HD participated in at least one of these components: 64% of the HD trainees enrolled in both the Build Your Own Business training and focused technical training in a specific work area, 11% enrolled only the former, and 4% only the latter. Nearly a half of all individuals assigned to the

HD arm completed the technical training workshops.

After completing their classroom training, HD students are typically placed in an internship or apprenticeship position with a local entrepreneur working in the selected employment sector. 39% of those in the HD arm undertook an apprenticeship during the study period, with the large majority of these in tailoring (53%) or hairdressing (22%). During this interval, students have regular check-ins with their trainers. Within a year of the initiation of training, students are considered "graduates" of HD.[6] This combination of several months of classroom training followed up by internships and workforce experience programs is typical of, if slightly more intensive than, comparable programs globally such as the *Jóvenes en Acción* program in Colombia (Attanasio et al., 2011).

Because the curriculum involves two substantial components of choice—whether to pursue vocational or small business training, and the employment sector in which to be trained—our experimental analysis will treat HD as a single intervention of which this choice is an integral component. We provide an analysis of the determinants of participation at each stage of the potential HD curriculum in Section 2.4.

GiveDirectly: Household grants program

To benchmark the impact of the HD program on cash, we worked with GiveDirectly, a U.S.-based 501(c)3 nonprofit organization. GiveDirectly specializes in sending mobile money transfers directly to the mobile phones of beneficiary households to provide large-scale household grants in developing countries including Kenya, Uganda, and Rwanda. GiveDirectly's typical model has involved targeting households using mass-scale proxy targeting criteria such as roof quality. GiveDirectly builds an in-country infrastructure that allows them to enroll and make transfers to households while simultaneously validating (via calls from a phone bank) that transfers have been received by the correct people and in a timely manner. Their typical transfers are large and lump-sum, on the order of $1,000, and the organization provides a programatically relevant counterfactual to standard development aid programs because it has a scalable business model that would be capable of providing transfers to the tens of thousands of households that are served by the HD program.

Since eligibility did not condition on having a cellphone, during the enrollment process individuals who did not themselves own a cell phone provided a number belonging to a trusted family member or friend, and transfers were sent to them through this intermediary. The payments were made to beneficiaries in two installments two months apart, with the first payment comprising 40% of the total to be paid to the beneficiary, and the second payment completing the transfer. After each payment is made, staff in the GiveDirectly call center team in Kigali contact every recipient to verify that payments have been received.

[6] Additional components of the broader HD curriculum include assisting students with access to finance through assistance in the formation of Savings and Internal Lending Communities and access to bank financing, and the use of a job matching resource that maintains a list of open positions and attempts to match graduates to them. These components of HD were not operative in the study districts at the time that we ran this evaluation. Also, HD typically includes prior beneficiaries of *Akazi Kanoze* but these individuals were not included in the study.

In terms of implementation timing, GD orientation commenced immediately after lotteries to notify youth randomized to receive a household grant and introduce them to the program. The value of household grants was not disclosed at the time of the lottery. GD treatment (where transfer values were disclosed to recipients) did not commence anywhere until the lotteries have been conducted everywhere in the district so as to avoid emphasizing the cash treatment prior to the completion of recruitment. A detailed timeline showing the relative timing of measurement activities and intervention implementation across the arms of the study is presented in Figure A.1.

2.2 Enrollment criteria

The study recruits youth from 13 geographic 'sectors' in the districts of Rwamagana, Muhanga and Nyamagabe.[7] Study participants had to be eligible for Huguka Dukore, to attend an informational session about Huguka Dukore, to enroll in a lottery to determine participation in that program following that informational setting, and to be traceable to a residence in a village in the sector where they were recruited. Attendance in person at the public lottery was not required for program enrollment. The study enrolled in its sample all individuals who met the criteria for treatment by Huguka Dukore in the study sectors. More detail on sample recruitment and the conduct of the lotteries is provided in Online Appendix D.

Table A.1 shows the process by which we moved from the original oversubscription universe to the final sample of 1,848 individuals deemed as fully eligible who were recruited into the study and randomized. Of 2,275 individuals who attended an orientation meeting and signed up for HD, 1,967 were found based on administrative review to meet the eligibility criteria. A further 119 could not be located either in the village of their stated residence, or were found to be resident outside the sector entirely, and consequently were deemed ineligible for intervention and the study.

There were no survey refusals at baseline, so our study sample reflects the full population of individuals who were assigned to treatments. The final study sample therefore consists of the universe of all individuals who met the enrollment criteria for Huguka Dukore, who attended an information session, agreed at that information session to be included in the assignment lottery, and were found resident in the relevant sector at baseline.

Demographic and employment characteristics (the latter of which will be defined in greater detail in Section 2.8 below) of the study participants are detailed in Table 3. Consistent with Huguka Dukore's 'soft' targeting criteria, the sample is 59% female with an average age of 23.5 (among the random sample assigned to control). They have an average of 7.6 years of education and typically live in households of approximately five individuals.

Although HD seeks to bolster employment opportunities for underemployed youth, it does not employ a hard criterion regarding employment status for eligibility. Consequently, it is not unusual for individuals to report that they are employed: 33 percent of (control-group) respondents reported being employed at baseline, using a definition that *excludes* agricultural work on a farm belonging

[7] In Rwanda, the *sector* is the geo-political unit below the district. There are 30 districts in Rwanda, and 416 sectors in total across those 30 districts.

to their own household (see Section 2.8 for more details). By endline the employment rate had risen to 48% in the control group.

Nonetheless, individuals in the study population are quite poor. 32% reside in households that the Government of Rwanda categorizes as Ubudehe I—its lowest socio-economic category, denoting a condition of 'extreme poverty'. Median consumption per adult equivalent is 5,879 RWF per month, which in 2018 PPP terms translates to a consumption level of USD 0.66 per day.

2.3 Assignment protocol

A public lottery was selected as the assignment mechanism for the study given the large sums of money being transferred and the desire by all parties to ensure that the assignment was considered fair and impartial by the research subjects. Participants drew their own treatment status as tokens of different colors from a sack, where each token corresponded to a given treatment arm and the number of tokens in the hat was determined by IPA according to the number of participants with fixed proportions assigned to each treatment. The proportion of individuals assigned to each treatment was fixed within each sector-level lottery, resulting in a standard block-randomized structure across the 13 blocks in the study.[8]

As illustrated in Table 1, this mechanism was used to assign individuals to one of four broad categories: a control arm, the HD program, the GD-administered cash transfers, or a *Combined* arm. In total, 485 individuals were assigned to HD, 672 to GD, 203 to the Combined arm, and 488 to control. Within the cash transfer arm, individuals were randomly assigned to the three bracketing transfer amounts (*GD-Lower*, receiving \$317.34; *GD-Middle*, receiving \$410.19; and *GD-Upper*, receiving \$503.04), or the (*GD-Large*) arm, receiving \$750.

The Combined arm received the HD treatment as well as the GD-Middle transfer. Both interventions were received at the same time as others in their same sector, meaning that they typically started the HD treatment several months before they would receive the household grant from GD. Receipt of cash transfers in the Combined arm was not made conditional on participation in HD training, a point which was emphasized at the lotteries.

Each program was described in detail at the lottery and representatives of both GD and HD (or its local partner) were present at every lottery. Individuals were notified whether they have been assigned to the GD, the Combined arm, or the HD arm at the time of the lottery. Individuals assigned to GD received tokens in a variety of colors which corresponded to different transfer amounts. The exact financial amounts were not discussed at the time of the lotteries. GD explained that youth randomized to GD would be contacted soon after the lottery to orient them to the program, and visited at their place of residence to undertake the enrollment process. Given the

[8] In the first phase of lotteries, comprising 792 study participants—we randomized purely at the individual level, as the study design did not anticipate multiple enrollees from the same household. In fact, the 792 participants in the first tranche of lotteries comprised 732 unique households. This resulted 34 households in which individuals in the same household were assigned to different treatments (at the level of the major arms of the study). Having recognized this issue, we altered the protocol in the second phase of lotteries and assigned treatment at the household level, where the 1,056 study members comprise 952 unique households. To reflect this issue we cluster standard errors at the household level.

public nature of the lottery assignment, the study was not blinded either to participants or to the survey firm. To avoid expectancy biases, we made it clear to the subjects at the time of the lottery that there would be no subsequent treatment by these implementers in the area.

2.4 Program Participation

Compliance with GiveDirectly treatment was nearly perfect. One individual in the middle GD arm was found to be ineligible by Ubudehe status and was not treated, and one individual assigned to the lower GD arm actually received the upper GD treatment. For GD the Intention to Treat (ITT) is therefore effectively the average treatment effect.

As anticipated, HD was most successful in achieving participation in its initial 10-week training (Work Ready Now, or WRN). 86% of the full HD treatment group (both HD-only and Combined arms) were counted as enrolled according to the contractual definition (attending the end of the first week of WRN training). This is the rate that the costing exercise uses since it alone determines the amount paid from USAID to the local implementing partner. But we can use institutional data from the HD program to examine participation in more detail.

Retention during the course of WRN is high; 79% of the overall sample completes this 10-week training program, which focuses on general workforce readiness. 69% of the of the HD sample complete the Build Your Own Business class (which is focused on entrepreneurship and self-employment); 13 individuals who did not take WRN did then go on to enroll in Build Your Own Business (BYOB). Finally, the Technical Training component of the HD intervention provides focused vocational instruction in a specific job sector, and was offered as a complement to BYOB. 48% completed the Technical Training component of the program. In the combined arm, participation with each of these components is about 5 pp higher than in the HD-only arm. Again, participation in Technical Training is not strictly confined to individuals who participated in any of the previous combinations of treatments.[9]

To understand the beneficiary characteristics associated with selection into the various components of HD, we regress participation in each component of the program on a battery of baseline characteristics, pooling the entire HD treatment group (HD-only and Combined). The results are presented Table A.3. In general compliance is relatively similar across observed beneficiary characteristics; older individuals are slightly more likely to complete the entrepreneurship training, but not technical training. There is a modest, negative association between working more at baseline and the likelihood of completing each stage of training; each additional hour of productive time use at baseline is associated with a reduction of approximately 0.4 percentage points in completion at all stages. Conversely, higher debt stocks at baseline are associated with greater completion rates. Both of these effects appear to be primarily driven by initial compliance, with measured attributes having little ability to predict subsequent compositional changes.

[9] The formality of the job sectors towards which the technical training is geared varies, ranging from formalized (hospitality) to quasi-formal (tailoring, hairdressing) to more agricultural forms of self-employment (poultry, pig rearing). Tailoring and poultry make up almost 75% of the trainings.

2.5 Cost Equivalence, Before and After the Fact

The costing exercise in the study utilized the 'ingredients method' (for more discussion, see Dhaliwal and Tulloch, 2012; Levin and McEwan, 2001; Levin et al., 2017; Walls et al., 2019). The policy question is asked from the perspective of the donor (in this case, USAID). To inform this objective, we conducted two different costing exercises at two moments in time. The ex-ante exercise, which was based on projected budgets and staffing costs, was used to predict the cost at the time of the study design and to choose the ranges over which GiveDirectly transfer amounts would be randomized. Then, a rigorous ex-post costing exercise was conducted for both programs once study implementation was complete, using actual budgets and expenditures.

Costing at Scale

Since the national HD program is eventually to cover twenty-three districts (e.g. much larger than the study population only) we attempt to cost the full national program (not just the study sample), inclusive of all direct costs, all indirect in-country management costs including transport, real estate, utilities, and the staffing required to manage the program, and all international operating costs entailed in managing the HD program.[10] Because we do not want differences in scale to drive differential costs per beneficiary, we asked GiveDirectly to artificially scale up their operations and provide us with numbers reflecting the costs per beneficiary if they were running a national-scale program across eight districts, including 40,000 beneficiary households like HD. This is the relevant scale for a USAID program officer contemplating commissioning a program to move the outcomes studied. Beneficiary identification costs, incurred partly by the survey firm and partly by HD, are calculated on a per-head basis and added to the costs of both implementers equally.[11]

We costed each GD arm separately, asking what the operating costs would have been if GD had run a national program at the scale of HD, giving only transfers of that amount. The ex-ante projected total cost of providing HD was \$452.47. The bracketing amounts were derived by supposing that the number of beneficiaries for the year two tranche of HD funding nationwide might vary between 8,000 and 12,000 beneficiaries, meaning that the per-capita cost would vary between \$377.05 and \$565.58. Note that because many of GD's costs are fixed at the individual level, the share of cost devoted to overhead falls as transfer amounts rise. This national-scale costing of each cash arm allows us to conduct the benefit/cost comparisons 'at scale', rather than having the artificial, multi-amount environment of the study contaminate the costing exercise across arms.

[10] Attributed shares of management and facilities costs are in proportion to the share of individually assignable activities in the total project budget; that is, we deduct a share of these central costs to reflect those HD activities that do not contribute to our estimates of benefits.

[11] This means that the operating costs for both implementers are slightly higher than they would have been absent the study-driven beneficiary identification costs, but these expenses drop out of the *comparative* costing analysis.

Differential Compliance

Given that the Intention-to-Treat is the heart of the experimental analysis, we construct the *realized cost per study subject* that corresponds to the spend on the sample over which the ITT is estimated. The raw costing returns the cost per beneficiary, but less than this is spent per study subject to deliver the ITT if compliance is less than complete. Both implementers face a relatively simple relationship between cost and compliance. For GD, individuals not treated cost nothing. Similarly for HD, their rules stipulate that they pay sub-implementers a fixed amount based on enrollment at the end of the first of WRN training to then follow through and offer all appropriate subsequent classes in the curriculum; compliance at this first hurdle therefore determines costs to USAID as well. These costs are almost exclusively based on offering the courses and do not scale sharply with class size. Hence, we consider all costs as 'averted' for non-compliers, and for each arm we calculate the ITT-comparable cost by multiplying the compliance rate times the cost per beneficiary.

Ex-post Costing Results

Table 2 shows the evolution of the costing analysis. As described above, the study design was built around the ex-ante HD cost figure of $464.25 per beneficiary.

Following the study intervention period, we undertake an ex-post costing exercise to determine actual expenditures on costed ingredients and the consequent at-scale costs to USAID of each study arm. These figures show that HD was less expensive than anticipated, and GD operating costs were slightly higher than anticipated. This means that the amount USAID spent per *beneficiary* was only $388.32, while the spending for the GD middle arm was $493.96. The inclusion of non-compliance further widens this gap, meaning that USAID cost per *study* household in the HD arm was $332.27, while in the GD arms it was $394.93, $490.99, $590.41, and $846.71, respectively. The combined arm, incorporating compliance with both components of the combined treatment, ended up costing USAID $840.20 per study individual, an amount similar to the GD Large arm. These are the numbers used in the cost equivalence analysis.

Because the costing exercise is undertaken from the funder's perspective, these estimates do not directly account for the opportunity costs of beneficiary time. Such opportunity costs are more obvious in the case of the HD training, given its classroom component, but the use of cash grants to start a business may involve forgoing leisure or other sources of income as well; such subtleties make a direct accounting for hours difficult to do in a way that is comparable across treatment arms. To the extent that such foregone income is capitalized in our measures of post-intervention savings and physical asset stocks, impacts on beneficiary time are captured in the benefit rather than cost side of our analysis.

In sum, our study ends up with even the smallest of the bracketing cash arms costing more than HD, while the GD Large arm provides a very close cost counterfactual to the Combined arm.[12] It

[12]The lower final costs arise primarily from two factors. First, compliance was lower than expected given that we were working with a group who had expressed willingness to participate in HD. Second, while the ex-ante costing had been based heavily on the spend in the study year and the number treated in the study year, the ex post

is important to remember in looking at our results, then, that the GD arms cost more than HD, and only through the linearity assumption in our cost-equivalence comparison can we recover the expected impacts of cash transfers at the cost-equivalent benchmarking amount. An implication for future work is that using wider brackets, and incorporating bounds on the take-up rate into those brackets, may be reasonable given the substantial uncertainty we have uncovered moving from ex-ante to ex-post cost estimates.[13]

2.6 Surveys and Outcome Measurement

Because we were interested in understanding both individual-level and household-level impacts, we used two distinct instruments within each round of data collection. A household survey was administered to the household head and a beneficiary survey was administered to the beneficiary. For beneficiaries who lived on their own or who headed their own household, these instruments coincided. We provide an overview of the contents of each instrument in Appendix Table A.2, and the construction of study indicators follows our Pre-Analysis Plan.

We have five primary outcomes for the study. Employment is a binary measure indicating that the individual spent more than 10 hours in the prior week in paid work or as the primary operator of a micro-enterprise. Productive Hours is the number of hours in the prior week spent in off-farm paid work or in micro-entrepreneurship. Both measures exclude own-farm agricultural work. Monthly income is the total amount earned over the prior month, including enterprise revenue. Productive asset stocks and household consumption per adult equivalent round out the primary outcomes.

Our secondary outcomes are divided into three groups. Additional measures of beneficiary welfare are subjective well-being and mental health, as well as the personal consumption of the beneficiary. Household wealth is measured using net non-land wealth, livestock wealth, and the stocks of savings and debt. Cognitive and skills dimensions are measured using Locus of Control, the Big Five index, as well as measures of the aspirations, business knowledge, and business attitudes of the beneficiary. Following our pre-analysis plan (AEARCTR-0004388), all monetary outcomes, both primary and secondary, are winsorized at 1% and 99% and measured in inverse hyperbolic sine (so that marginal effects can be interpreted as approximate percent changes).[14]

costing revealed a larger than expected share of costs in the early years of the HD budget being spent on curriculum development and implementer training. Because these costs are amortized over beneficiaries for the full five years of the program, they pushed down the spend per beneficiary in this early year of the study.

[13] An alternate possibility would be to conduct benchmarking using the Treatment on the Treated or Local Average Treatment Effects. While this is desirable from the costing side (given that cost per actual beneficiary is almost always easier to calculate) it has the downside of requiring the standard exclusion restriction assumptions on the outcome side for the LATE.

[14] We note that the share of 'zeros' among IHS-transformed primary outcomes at endline are moderate: 14, 31, and 0 percent respectively for income, productive assets, and household consumption. This puts them below the suggested rule-of-thumb threshold of one third zeros for use of the elasticity interpretation in Bellemare and Wichman (2020).

2.7 Attrition

We attempted to follow up with all study beneficiaries 18 months after baseline.[15] The tracking protocol for the post-treatment round was designed around the individual beneficiary, following him or her to whatever the relevant household was at that time (rather than tracking the baseline household). The interventions studied in this trial have the possibility of inducing migration; consequently it was particularly important to have a strategy to address attrition.

Our tracking strategy proceeded in two phases. First, we attempted to track all individuals who were still residing in any study district or in Kigali. Once we had completed this exercise we were left with 122 baseline individuals who we had not yet found. We then randomly sampled half of these individuals (blocking on treatment status), and began an 'intensive tracking' phase that spent substantial resources to track them wherever they had gone, including migrating out of the country, and survey them. This exercise resulted in IPA finding and surveying all 60 living beneficiaries in the intensive tracking sample (one had passed away). Given this remarkable rate of contact, we have an unusually convincing ability to correct for attrition by simply giving the intensive tracking sample weights of 2.

To verify that the data weighted in this way recovers the missing potential outcomes, we should establish whether the intensive sample that we drew was representative of the universe of early attritors. We can analyze this by a balance test of the intensive tracking sampling across the baseline outcome for all the early attritors. The sample for this is small (122) but in Table A.4 we find no evidence of systematic problems with this sampling (2 outcomes out of 20 unbalanced with p-values below 0.10 prior to correction for the False Discovery Rate, and none significant once we have corrected). These two pieces of evidence—representative sampling in intensive tracking and near-perfect tracking rate—suggest that we have an endline sample that is uniquely representative of the randomized universe.

2.8 Balance

The next step is then to establish whether the attrited and reweighted sample used for analysis was balanced at baseline. To ask this question, we estimate a balance table using baseline outcomes but only for the attrited endline sample, and with the weights, blocking, and clustering used in the endline analysis. This makes the balance test mimic the impact analysis we will run as closely as possible.

These results are presented in Table 3. The experiment appears well balanced (note that this is also the case if we simply use the full unweighted baseline sample), with rates of rejection consistent with random noise and none of the joint F-tests of all treatments indicating imbalance. We therefore proceed to the analysis of impacts with confidence that the study is internally valid.

[15] A 40-month endline is also planned for the study, but the advent of COVID in the interim will complicate interpretation of those results for "business as usual" conditions.

3 Results

3.1 Overall ITT Impacts

The data from the study are analyzed in a manner consistent with the design of a multi-arm, household-randomized program. Let the subscript i indicate the individual, h the household, and b the randomization block (lottery groups within which the randomization was conducted). For outcomes observed both at baseline (Y_{ihb0}) and at endline (Y_{ihb1}), we conduct ANCOVA analysis including the baseline outcome and fixed effects for the sector-level assignment blocks within which the randomization was conducted μ_b. We also include a set of baseline control variables, X_{ihb0}, selected from the baseline data on the basis of their ability to predict the primary outcomes, following the post-double-LASSO approach of Belloni et al. (2014b); further details of this procedure are provided in Online Appendix C. Base regressions to estimate the Intention to Treat Effect include indicators for the HD treatment T_{ihb}^{HD}, a vector of indicators for *each* of the three GD 'small' treatment values, T_{ihb}^{GDS1}, T_{ihb}^{GDS2}, and T_{ihb}^{GDS3}, an indicator for the GD 'large' treatment T_{ihb}^{GDL}, and an indicator for the combined arm T_{ihb}^{COMB}:

$$\begin{aligned} Y_{ihb1} &= \delta^{HD}T_{ihb}^{HD} + \delta^{GDS1}T_{ihb}^{GDS1} + \delta^{GDS2}T_{ihb}^{GDS2} + \delta^{GDS3}T_{ihb}^{GDS3} \\ &\quad + \delta^{GDL}T_{ihb}^{GDL} + \delta^{COMB}T_{ihb}^{COMB} + \beta X_{ihb0} + \rho Y_{ihb0} + \mu_b + \epsilon_{ihb1} \end{aligned} \tag{1}$$

Standard errors are clustered at the household level because the second tranche of treatment was assigned at the household level. For outcomes that are collected at endline only, we cannot include the lagged outcome to run the ANCOVA regression, and so use the simple cross-sectional analog to Equation (1).

To mitigate risks of false discovery across multiple outcomes and treatments, we report Anderson's 2008 False Discovery Rate to adjust p-values within each of the four relevant families (primary outcomes and the three families of secondary outcomes outlined in Section B.2), ensuring that the False Discovery Rate at the family level is controlled at five percent. This follows the procedures described in our Pre-Analysis Plan . For each point estimate we present both the unadjusted standard error (in parentheses) as well as the False Discovery Rate adjusted q-value [in brackets]. The stars on the coefficients are based on the adjusted q-values.

Tables 4 and 5 present the results of this analysis for primary and secondary outcomes, respectively. Rows of each table represent distinct outcomes, with columns presenting the corresponding estimates for impacts of the core treatment arms of the study: HD, each of the three smaller GD arms that were designed to be cost equivalent, the large cash transfer arm (GD-Large), and the arm that receives both the medium GD cash transfer amount and HD (Combined). Appendix Tables A.5 and A.6 present a more parsimonious, pre-specified specification that pools the three smaller GD transfer amounts into one arm. We focus discussion here on the specification disaggregating GD transfer values because the power gains from pooling are modest, because the average transfer cost exceeds the costs of HD, and because estimates in the pooled specification effectively provide an overview of the average effects across the smaller arms.

Beginning with Table 4, we see that none of the programs had statistically significant impacts on the core outcome of employment rates. Relative to a control group employment rate of 48%, neither HD, cash, nor the combination saw improvements of more than 5 percentage points. HD and the GD arms would have required point estimates of 6 and 10 percentage points, respectively, to reject the null of zero effect; these modest improvements are not close to being significant.[16] However, 95 percent confidence intervals for these estimates do not rule out impacts of a magnitude consistent with those observed in other evaluations of active labor-market programs (McKenzie, 2017). Moreover, as we will discuss in greater detail below, the absence of significant impacts on this omnibus employment indicator masks distinct patterns of occupational changes induced by the GD and HD treatments.

Moving to the continuous measure of the number of productive hours worked in a week, we see more promising impacts. Here, HD was successful in driving a 3 hour increase off a base of 18.4, an improvement of 16%. Over the 18 months since baseline the trend in the control group shows substantial increases in employment rates (from 33 to 48%) and productive hours (from 11 to 19 hours per week), and treatment effects relative to this large secular improvement are modest.[17]

The GD transfers appear to have had a non-linear effect on productive time use. The strongest effect on hours worked is found for the middle transfer (an increase of 6.5 hours per week). Both lower and upper transfers are substantially smaller in estimated impact—around half the magnitude—and not statistically significant, though we cannot reject equality between these arms and the middle transfer. On the other hand, the GD Large arm has an impact of only 1.1 hours relative to control, which is not only not statistically significant relative to control, but significantly different from the Middle arm: we reject equality of impacts between the Middle and Large cash transfer arms, with a p-value of 0.05. This is the first evidence of an apparent non-monotonicity in the impact of cash, suggesting that moderate transfers induce an increase in labor inputs, but once transfers become sufficiently large they begin to shift the opportunity cost of leisure, or into less labor-intensive occupations (as discussed below), discouraging labor inputs. This finding contrasts with the absence of labor-supply effects across seven cash transfer programs—with much smaller transfer sizes—studied by Banerjee et al. (2017).

The Combined arm, far from showing signs of a complementarity between the two programs, suggests that when individuals receive both cash and training there is no corresponding improvement in labor inputs, while had they received either alone a benefit would have been observed. In the following subsection we provide a more detailed analysis of the employment impacts of the

[16] In the specification pooling the three bracketing transfer sizes reported in Appendix Table A.5, the pooled estimate of these transfers is 0.03, with a standard error of 0.03.

[17] This pattern is consistent with an "Ashenfelter dip", a phenomenon long understood to exist in the evaluation of job training programs, where those interested in entering them have in many cases systematically experienced negative wage outcomes (Ashenfelter and Card, 1985). In the HD case, such a pattern in the control group may also be caused by the targeting of youth, who might be expected to be increasingly likely to be employed as they age. Either phenomenon would systematically bias pre-post comparisons, making RCTs of such programs particularly valuable.

different programs.

Next we consider the primary household outcomes of monthly income, productive assets, and consumption per capita. All of these outcomes are measured as the inverse hyperbolic sine, meaning that impacts can be interpreted as approximate percent changes.[18] Beginning with the impact of HD, we see a 31% increase in income (not significant), no meaningful change in consumption, but a highly significant surge in productive assets; these rise to be 151% higher than the control group mean. For a program that has made no material transfers to the beneficiaries, this improvement is substantial and impressive. When we look at the impact of cash transfers, we replicate the results of many other studies, showing that cash can lead to substantial improvements in these indicators. Measuring outcomes roughly 14 months after the transfers were made, we find the GD treatments lead to a doubling of monthly income, again non-monotonic with intermediate transfers appearing to be most effective. There is a quadrupling of productive asset values, and household consumption increases by 20-36 percent. For the Large arm (whose value received is 80% larger than the GD-Middle arm), we see outcomes that are in general very similar to the smaller arms; slightly smaller impacts on income, essentially identical impacts on productive asset values, and slightly larger effects on consumption. So again, these results are consistent with the idea that the Large arm induces a smaller productive effect than the Middle and Upper bracketing transfer amounts, leading to less earned income (see penultimate column of Table 4 for significance tests). Given this, the superior consumption for the Large arm appears to be being generated by the spending down of transferred resources. In all cases, the Combined arm has impacts that look relatively similar to the Middle arm (despite having cost almost twice as much), suggesting a lack of complementarity.

For the GD arms where compliance is perfect, the Intention to Treat measures effect of actually receiving the program, namely, the Treatment on the Treated (ToT). For the HD arm where the loosest measure of compliance is 85.6%, if we are willing to assume that those not participating received no indirect effect of being included in the treatment, then we can back out the ToT by dividing by the compliance rate. The resulting ToT estimate would 17% larger than the ITT for each variable, with the same significance level.

ITT effects on secondary outcomes

In Table 5 we present estimates of impacts on three families of secondary outcomes: beneficiary welfare, household wealth, and cognitive and non-cognitive skill development.

Beginning with beneficiary welfare, this outcome family includes measures of mental health and beneficiary-specific consumption. We have two ways of measuring mental health. The first of these is a composite of the answers to two simple Likert-scaled questions about subjective well-being, one on happiness, and one on life satisfaction. For this outcome, we see every arm improving subjective well-being, with effects in all cash arms being more than twice as large as the HD arm. Our second measure of mental health is built around reporting on a set of potential mental health

[18] Note that none of the recall windows include the period in which the transfers were received, and so our income measures are not mechanically picking up the receipt of the transfers themselves.

issues over the past two weeks, including stress, ability to concentrate, losing sleep, confidence, and feelings of worthlessness. Interestingly, none of the interventions improve this measure, suggesting that we are seeing improvements in general life satisfaction but no decrease in specific negative symptoms of poor mental health. Our final measure of beneficiary welfare is the inverse hyperbolic sine of the consumption of the specific beneficiary (as opposed to the primary outcome of household consumption). Here we see no improvements under HD, and increases of approximately 50% over the control outcome for any arm that receives cash (again larger for the middle-sized transfers).

Next we examine measures of household wealth. Perhaps unsurprisingly, cash transfers move these outcomes powerfully. Once again we see an impact that appears to be relatively homogeneous and somewhat non-monotonic in all arms that receive cash transfers, with net non-land wealth more than doubling in all arms but the smallest, livestock wealth tripling, and savings doubling. In Appendix Figure A.2 we plot the Cumulative Density Functions of savings across the different treatments. All the treatment distributions first-order dominate the control, implying treatment effects for savers of all levels. While coefficients on debt are negative for all cash arms, there is no evidence of significant pay-down of debt arising from cash transfers over this time. HD exhibits interesting impacts, with no change in the core measures of household wealth, but an increase both in savings (103% increase, significant at the 1% level) and debt (41% increase, insignificant).

Finally, we examine cognitive and non-cognitive skills. Using the standard Locus of Control index we find our first evidence of complementarities, in that only the Combined arm moves this outcome. None of the arms influence measures of Aspirations or the Big 5 index.[19] Our measure of business knowledge is a score on a set of questions built to reflect the HD curriculum, and hence serves as a manipulation check on their beneficiaries having learned what was intended. The results show that any arm receiving HD (HD or Combined) has about a half of a standard deviation improvement in their performance on this test, a very sizeable effect. GD does not move this score at all. The business attitudes measure we use captures attitudes toward entrepreneurship; while none of the interventions significantly improve it, they all lead to an increase of about 0.1 SD and, prior to the correction for false discovery the smaller GD arms (which saw a surge in productive hours), had an effect on this measure significant at the 10% level. The improvement in business knowledge for HD participants does not translate into an increase in entrepreneurship.

Detailed Analysis of Employment Effects

Given the central importance of employment in this study, we now seek to dig deeper into the differential impact of the programs. More nuance can be provided for these results in a number of dimensions. First, we can examine how the treatments shift the full distribution of time use by plotting the densities of productive time use across arms. Figure A.3 shows a modest rightward

[19] The original aspirations survey which we borrowed from Bernard et al. (2014) used the gap between desired future income/wealth standing and the current standing to measure aspiration; we found that our interventions had substantial positive impacts on the current economic standing at the time of the survey and little effect on desired future standing. This showed up as a negative treatment impact on aspirations; we therefore have deviated from our PAP and present the aspirations results using only the desired future economic standing.

shift in the CDFs (implying an increase in hours worked) across most of the distribution. None of the interventions moved hours for those who would have worked least, the Combined arm is most effective around the middle of the distribution, and the HD Main arm is most effective for those working the most hours. This speaks to the enabling effect of additional capital being particularly important for those already heavily engaged in productive work.

In Appendix Table A.7 we present a more detailed analysis using the five disaggregated categories in which we asked the underlying time use questions. These report the hours beneficiaries worked in the previous week in agricultural wage labor, non-agricultural wage labor, non-agricultural enterprises, agricultural self-employment, or agricultural processing and trading. Interestingly, the HD treatment induces an increase in non-agricultural wage labor of an estimated 6 percentage points (against a control-arm counterfactual of 30 percent), as well as a modest increase in agriculture-related self-employment, while the impacts of cash transfers are concentrated in non-agricultural enterprise and self-employment labor, at the expense of agricultural (and, at high transfer values, other forms of) wage labor.

Next, we delve into the source of the dissonance between our binary measure of employment, which shows no impact, and the continuous measure of productive hours, which does. The obvious suspect here is the specific threshold used to define 'employed'. To examine this, we vary this threshold continuously from 5 hours per week to 40 hours per week, and present a visualization of employment rates across thresholds and across treatment arms. We use the disaggregated labor categories analyzed in Table A.7, which are presented in the Appendix in decreasing order of their importance to total employment. First, Figure A.4 presents effects on non-agricultural wage labor (the category with the most overall time use). In general, the treatment effects represent similar vertical shifts across all the employment thresholds, implying that the binary estimate of employment effects would not be sensitive to threshold. HD exerts a positive effect, and GD-Large a negative effect, on wage employment. Non-agricultural self-employment, as shown in Figure A.5, responds little to HD, and to cash amounts in a relatively monotonic way (bigger transfers mean bigger impacts). Agricultural wage employment, which is both arduous and low-paid, appears to demonstrate an income effect in being depressed by any kind of cash transfer (Figure A.6).[20] Agricultural self-employment is increased by all the treatments, and it is here that the non-monotonicity in transfer amounts is strongest, with GD-Main seeing a larger effect than GD-Large (Figure A.7). Finally, agricultural processing and trading appears to be an employment sector that HD encourages individuals to *leave*, helping to explain smaller total productive hours effects despite the surge seen in non-agricultural wage employment and agricultural self-employment (Figure A.8). Given the general invariance of the impacts to the thresholds used, this suggests that the dissonance arises from the fact that the treatment effects on productive hours are occurring

[20]Using our own data to divide the amount earned by the hours worked for each of the five labor sectors, we find agricultural wage labor having by far the lowest pay, 10 cents per hour. The pay in non-agricultural wage labor is almost three times as high (26 cents/hour). The hourly wage rates we record in self-employment are even higher (36 cents for agricultural self-employment, 56 for non-agricultural self-employment, and 71 cents for agro-processing), but these should be interpreted with some caution due to the complexity of telling net from gross income in self-employment.

among those who would anyways have been counted as employed.

Finally, we can exploit a different source of data, which is a series of phone calls made by IPA to all of the study participants over the period from November 2017 (around the end of the HD intervention) to May 2019 (a few months before the midline). The core purpose of these phone calls was to acquire tracking information for the in-person survey, but we block randomized the month in which we called each individual, and during the call asked the same time-use questions that would be included in the midline. This survey was successful in reaching 1,797 of our study subjects (97%) and, given the randomization by month, provides a clean way of estimating the time path of treatment effects. In Figure A.9 we show coefficient plots of ITT regressions run separately within each month's sample of surveys. The GD-Main treatment has a consistent positive effect on time use across every monthly sub-sample, while the coefficients on the other treatments are more unstable across time. A core purpose of this phone survey was to see what happened to employment as the apprenticeship phase of HD came to an end. In Figure A.10 we can indeed see the elevated rate of apprenticeships for HD participants as program participation came to an end (late 2018), which then contextualizes the uptick in productive hours visible in the prior table for HD beneficiaries in the early months of 2019.

Summarizing this analysis, we see reasonable labor market effects of the HD program in a number of dimensions. We would not expect productive time use to improve *during* the intervention, and our midline survey appears to come about six months after the labor market impacts emerge. HD focuses on job training and on helping beneficiaries start agricultural enterprises, and it is successful in both these endeavors. Because most of the individuals entering this work are leaving agricultural wage employment or agricultural trading (which we also count as employment), the headline employment rate does not shift substantially. Cash transfers, on the other hand, particularly strongly drive individuals away from agricultural wage work and into self-employment (both agricultural and non-agricultural). So these two interventions have the effect of pushing participants down somewhat different paths towards income generation.

3.2 Cost-Equivalent Benchmarking

We pre-specified a regression adjustment strategy for benchmarking HD at an exactly cost-equivalent level using the ex-post costing data from both programs. First, begin with the total GD donor cost per study subject within each transfer amount arm, denoted by t_c (this is the final column in Table 2). Subtract from this number the benchmarked HD cost per household C from the same column, and denote the difference $t_c - C = \tau_c$; this is the deviation (positive or negative) of each GD arm from the benchmarked HD cost. Set τ_c to zero in the control and HD arms. We can then re-run regression (1) above omitting the combined arm, and controlling for a linear term in τ_c, a dummy for either treatment, and a dummy for receiving HD:

$$Y_{ihb1} = \delta^T T_{ihb} + \delta^{HD} T^{HD}_{ibh} + \beta X_{ihb0} + \rho Y_{ihb0} + \gamma_1 \tau_c + \mu_b + +\epsilon_{ihb1} \tag{2}$$

In this specification, T_{ihb} is a dummy variable indicating that individual i in household h of randomization block b was assigned to any treatment (HD or GD). Subject to the assumption of linear transfer expenditure effects, the slope coefficient τ_c captures impacts arising from deviations in GD cost from HD cost, the coefficient δ gives the impact of GD at the cost of HD, and the dummy variable δ^{HD} provides a direct benchmarking test: the differential impact of HD over GD at the same cost per eligible. We impose the simple linear functional form to preserve as much statistical power as possible for the core cost-equivalent benchmarking comparison, although it is straightforward to make this more flexible.

Table 6 shows the cost-equivalent benchmarking results for primary outcomes, and Table 7 for secondaries. Beginning with the third column, we linearize across transfer arms and test the marginal effect of receiving an additional \$100 from GD. Remarkably, across all 17 outcomes we do not have a single case where we can detect a significant slope across transfer amounts; indeed, more than half of the transfer slope coefficients are negative.[21] The takeaway seems to be that this study features transfer amounts large enough to have cleared an impact threshold beyond which additional amounts of money do not lead to meaningfully better outcomes. The second column, *Cost-equivalent GD impact*, reports the coefficient on an indicator for receiving any treatment. This estimates the effect of GD at the cost-equivalent level. Significant here are the outcomes already shown to be moved by GD (hours worked, income, productive assets, consumption, subjective well-being, livestock wealth, and savings).

The first column, *Differential impact of HD*, presents the core of the comparative cost effectiveness exercise that this study was built to conduct. Using the linearized cost adjustment, it gives the differential impact of HD relative to a cash transfer of precisely equal amount. The results indicate that HD leads to somewhat lower monthly income and substantially fewer productive assets than a cost-equivalent cash transfer. In terms of secondary outcomes, HD also does significantly worse at driving subjective well-being, beneficiary consumption, and livestock wealth; and of all the outcomes in the study, only for the Business Knowledge index does HD do significantly better than cash.[22]

Appendix Figure A.11 provides graphical interpretation for our cost equivalent comparison, using subjective well-being and business knowledge as outcomes. The cost adjustment strategy first fits a linear regression of the outcome by transfer amount in the GD arms (solid black line), then extrapolates this line to the predicted value that would have obtained had the GD transfers exactly

[21] This results stands in stark contrast to our previous benchmarking study (McIntosh and Zeitlin, 2019), in which the GD-Main cash transfer amounts were much smaller (\$84 on average) and a sharp differentiation was visible between these small transfers and the Large arm (\$532).

[22] Recall that we do not directly account for the opportunity cost of the time taken in training in the HD arm, which is substantial. Since we study outcomes many months after the training has concluded it may be reasonable to think that any lost productivity from time in class is reflected in the comparative stocks of assets and flows of consumption. Further, in Table A.7, we see that only 29% of individuals in the control reported working more than 20 hours per week, suggesting that substantial slack time existed and the opportunity costs of training time may not have been large. However, in our endline survey HD beneficiaries report spending an average of 23 hours per week in class while attending trainings, and using \$1.80 of their own money on travel and other expenses to attend class. To the extent that we have missed important costs that subjects experienced from HD participation this presumably tips the scales further in favor of cash programming, which does not impose these costs.

equalled the ex-post HD cost (hollow circle). Mapping the picture to the regression estimands, then, the third column gives the slope of the line, the second column gives the vertical differential between the predicted hollow circle and the control, and the core cost equivalent test measures the vertical difference between the hollow circle and the black diamond. Similar figures for all primary outcomes, along with a side-by-side comparison to a more standard cost-effectiveness approach, are included in Figure 2 and discussed in Section 4.

These results illustrate the value of the double counterfactual created by the cash benchmarking design. Compared to what would have happened in the absence of the program HD is successful, leading to meaningful improvements in a number of the core outcomes it was designed to move. Considering the cost of the program, however, and comparing to what would have happened if this cost had been distributed directly to beneficiaries, the picture is less rosy. Given that the direct distribution of these costs would have led to a surge in consumption and investment, the hurdle for success is raised and we find the HD program falling short across most outcomes. Unless policymakers had a strong preference for the specific human capital built by HD and measured in our Business knowledge index, over the 18 month time horizon the benefits of cash would dominate. It is nonetheless impressive that HD managed to generate meaningful improvements in productive assets, savings, and subjective well-being without having made any material transfers to beneficiaries.

A natural question is the extent to which linearity in the cost-equivalence adjustment may be driving our results. To interrogate this, tables A.8 and A.9 present the estimated cost-equivalence comparison using a variety of different functional forms to control for cost, for primary and secondary outcomes respectively. In each table, Column 1 present the base linear case from the prior tables. Column 2 uses a quadratic, and column 3 a third-order polynomial, functional form to control for cost. Columns 4–7 then serially drop one of the GD transfer amount arms and present the cost equivalence comparison if that arm had not been in the study.

For outcomes for which the linear specification is strongly significant (such as productive assets, subjective well-being, livestock wealth, and debt), results are very consistent across specifications. For monthly income the significance is sensitive, but the sign remains the same across specification. Relatively weaker comparative results (consumption and wealth prove to be substantially more sensitive. The benefit of HD in driving business knowledge retains a very similar point estimate across specifications but particularly the cubic cost control damages the power and hence significance of the estimate. Overall, the main comparative results of the study prove to be robust to the specification of the cost control structure.[23]

[23]These tables also provide a way of thinking about the power tradeoffs in the use of the linear specification for cost adjustment. One way of conceptualizing this is to compare the SE of the core impact estimates in Table 4 to Table 6. The latter are 12-40% larger across the primary outcomes, giving a direct measure of the power loss from the extrapolated counterfactual relative to one that is directly experimental. This comparison captures the power cost of ex-ante uncertainty about the true cost of delivery. However, when we examine Table A.8, we see that the standard errors on the quadratic specification are typically 50% higher, and the cubic specification almost four times as high, as the linear specification. Hence, given the need to make some kind of cost adjustment after the fact, the linear approach taken here appears a reasonably well-powered bias-variance compromise.

3.3 Complementarities

Complex, multi-dimensional programs are often justified on the grounds that poverty presents a range of constraints, meaning that individuals are unable to benefit unless more than one constraint is relaxed at once. Our study design allows a canonical statistical analysis of such complementarities. In addition, it further allows us to understand whether evidence for those complementarities (or their absence) arises because of non-constant returns to resources directed at individual beneficiaries, or due to the complementary nature of the specific resources transferred—cash or training—per se. Our evidence suggests that the apparent absence of complementarities in a canonical test arises due to diminishing returns to resources in general: the canonical test confounds complementarities with resource intensity. An alternative, cost-equivalent test of complementarities suggests positive evidence of technological complementarities between cash and kind.

Given that the Combined treatment is comprised of the GD Middle transfer and the HD treatment, our ITT estimates (Equation 1) nest a canonical 2×2 cross-cutting design. This provides a natural test of complementarities: in the presence of positive complementarities, the impact of the Combined arm (who received HD and the middle GD transfer value) will be greater than the sum of the GD Middle and HD arms' impacts. We derive the difference between the Combined arm and the sum of the GD Middle and HD arms, and report the p-value from an associated F test for equality, in Complementarity Test (a) of Tables 8 and 9 for primary and secondary outcomes, respectively.

Across the board, the evidence from this canonical test in Tables 8 and 9 suggests that—far from finding positive complementarities—the whole appears to be less than the sum of the parts. The sign on the core test (Column a) is negative for all primary outcomes and for two thirds of secondary outcomes. The complementarity is significantly negative for productive hours and subjective well-being. Outcomes that are driven by cash are not further helped by HD, outcomes driven by HD are not further helped by cash, and for two of our key outcomes the combination is actually worse than what we would expect from the independent effect of the two programs.

The fact that the GD Large and Combined arms turn out to have virtually identical cost allows us to make a second form of cost-equivalent comparison between training and cash. Conditional on receiving the GD Middle transfer, are individuals made better off by having the HD training added on top of that, or having the cost of the HD training added to the cash transfer they are already receiving? If strong complementarities were present there would be a logic for adding this extra cost over a basic cash transfer in the form of training rather than further money. Column (b) of Tables 8 and 9 compares the impacts in these two arms. Noting that these arms are both relatively small and hence this comparison is less well powered than others presented in the paper, the only significant comparison is for business knowledge, where those in the Combined arm learn in a manner very similar to the HD-only arm but very different from those who only receive cash. There is suggestive evidence, not quite significant at the 10% level, that the Combined arm is superior at driving beneficiary consumption, subjective well-being, and the Locus of Control.

How can we reconcile the absence of complementarities in the canonical test and the superior impacts of cash over training with the fact that the Combined arm appears to be a better use of resources than the Large cash transfer? The canonical complementarities test implicitly embeds returns to scale since the combined arm costs more than either independent arm. One explanation for a lack of evidence of complementarities would simply be diminishing returns to the global scale of investment; we have already seen evidence of diminishing marginal returns to cash in both the comparisons of cost-effectiveness ratios between GD Lower and GD Large (Tables 4 and 5, column b) and in the slope of impacts with respect to incremental transfer values (Tables 6 and 7). The comparison of the Combined arm and the GD Large arm presented in the prior paragraph holds scale constant but does not isolate complementarities.

We can combine cells from our rich experimental design to illustrate the role of complementarities and diminishing returns in the relative performance of the Combined interventions versus GD Large in a manner that is, to our knowledge, unique to the literature. Specifically, to do so we form a kind of cost-equivalent difference-in-differences examining the comparative benefits of cash vs training over the top of GD Middle, as compared to cash versus training over the top of nothing. In the terms of Equation (1), this amounts to a test for the null of equality between two differences: $\delta^{COMB} - \delta^{GDL} = \delta^{HD} - \delta^{GDS1}$. The comparison between these two differences arises from the joint product of complementarities (on the left) and differential marginal returns of additional cash (above GDS2 on the left, and above zero on the right). Because each half of the equation is internally cost equivalent, we rigorously hold scale constant while examining the differential effectiveness of training in the presence and absence of cash transfers.

Complementarity test (c) of Tables 8 and 9 presents estimates of this difference-in-differences measure of complementarities, and associated p-values from an associated F test. Results are strikingly different than the canonical complementarities test: we find weakly positive estimates for all primary outcomes and for eight out of ten secondary outcomes. Estimated effects are statistically significant at the 1 percent level for productive assets, beneficiary-specific consumption, and household livestock wealth. A base transfer of cash appears to close the gap in impacts between cash and training.

This finding allows us to make a more nuanced set of benchmarking comparisons. Given the strong returns to moderate cash transfers, cash proves superior to training across a relatively broad set of outcomes starting from zero. The effects of HD, however, appear to be only very weakly lower over the top of a base cash transfer, while further cash above this base is quite cost-ineffective. So if the question on the table is 'how should we concentrate a large amount of funding on one individual', this combination of results says that HD over the top of base cash benchmarks more favorably against cash.

3.4 Analysis of Heterogeneity

We analyze four baseline dimensions over which we pre-committed to estimate heterogeneous treatment effects. These are gender, household consumption per capita, risk aversion (measured by the choices in a Binswanger–Eckel–Grossman lottery) and local labor market conditions (the employment rate within each of the 67 'cells' of the study).[24] We use a standard interaction between treatment indicators and baseline characteristics to study heterogeneity across these baseline covariates, and examine only primary outcomes. The covariates are demeaned prior to interaction so that the uninteracted coefficient should be interpreted as impact at the mean of the interaction variable. To facilitate interpretation of the distributional incidence of outcomes at endline across these baseline attributes, we omit both the baseline outcome (ANCOVA) and also the LASSO-selected covariates that are included but not reported for all the prior regressions; this allows us to interpret the descriptive variation between baseline attributes and outcomes as reflecting the bivariate correlation between these variables, rather than reflecting a partial association.

Results of subgroup analyses by gender, risk, consumption, and cell-level employment shares are presented in Appendix Tables A.10, A.11, A.12, and A.13, respectively. Overall, we find very little evidence of meaningful heterogeneity in the impact of the program across these four dimensions. Gender itself has a huge effect (female beneficiaries are less likely to be employed, put in fewer productive hours, and have lower levels of income and assets), but none of the interventions affect women in a manner significantly different than men. It is important to recognize, however, that power starts to become a greater concern as we split the study into smaller cells. Reading point estimates, we see that in fact HD does lead to a 9 pp increase in employment rates among men, while women see a 3 pp *deterioration* in employment rates during HD; however, neither the effect for men nor the differential effect for women is statistically significant. Table A.11 shows the risk averse being somewhat better at translating the interventions into income, with few other differences. Encouragingly, Table A.12 shows impacts that are relatively invariant to baseline consumption, indicating that both cash and HD are equally effective for the very poor (the GD-Large arm has slightly larger benefits, and the Combined arm slightly larger benefits, for those with higher consumption at baseline). Given that many related programs have had an easier time creating benefits for the non-poor, this suggests that both of the interventions studied here should be considered good candidates for heavily poverty-targeted programs. Baseline local employment rates, studied in Table A.13, are not only not driving the impacts of the program but appear to be completely uncorrelated with outcomes overall. Finally, based on feedback from the Rwanda USAID Mission, we included age as a dimension of heterogeneity, and again in Table A.14 find no evidence of differences. The bottom row of all of these tables shows the p-value on F-tests of joint significance across the four interaction terms (note these are based on the non-FDR-adjusted significance rates); we present 25 such omnibus tests and find two of them to be significant at the 10% level, fewer than we would expect by random chance.

Hence, taken together, the analysis of heterogeneity suggests programs that are having consistent and similar effects across different types of beneficiaries and across local labor market condi-

[24] We filed our initial PAP with the AEA and then subsequently submitted to the *Journal of Development Economics'* pre-results publication facility. In the time between these two submissions we came to realize that the data from the Convex Time Budget exercise we conducted at baseline (Andreoni and Sprenger, 2012) had not produced meaningful results. As a result we dropped the analysis of heterogeneity using discount rates and hyperbolicity derived from the CTB, and present here only the heterogeneity tests included in the later pre-results document accepted by the JDE.

tions. An important caveat to this result is that our study selection criteria are stringent (imposing a relatively narrow age range, education level, poverty status, as well as a stated desire to engage in job training) and hence may have generated a sample that has unusually little internal variability.

3.5 Spillovers

Spillovers are of central interest in this project for several reasons. First, for both of the programs being studied here, recent literatures suggest that we should be concerned with impacts on non-beneficiaries. Crépon et al. (2013) show that most of the benefits of job training programs in France come from diverting a fixed set of job opportunities towards treated individuals and away from untreated ones. Cash transfer programs appear to have complex spillovers on non-beneficiaries, with Angelucci and De Giorgi (2009) and Egger et al. (2019) finding potential *positive* spillovers through family or labor market mechanisms, and other studies suggesting negative spillovers to non-beneficiaries (Haushofer and Shapiro, 2016), particularly in terms of mental health (Baird et al., 2013). Because our study uses an individually randomized design, these spillovers are a direct threat to internal validity, hence this test is critical.

We look for spillovers both on program participation and on primary outcomes. While in principle there may be externalities of each program at several levels of contact, we focus on spillovers that are *local*, in the sense that they occur between individuals who reside in the same village at baseline.[25] The reason for doing so is both substantive—this is plausibly the level at which such interactions are most salient—as well as practical: since the randomization is blocked at the sector level, it provides no variation in treatment saturation at that or higher levels.

Let T_{ivb}^{w} denote the assignment of individual i in *village* v to treatment $w \in \{\text{GDM}, \text{GDL}, \text{HD}\}$, where we pool the three smaller cash-transfer values into a single arm, $w = \text{GDM}$, as distinct from the larger transfer value, $w = \text{GDL}$. Individuals in the combined arm have $T_{ivb}^{GDS} = T_{ivb}^{HD} = 1$. Define $T_{ivb} = [T_{ivb}^{GDM}, T_{ivb}^{GDL}, T_{ivb}^{HD}]$ as the vector denoting individual i's treatment status. Finally, we let the vector $\bar{T}_{-i,vb} = [\bar{T}_{-i,vb}^{GDM}, \bar{T}_{-i,vb}^{GDL}, \bar{T}_{-i,vb}^{HD}]$ denote the average treatment status of study individuals *other than* individual i in village v for each of the three treatments (that is, the saturation of each treatment among others in the village), and we adopt the convention that $\bar{T}_{-i,vb} = 0$ if individual i is the only study participant in village v. This vector of village-level saturations is randomly assigned through the household-level lottery, and is independent of the own-treatment terms because we calculate saturations among others in the village. The densities of the treatment saturations are plotted in Figure A.12.

[25] We conducted a social network survey measuring connections to other individuals in the study at baseline that we had intended to use for this analysis, but we found that a) networks within villages are typically completely connected, and b) we were unable to collect this data for a small subset of beneficiaries. Because the simple treatment saturation in the village maps almost perfectly to the saturation in the social network and is universally observable, we adopted the village connection metric in our pre-analysis plan.

Spillovers on Compliance

Using this notation, we can represent the three types of spillover analysis conducted, in increasing order of complexity. First, we analyze whether there are spillover effects on the rate at which individuals choose to participate in Huguka Dukore (this question is not interesting for GD because compliance is so close to universal). A major concern during the study design phase was that the assignment of one's peers to cash would discourage participation with HD. To do this, we use only the HD arm and estimate the following linear probability model:

$$\mathrm{E}[P^w_{ivb}|\bar{T}_{-i,vb}] = \mu^w_b + \phi^w \bar{T}_{-i,vb} \tag{3}$$

where P^W_{ivb} is a measure of individual i's participation in treatment $w \in \{HD, Combined\}$.

Table 10 illustrates that the density of GD treatment in a village does not drive compliance. The first column analyzes compliance within the standalone HD arm, and the second column within the Combined arm. The point estimates on the GD saturation rates are zero or positive, and never significant. Instead, this table shows that HD compliance is driven by the *Huguka Dukore* treatment saturation; the point estimate in the standalone HD arm is significant at the 5% level and suggests that as an HD participant goes from having no one else in the village treated to everyone else in the village assigned to that treatment, we can expect compliance to increase by 27 pp. Since HD in the absence our of experiment would naturally attempt to treat 100% of the willing and eligible individuals considered in this study, our study features a compliance rate that may be slightly too low relative to the rate that would naturally occur. However, since the costing estimate is multiplied by the observed compliance rate and the ITT is similarly a function of observed compliance in our study, compliance falls out of the benefit/cost comparisons that we make (it is in both the numerator and the denominator). This suggests that our estimates are still likely to be informative about treatment effects at alternative saturation levels, absent substantial compositional effects on essential heterogeneity in the sense of Heckman and Vytlacil (2005). Tests below for interference on downstream outcomes will help to rule out such heterogeneity.

Spillovers on Outcomes

Next, we can take our study of spillovers to the primary outcomes used in the study. We first use a more parsimonious specification that looks for average spillover effects *from* each type of treatment saturation on the other members of the village. To do so, we modify the specification of equation used to estimate ITT effects (Equation 1) for outcome Y_{ivb1} as follows:

$$Y_{ivb1} = \delta_1 T_{ivb} + \delta_2 \bar{T}_{-i,vb} + \beta X_{ivb0} + \rho Y_{ivb0} + \mu_b + \varepsilon_{ivb1}. \tag{4}$$

Table 11 conducts this analysis and uncovers no evidence of spillovers in outcomes; not only are none of the saturation rates for any of the three treatments significant, but the sign of the coefficients alternate signs across outcomes for all three treatment saturations. As would be expected in a study for which this is not the primary question, these spillover effects are less precisely estimated than the direct effects of each treatment: typically, standard errors associated with the saturation effect are approximately twice those of standard errors for the direct effect. But on outcome dimensions where HD or GiveDirectly have direct effects that are statistically distinguishable from zero, we are able to rule out spillover effects of a magnitude large enough to fully offset these impacts among the eligibles.

Finally, we use the full model from our PAP that allows for the estimation of spillovers both *from* each treatment arm, and *on to* each treatment arm:

$$Y_{ivb1} = \delta_1 T_{ivb} + \delta_2 \bar{T}_{-i,vb} + \delta_3 T_{ivb} \bar{T}_{-i,vb} + \beta X_{ivb0} + \rho Y_{ivb0} + \mu_b + \varepsilon_{ivb1}. \tag{5}$$

In Equation (5), the three coefficients in δ_2 provide a test of the spillover effects of each of the three treatments onto control individuals, and the nine coefficients in the vector δ_3 test for whether the spillover effects *from* the saturations of any of the three treatments *on to* individuals directly receiving each of the three treatments differ from the control.[26]

The five Appendix Tables A.15, A.16, A.17, A.18, and A.19 present this analysis. Because of the large number of hypotheses being tested, we present the analysis for each outcome in a separate table. Each of these analyses shows results from a single regression with the three different sets of treatments by saturation interactions stacked as adjacent columns. Again, this analysis is remarkably clear and consistent in showing an absence of spillover effects. Using significance levels derived from within-regression false discovery corrections, we do not have a single significant spillover effect in the control, or differential effect for any treatment, across any of the 60 spillover tests performed here. Using unadjusted p-values, we find 4 of these comparisons significant at the 10% level and none at the 5% level, in line with random chance. While this spillovers analysis has limited power compared to the main impact estimates, a simple way to illustrate that we see little evidence of strong contamination is as follows. For outcomes where we have significant direct effects, such as the productive assets outcome shown in Table A.18, our point estimates for the zero-saturation effect, δ_1 of GD Main (point estimate 4.35) exceed the magnitude of the minimum detectable level of interference (this is approximately 2.8 times the standard error on parameter δ_3) for all but the interference effect of GD Large. In such cases, we are powered to rule out forms of interference that were sufficiently large to offset the zero-saturation direct treatment impact.

Figure 1 provides a graphical take on this analysis, showing the predicted outcome for each treatment group (in rows) and primary outcome (in columns). Using the observed treatment saturations and the estimated saturation slopes, this exercise predicts the outcome we would expect to see within each arm and outcome as the local intensity of treatment changes, with the specific treatment saturations generating the spillovers plotted as different colored fitted lines in each graph. The only visual signs of spillovers are restricted to the GD-Large arm (which in reality has a

[26]We report cluster-robust standard errors for each of the coefficients, clustering at the village level. In addition, given the large number of hypotheses tested in these regressions (sixteen) we correct the p-values in these regressions using Anderson's (2008) False Discovery Rate correction across all coefficients within each regression.

limited variation in the saturations), while the HD and GD Main arms which are the core of the benchmarking exercise appear completely invariant to local intensity of treatment.

It therefore appears that we can quite simply conclude that our study did not generate detectable within-village spillover effects on outcomes. While our blocked lotteries do not allow us to study the kinds of cross-village General Equilibrium (GE) effects generated in Egger et al. (2019), the sparse geographic distribution of our study sample implies the total value of our transfers are small relative to local GDP. It appears unlikely that we would have generated meaningful local GE effects. We do not see the kind of zero-sum diversionary treatment effects uncovered by job-training programs in more formalized labor markets (Crépon et al., 2013), presumably because our overall impact on employment was small, and because the informal labor markets in this context do not demonstrate a hard capacity constraint in the same way. The invariance of outcomes to HD saturations suggests that the response of HD compliance to HD saturations is not generating a study ITT that is driven by the lower treatment saturations induced by the lottery. Comfortingly, the overall takeaway is that our analysis remains a valid estimate of treatment effects that appear homogeneous across the support of saturation levels in the study, and there is no need to attempt to correct for spillovers.

3.6 Tracing cash flows

In this question, we ask a question specific to the cash-transfer arms of our study: *How have participants used the funding they received?* Tracing these cash flows facilitates understanding in several ways. First, variation in the use of funds by the amount of cash received, and by the provision of HD training alongside cash in the combined arm, may help us to understand the mechanisms by which these arms deliver distinct impacts. Second, the extent to which we can account for the full values of transfers received sheds light on whether there might be 'missing' dimensions of impact not accounted for in our evaluation. And third, the extent to which observed patterns of spending are consistent with a simple spend-down of cash grants—and not a set of investments likely to deliver future income gains—is indicative of the sustainability of the impacts of cash.

Accounting for cash flows requires us to model both the inflows and the outflows induced by each active treatment arm. Clearly, the core of the induced inflow is the value of cash itself received by the beneficiary. But we also need to account for other income gains that each intervention can cause. To do so, we estimate each arm's impacts on beneficiary income and on transfers *received* by the beneficiary household, and we include these as inflows. We then compare these impacts on inflows with estimated impacts on expenditures, where our expenditure measures are a mixture of flows—household consumption, and transfers and loans made to other households—as well as stocks—savings, debt, livestock, and other productive assets. Flow measures vary in their recall periods, with income and consumption measured over comparatively short recall periods (as analyzed in Section 3.1, these are constructed as estimates for the month prior to the follow-up survey) and transfers measured over a 12-month recall period.

We value impacts on each of these dimensions by multiplying estimated *proportional* impacts of each program, from the specification in equation (1), by the average *level* of the corresponding outcome at follow-up in the control group. We emulate the results of Section 3.1 for household transfers in Appendix Table A.23. Together with the ITT results on primary and secondary outcomes, and combined with estimates of control means, this allows us provide the estimated financial impact of each program on both inflows and outflows.

The results of this exercise are presented in Table 12. In Panel A of that table, we estimate impacts of each cash-transfer arm on households' total income. This comprises not only the direct value of transfers received, but also induced increases in beneficiary income and in transfers received from other households. Since beneficiary income is measured over a short recall, we need to make an assumption about its time path, and so we extrapolate over the 12 months since the cash transfer assuming a constant impact in all months. Whether this under- or over-states true financial inflows will depend, among other things, on whether beneficiary income measures represent true income effects or the spend-down of business stocks. We compare these impacts on cash inflows with financial outflows, measured in Panel B. Those outlays include measures of flows including household consumption, impacts on which are extrapolated as constant over the period since the cash transfer, as well as loans and transfers made by the household, which are measured with a 12-month recall, and so do not need to be extrapolated. They also include values of stocks at follow-up of livestock, other productive assets, savings, and debt. Our total outflow measure is constructed as the sum of the financial value of impacts on each of these categories, with impacts on debt entering negatively. We then construct the share of cash inflows that can be accounted for by these measured dimensions as the ratio of total treatment-induced outflows to total treatment-induced inflows for each arm.

We draw three basic conclusions from this exercise. First, the fact that induced outflows constitute a large share of induced inflows in general suggests that our measures of financial assets and expenditures are relatively complete, at least as far as is relevant to the impacts of these transfers. Second, expenditure patterns do exhibit some differences across arms. For example, investments in livestock and in other productive assets do not rise proportionally with the value of transfers. Further, comparison between expenditure patterns in the Middle and Combined arms—for which cash transfer values are equal—suggests that there are modest differences in the application of cash induced by HD training. The Combined treatment seems to divert the flow of cash to livestock and productive assets to a greater extent than the Middle arm, at the expense, in part, of debt repayment. While these investments have not delivered increases in income at the time of follow-up, it suggests that there remains a possibility that the Combined arm will cause differences in incomes as these investments deliver returns over the long term.[27] Third, while our ability to draw inferences about the sustainability of cash-transfer impacts beyond the follow-up period is necessarily speculative, there are two features of this exercise that point to sustainability. First, the consump-

[27]The differences in investments induced appear too small to have created meaningful differences in income over the period studied. For example, if the *differential* livestock investment in the Combined arm versus the Middle arm, of $262-$218=$44, was undertaken immediately after the transfer and paid an annuity value of 5 percent, the resulting difference in incomes would have been just over $2 in total over the year since transfers occurred.

tion impacts of the cash transfers are smaller than their income impacts in absolute terms. And to the extent that a rapidly declining consumption path over the period prior to follow-up period would have been indicative of a lack of sustainability, we note that under the current (generous) assumption about total impacts on inflows, there is little unaccounted-for expenditure that could have been part of a downward-sloping consumption pattern over time.

More detail on the question of 'where did the money go' is provided in Table A.20. This analysis sums across all enterprises reported on in the beneficiary survey (which encompass both household and beneficiary-owned businesses) to provide richer detail on entrepreneurship impacts. The average control individual reports 1.4 businesses operating, a number which is not changed by the HD treatment but increases by about .5 for any arm including cash. Interestingly, all interventions display an increase in own hours worked in entrepreneurship (as has been shown using other survey questions), but here we also see a corresponding increase in the reported participation of *other* household members induced by all of the treatments, particularly those involving large cash transfers. This provide evidence of complementarities between physical capital and labor crowding in household effort beyond the beneficiary him/herself, and helps to explain what appear to be sustained income increases at the household level.[28] Customers and sales rise across the board, but only significantly so in the arms that include cash. Self-reported profits rise strongly across all arms, with a bump of $6.22 per month from HD and more than $12 per month in the larger of the small GD arms. Given that the profits from entrepreneurship alone (extrapolated over 12 months) make up more than half of the income increases and the lion's share of the consumption increases reported across arms in the prior table, these results help to flesh out the story of how the large observed increases in productive assets translate into improvement in economic well-being.

4 Value for Money

There are multiple ways that one can pose that most basic question in cost effectiveness: how can policy spending achieve the greatest effect? Our study is designed to emphasize one comparison, namely the *cost equivalent* one: if a comparable amount of money is to be spent per beneficiary across programs, which achieves the greatest benefit? This approach holds both the beneficiary pool and the spend per beneficiary fixed, and asks about comparative effectiveness.

A related but different question can be asked if one is willing to concentrate spending on a subset of the beneficiary pool. This is *comparative cost effectiveness*: how can money be spent to create the largest total benefit across the pool for a fixed overall budget? In Tables 4 and 5 when we test for differences in the ratio of the effect sizes to the cost of the arm, this is the question we are asking.

A visual comparison between the cost equivalence and the cost effectiveness approaches is provided in Figure 2. In the left-hand column of figures, we illustrate the cost-equivalent comparison for the five primary outcomes. Here, the question is focused on a specific point on the x-axis, namely

[28]In more detailed analysis not shown, we find these increases in other household labor to be entirely confined to the new businesses started by beneficiaries, and do not take place in pre-existing household-owned enterprises.

the cost of HD (the black diamond), and the GD arms are pooled to estimate one counterfactual, which is the predicted value at the HD cost. The Control need not even be included to execute this comparison.

The right-hand column of Figure 2 plots the identical outcomes by arm, but uses them to instead represent the cost effectiveness question in a visual way. The term *value for money* can be seen in the slope of the line relative to the counterfactual in benefit-cost space; the value generated in improved outcomes (y axis) over the money spent (x axis). Hence, the steepest slope of the line between the Control outcome (where no money was spent) and the outcome in each active treatment arm in this space represents the highest value for money (assuming that the donor is indifferent to reassigning benefit across individuals)[29]

In our study, the cost-equivalent question can be posed at one cost by design (the ex-post cost of HD, compared to the extrapolated benefit of GD at that cost), and at another cost by accident (because the Combined and GD-Large arms turn out to have virtually identical costs). The comparative cost-effectiveness question, on the other hand, is easily posed across costs since it essentially compares the gradient of additional money at each modality and total cost. Comparative cost effectiveness is therefore an attractive way to horse-race the GD arms against each other.

Given these two potential approaches, what does our study say about value for money? Beginning with the central cost-equivalent comparison between HD and cash, the main results of the paper suggest that for the $332 price tag of HD, one could produce significantly better outcomes via cash transfers across productive assets, beneficiary consumption, and livestock wealth, and for this amount of money HD is superior only at producing business knowledge. This result can help policymakers with different objectives think about which type of intervention will best produce the outcomes that they want to see.

Appendix Tables A.21 and A.22 provide estimates of cost effectiveness for primary and secondary outcomes, respectively, by dividing the ITT estimates by the cost of the arm in hundreds of dollars, and testing for the differences across arms (this is the slope coefficient visually represented in the right-hand panel of Figure 2). Seen in comparative cost effectiveness terms, the GD-Middle transfer appears to be superior overall. It has the highest benefit/cost ratio across four of the five primary outcomes, losing only to the GD-Lower arm on productive asset value. Among secondary outcomes, GD Middle wins or ties in terms of cost effectiveness for subjective well-being, beneficiary consumption, and net household wealth. HD has higher cost effectiveness in driving savings, and for business knowledge.

When we make the cost-equivalent comparison between the Combined arm and the GD Large arm, we are asking the following question: given that we have already spent the $494 dollars to deliver the GD Main arm, is it then better to spend another $350 to deliver HD on top of that cash, or should that additional spending be used to amplify the cash transfer? Here, the answers

[29]In the absence of general equilibrium effects, all points along these lines connecting average outcomes in active treatment arms with the control group are achievable by mixing that treatment with a fraction of individuals left untreated. In this sense, there is a lesser degree of extrapolation required for comparisons based on cost effectiveness than for those based on cost equivalence.

look quite different. There is now no outcome for which the GD Large arm is significantly better than Combined, and the Combined arm continues to demonstrate HD's advantage at producing business knowledge. Why these divergent results? The answer appears to be evidence of diminishing marginal benefits of cash; because HD is effective at driving many of the economic outcomes and additional cash is having a weaker effect, we find these two interventions to have similar cost-equivalent benefits even for outcomes on which cash 'beat' HD starting from zero. In spite of the greater incremental impact of HD over and above a base transfer, however, even in this context HD delivers a smaller benefit than cost-equivalent cash delivers relative to the counterfactual of no intervention. The fact that, in general, the less expensive interventions produce superior outcomes in cost-effectiveness terms implies that we uncover no evidence to justify concentrating spending per person above the levels seen in the GD Middle arm.

5 Conclusions

This study undertakes an exercise increasingly called for in recent years: namely, a direct comparison of in-kind development aid to cash. The comparison program is a well-established workforce readiness intervention, and the beneficiary group is one in which both interventions have straightforward pathways to generate long-term improvements in welfare. Our design uses randomized variation in cash transfer amounts to identify cost-equivalent comparisons.

Our findings on the impacts of Huguka Dukore contribute to the evidence base on supply-side active labor market programs, of which it is typical in design. Huguka Dukore appears successful in increasing beneficiaries' productive time use, as well as the stock of productive assets that they accumulate. 18 months after program assignment, its participants report higher levels of subjective well-being, demonstrate improved business knowledge, and have increased their stocks of savings considerably. HD has no significant effects on employment rates or beneficiary incomes. Given a control-group endline average income of approximately USD 17.45 per month, our point estimates on HD's impacts on beneficiary income imply that approximately 61 months of earnings at this rate would be required to pay back the realized costs to beneficiaries, consistent with supply-side labor market policies elsewhere (McKenzie, 2017). But looking beneath the surface, we do see exploratory evidence of shifts to microenterprise and even non-agricultural wage labor, which may portend long-term labor-market impacts.

On the other hand, cash transfers to this population appear to have moved a wide range of outcomes, including productive hours, incomes, productive assets, and household consumption. These impacts are substantial for beneficiaries and provide a meaningful return on the costs of intervention: for example, the cost to USAID of the middle transfer would be recuperated in beneficiary income impacts alone after approximately 26 months. Secondary measures including subjective well-being, household wealth, and savings are all meaningfully moved by these transfers. Several lessons about the design mix of transfer programming are evident: there is no evidence of complementarities between these cash-transfer impacts and the skills provided by HD, and increases in transfer size above the middle transfer amount of $411 included in our study do not appear justified in cost-effectiveness terms. If resources are to be added beyond the cost of the GD Middle transfer, our results suggest that it is weakly preferable to provide these in the form of training rather than adding to the amount of cash provided.

It is worth noting a number of study limitations. First, the final costing number is lower than anticipated ex-ante, meaning that the cost equivalent analysis must extrapolate to a cost lower than the costs of cash transfers observed in the GD arms. Second, in order to achieve the benchmarking, we confine both implementers to somewhat unnatural sample selection rules. An organization providing cash transfers would never have a reason to target only individuals who express interest in a training program. For HD, the study constrained them to study only poor individuals (they do not normally use Ubudehe status as a targeting criterion). We examine only the individually assignable aspects of HD, and we miss any environmental benefits to the employment landscape caused by HD's capacity building and job placement work. Also, because we induced HD to treat at a lower intensity than they normally would (they would typically have treated *all* the individuals in our study), we may not have captured the effect of a program running at greater intensity. Nonetheless, the internal variation in our sample suggests that these issues would generate limited bias in our study: outcomes are flat across transfer amount in GD (costing error), impacts are homogeneous across a range of beneficiary characteristics (differential targeting rules), and we find no evidence of saturation effects (HD treatment intensity). It therefore appears likely that our study has reasonable external validity.

The comparative evaluation of cash and in-kind modalities in this study—together with experimentally induced variation in cash transfer sizes—allows us to speak to the cash-benchmarking question increasingly called for in recent years. How would beneficiary outcomes change if a standard and widespread development intervention were simply distributed to the beneficiaries in the form of a mobile money transfer? Proponents of cash transfers have suggested that they should be considered the 'index funds' of international development, meaning a benchmark to which other programs are compared (Blattman and Niehaus, 2014). A wide range of programs may have economic welfare as their ultimate objective. But cash benchmarking provides the most policy-relevant counterfactual when the programs share an overlap in their target populations and immediate outcomes, and drive private benefits (as opposed, for example, to transport infrastructure or security/policing expenditure). Huguka Dukore's aims of improving productive employment, assets, income, and consumption, together with accumulated evidence suggesting that cash transfers might also impact these outcomes for a comparable population of youth, make this case well suited for benchmarking against the impacts of cash transfers.

We estimate that at transfer values that are cost-equivalent from the donor perspective, the impacts of cash transfers exceed that of HD by a statistically significant margin on two of five primary outcomes: monthly income and productive assets. On the other hand, given the substantially superior performance of HD at producing the human capital measured by the business skills index, we can provide an exact exchange rate that quantifies the tradeoffs policymakers must be willing to make. Focusing on the comparative impacts on productive assets versus business skills, to prefer HD at 18 months policymakers must be willing to forego a 200% increase in beneficiary productive assets in order to obtain a 0.5 SD increase in human capital.

This approach to cash benchmarking holds donor expenditure per beneficiary constant—in effect, restricting comparisons of in-kind programs to cash transfers that could reach an equivalent number of beneficiaries with equal-sized transfers. This need not be the optimal intensity of transfers. A donor who seeks to maximize an additive social welfare function in any of the outcomes considered, for example, would prefer to choose programming that maximizes the benefit-cost ratio. Summarizing across outcomes, the middle of the cash transfer sizes considered in this study seems to perform best by this metric.

One of the most surprising results of our study is the apparent lack of complementarities between HD's training and the cash transfers in a canonical two-by-two comparison. We designed the Combined arm to provide the cleanest test of structural complementarities between physical and human capital from the beneficiary side, which required that we implemented each arm precisely as in the standalone case. In reality, however, these two programs could be interwoven in a number of deeper ways. At the very least, regular programming intended to add capital over training would typically only do so at the end of the training, while our Combined arm received their cash midway through HD. More fundamentally, one might think about making the cash component conditional on participation in HD, which we did not do. So our results should not be taken to suggest that it is possible to design cash and training programs in a complementary manner, but rather than simply providing them together does not automatically generate a whole greater than the sum of the parts, and that the diminishing returns to total investment in individual beneficiaries can undermine the value of such an approach. Given the many complex, expensive, and multi-dimensional programs pushed by the donor community, this result is worth paying attention to. Moreover, our study's ability to make cost-equivalent comparisons between the Combined arm and the Large transfer shows that, even under the current intervention design, what drives the absence of complementarities between these programs may not be the absence of coincident human capital and liquidity constraints per se, but rather the diminishing marginal returns to resources invested in any one individual.

The lessons of this study add to the evidence base not just on specific programs, but on the relative importance of the underlying constraints they seek to address. HD does appear to have been effective in alleviating a human capital constraint: it improves business knowledge, though non-cognitive dimensions such as aspirations and beneficiaries' self-efficacy appear to have been harder to move. In this context, the returns to raising business knowledge appear not to have been as impactful as the returns to alleviating liquidity constraints through cash transfers. And the liquidity constraints faced by these individuals appear not to be exceedingly large, given that the returns to increases in cash-transfer size are modest. A companion cash-benchmarking study in Rwanda (McIntosh and Zeitlin, 2019) that targeted households with malnourished children also focused on Rwandans within Ubudehe 1 and 2, and so provides an interesting point of comparison to the results here. That study was generally unable to reject constant benefit/cost ratios comparing transfers around \$100 to transfers of \$500, while here with bigger and more evenly spaced transfer amounts we find that transfers larger than USD 400 have limited additional value. This helps to identify the 'sweet spot' for cash in this context. Moreover, the present study shows that the impact of human-capital and liquidity improvements flow through different channels, in particular as the HD program does move beneficiaries in the direction of wage employment, whereas cash transfers support self-employment. How these different mechanisms shape the long-run impact of interventions that provide skills versus those that provide cash remains an important question.

References

Aguero, Jorge, Michael Carter, and Ingrid Woolard, "The impact of unconditional cash transfers on nutrition: The South African Child Support Grant," 2006.

Ahmed, AU, JF Hoddinott, S Roy, E Sraboni, WR Quabili, and A Margolies, "Which Kinds of Social Safety Net Transfers Work Best for the Ultra Poor in Bangladesh," *Operation and Impacts of the Transfer Modality Research Initiative*, 2016.

Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney, "The long-run impact of cash transfers to poor families," *The American Economic Review*, 2016, *106* (4), 935–971.

Alcid, Annie, "A randomized controlled trial of Akazi Kanoze youth in rural Rwanda," Report submitted to USAID. Waltham, MA: Education Development Center 2014.

Alfonsi, Livia, Oriana Bandiera, Vittorio Bassi, Robin Burgess, Imran Rasul, Munshi Sulaiman, and Anna Vitali, "Tackling youth unemployment: Evidence from a labor market experiment in Uganda," London School of Economics, STICERD Working Paper number 64 2019.

Anderson, Michael L, "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*, 2008, *103* (484), 1481–1495.

Andreoni, James and Charles Sprenger, "Estimating time preferences from convex budgets," *American Economic Review*, 2012, *102* (7), 3333–56.

Angelucci, Manuela and Giacomo De Giorgi, "Indirect effects of an aid program: how do cash transfers affect ineligibles consumption," *American Economic Review*, 2009, *99* (1), 486–508.

Araujo, M Caridad, Mariano Bosch, and Norbert Schady, "Can Cash Transfers Help Households Escape an Inter-Generational Poverty Trap?," in "The Economics of Poverty Traps," University of Chicago Press, 2017.

Ashenfelter, Orley and David Card, "USING THE LONGITUDINAL STRUCTURE OF EARNINGS TO ESTIMATE THE EFFECT OF TRAINING PROGRAMS," *The Review of Economics and Statistics*, 1985, *67* (4), 648–660.

Attanasio, Orazio, Adriana Kugler, and Costas Meghir, "Subsidizing vocational training for disadvantaged youth in Colombia: Evidence from a randomized trial," *American Economic Journal: Applied Economics*, 2011, *3* (3), 188–220.

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.

__ , **Craig T McIntosh, and Berk Özler**, "When the money runs out: Do cash transfers have sustained effects on human capital accumulation?," *Journal of Development Economics*, 2019, *140.*

__ , **Jacobus De Hoop, and Berk Özler**, "Income shocks and adolescent mental health," *Journal of Human Resources*, 2013, *48* (2), 370–403.

Balboni, Clare, Oriana Bandiera, Robin Burgess, Maitreesh Ghatak, and Anton Heil, "Why do people stay poor?," Technical Report, Technical Report, London School of Economics and Political Science, London ... 2019.

Banerjee, Abhijit V., Rema Hanna, Gabriel E. Kreindler, and Benjamin A. Olken, "Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs," *World Bank Research Observer*, 2017, *32* (2), 155–184.

Barham, Tania, Karen Macours, and John A Maluccio, "Schooling, Learning, and Earnings: Effects of a 3-Year Conditional Cash Transfer Program in Nicaragua After 10 Years," *La Plata, Argentina: Centro para los Estudios Distributivos, de Trabajo y Sociales. Disponible en: http://www. cedlas-er. org/sites/default/files/aux_files/barham-maluccio-macours_. pdf*, 2014.

Bellemare, Marc F and Casey J Wichman, "Elasticities and the inverse hyperbolic sine transformation," *Oxford Bulletin of Economis and Statistics*, 2020, *82* (1), 50–61.

Belloni, A, D Chen, V Chernozhukov, and C Hansen, "Sparse models and methods for optimal instruments with an application to eminent domain," *Econometrica*, November 2012, *80* (6), 2369–2429.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "High-dimensional methods and inference on structural and treatment effects," *The Journal of Economic Perspectives*, 2014, *28* (2), 29–50.

__ , __ , **and** __ , "Inference on treatment effects after selection among high-dimensional controls," *Review of Economic Studies*, 2014, *81*, 608–650.

Bernard, Tanguy, Stefan Dercon, Kate Orkin, Alemayehu Taffesse et al., *The future in mind: Aspirations and forward-looking behaviour in rural Ethiopia*, Centre for Economic Policy Research London, 2014.

Blattman, Christopher and Laura Ralston, "Generating employment in poor and fragile states: Evidence from labor market and entrepreneurship programs," Working paper, University of Chicago 2015.

__ **and Paul Niehaus**, "Show them the money: why giving cash helps alleviate poverty," *Foreign Affairs*, 2014, *93*, 117.

_ , **Nathan Fiala, and Sebastian Martinez**, "The long term impacts of grants on poverty: 9-year evidence from Uganda's Youth Opportunities Program," Technical Report, National Bureau of Economic Research 2018.

Bongaarts, John, "Development: Slow down population growth," *Nature News*, 2016, *530* (7591), 409.

Brudevold-Newman, Andrew Peter, Maddalena Honorati, Pamela Jakiela, and Owen Ozier, "A firm of one's own: experimental evidence on credit constraints and occupational choice," World Bank Policy Research Working Paper no. 7977 2017.

Campos, Francisco, Michael Frese, Markus Goldstein, Leonardo Iacovone, Hillary C. Johnson, David McKenzie, and Mona Mensmann, "Teaching personal initiative beats traditional training in boosting small business in West Africa," *Science*, September 2017, *357* (63357), 1287–1290.

Card, David and Alan B Krueger, "Time-series minimum-wage studies: a meta-analysis," *American Economic Review*, May 1995, *85* (2), 238–243.

Casey, Katherine, Rachel Glennerster, and Edward Miguel, "Reshaping institutions: Evidence on aid impacts using a preanalysis Plan," *Quarterly Journal of Economics*, 2012, *127* (4), 1755–1812.

Chowdhury, Reajul, Elliott Collins, Ethan Ligon, and Kaivan Munshi, "Valuing Assets Provided to Low-Income Households in South Sudan," 2016.

Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora, "Do labor market policies have displacement effects. Evidence from a clustered randomized experiment," *The quarterly journal of economics*, 2013, *128* (2), 531–580.

Cunha, Jesse M, Giacomo De Giorgi, and Seema Jaychandran, "The price effects of cash versus in-kind transfers," *Review of Economic Studies*, forthcoming.

Dhaliwal, Iqbal and Caitlin Tulloch, "From research to policy: using evidence from impact evaluations to inform development policy," *Journal of Development Effectiveness*, 2012, *4* (4), 515–536.

Diaz, Juan Jose and David Rosas, "Impact evaluation of the job youth training program Projoven," Technical Report, IDB Working Paper Series 2016.

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael W Walker, "General equilibrium effects of cash transfers: experimental evidence from Kenya," Technical Report, National Bureau of Economic Research 2019.

Fernald, Lia CH, Paul J Gertler, and Lynnette M Neufeld, "10-year effect of Oportunidades, Mexico's conditional cash transfer programme, on child growth, cognition, language, and behaviour: a longitudinal follow-up study," *The Lancet*, 2009, *374* (9706), 1997–2005.

Fox, Louise and Upaasna Kaul, "What Works for Youth Employment in Low-Income Countries?," 2017.

_ , Lemma W Senbet, and Witness Simbanegavi, "Youth employment in Sub-Saharan Africa: challenges, constraints and opportunities," *Journal of African Economies*, 2016, *25* (suppl_1), i3–i15.

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), 164–192.

Giné, Xavier and Ghazala Mansuri, *Money or ideas? A field experiment on constraints to entrepreneurship in rural Pakistan*, The World Bank, 2014.

Government of Rwanda and World Bank Group, "Future Drivers of Growth in Rwanda: Innovation, integration, agglomeration, and competition," International Bank for Reconstruction and Development / The World Bank 2019.

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.

Heckman, James J and Edward Vytlacil, "Structural equations, treatment effects, and econometric policy evaluation 1," *Econometrica*, 2005, *73* (3), 669–738.

Hidrobo, Melissa, John Hoddinott, Amber Peterman, Amy Margolies, and Vanessa Moreira, "Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador," *Journal of Development Economics*, 2014, *107*, 144–156.

Hoddinott, John, Susanna Sandström, and Joanna Upton, "The impact of cash and food transfers: Evidence from a randomized intervention in Niger," 2014.

Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond, "Long-run impacts of childhood access to the safety net," *The American Economic Review*, 2016, *106* (4), 903–934.

Jones, Damon, David Molitor, and Julian Reif, "What do workplace wellness programs do? Evidence from the Illinois Workplace Wellness Study," *The Quarterly Journal of Economics*, 2019, *134* (4), 1747–1791.

Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Christopher Udry, "Agricultural decisions after relaxing credit and risk constraints," *Quarterly Journal of Economics*, 2014, *129* (2), 597–652.

Kluve, Jochen, Susana Puerto, David Robalino, Jose Manuel Romero, Friederike Rother, Jonathan Stoeterau, Felix Weidenkaff, and Marc Witte, "Interventions to improve the labour market outcomes of youth: A systematic review of training, entrepreneurship promotion, employment services and subsidized employment interventions," *Campbell Systematic Reviews*, 2017, *13* (1), 1–288.

Leroy, Jef L, Paola Gadsden, Sonia Rodríguez-Ramírez, and Teresa González De Cossío, "Cash and in-kind transfers in poor rural communities in Mexico increase household fruit, vegetable, and micronutrient consumption but also lead to excess energy consumption," *The Journal of nutrition*, 2010, *140* (3), 612–617.

Levin, Henry M and Patrick J McEwan, *Cost-effectiveness analysis: Methods and applications*, Vol. 4, Sage, 2001.

_ , _ , Clive Belfield, A Brooks Bowden, and Robert Shand, *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis*, SAGE publications, 2017.

Manley, James, Seth Gitter, and Vanya Slavchevska, "How effective are cash transfers at improving nutritional status?," *World development*, 2013, *48*, 133–155.

McIntosh, Craig and Andrew Zeitlin, "Benchmarking a Child Nutrition Program against Cash: Experimental Evidence from Rwanda," Technical Report, Working Paper 2019.

McKenzie, David, "How effective are active labor market policies in developing countries," World Bank, Policy Research Working Paper 8011 3 2017.

_ , "Small business training to improve management practices ind eveloping countries: re-assessing the evidence for 'training doesn't work'," *Oxford Review of Economic Policy*, 2021, *37* (2), 276–301.

Mel, Suresh De, David McKenzie, and Christopher Woodruff, "One-time transfers of cash or capital have long-lasting effects on microenterprises in Sri Lanka," *Science*, 2012, *335* (6071), 962–966.

Molina-Millan, Teresa, Tania Barham, Karen Macours, John A Maluccio, and Marco Stampini, "Long-Term Impacts of Conditional Cash Transfers in Latin America: Review of the Evidence," Technical Report, Inter-American Development Bank 2016.

Pega, Frank, Stefan Walter, Sze Yan Liu, Roman Pabayo, Stefan K Lhachimi, and Ruhi Saith, "Unconditional cash transfers for reducing poverty and vulnerabilities: effect on use of health services and health outcomes in low-and middle-income countries," *The Cochrane Library*, 2014.

Republic of Rwanda, "Seven Years Government Programme: National Strategy for Transformation, 2017–2024," Office of the Prime Minister, Government of Rwanda September 2017.

Samuels, Fiona and Maria Stavropoulou, "'Being Able to Breathe Again': The Effects of Cash Transfer Programmes on Psychosocial Wellbeing," *The Journal of Development Studies*, 2016, *52* (8), 1099–1114.

Schwab, Benjamin et al., "In the form of bread? A randomized comparison of cash and food transfers in Yemen," in "Agricultural & Applied Economics Association 2013 AAEA & CAES Joint Annual Meeting" 2013, pp. 4–6.

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

Seidenfeld, David, Sudhanshu Handa, Gelson Tembo, Stanfield Michelo, Charlotte Harland Scott, and Leah Prencipe, "The impact of an unconditional cash transfer on food security and nutrition: the Zambia child grant programme," 2014.

Skoufias, Emmanuel, Susan W Parker, Jere R Behrman, and Carola Pessino, "Conditional cash transfers and their impact on child work and schooling: Evidence from the progresa program in mexico [with comments]," *Economia*, 2001, *2* (1), 45–96.

Walls, Elena, Caitlin Tulloch, and Christine Harris-Van Keuren, "Cost Analysis Guidance for USAID-Funded Education Activities," Technical Report 2019.

Young, Alwyn, "Consistency without inference: Instrumental variables in practical application," September 2019. Unpublished, London School of Economics.

Tables and Figures

Table 1: Study Design

Sector	Control	Huguka Dukore	GiveDirectly				Combined
			317.16	410.65	502.96	750.30	HD + 410.65
Kaduha	63	60	21	21	22	22	26
Kibumbwe	32	37	10	10	12	13	13
Kigabiro	14	12	4	5	4	5	5
Kiyumba	17	17	6	6	6	6	8
Mugano	51	51	18	18	18	18	22
Muhazi	39	40	13	19	13	18	17
Munyaga	34	34	10	10	10	12	14
Munyiginya	25	25	8	8	8	10	10
Musange	30	29	10	10	10	9	12
Mushishiro	24	23	6	6	6	9	8
Nyakariro	49	50	16	17	19	17	22
Nyarusange	57	54	21	20	19	19	24
Shyogwe	53	53	18	18	18	20	22
Total	488	485	161	168	165	178	203

Note : This table gives the number of study individuals assigned to each treatment arm in each of the 13 sectors within which lotteries were conducted. The lotteries were blocked so that fixed fractions of individuals are assigned to each arm.

Table 2: Results of Costing Exercise

Treatment Arm:	Ex Ante Cost	Value received	Ex Post Cost	Fraction operating cost	Compliance Rate	Cost per study household
Huguka Dukore	\$464.25	\$153.47	\$388.32	60.5%	85.6%	\$332.27
GD lower	\$377.03	\$317.16	\$394.39	19.6%	100%	\$394.39
GD mid	\$464.25	\$410.65	\$493.96	16.9%	99.4%	\$490.99
GD upper	\$571.74	\$502.96	\$590.41	14.8%	100%	\$590.41
GD large	\$828.47	\$750.3	\$846.71	11.3%	100%	\$846.71
Combined	\$928.5	\$561.11	\$885.64	36.3%	89.6%(HD), 100%(GD)	\$840.20

Note: The first column shows the ex-ante costing data on which study was designed; the core number is the HD cost around which the GD actual transfer amounts in column 2 were designed. Column 3 shows the results of the ex post costing exercise. Column 4 provides the share of spending that did not reach the beneficiaries either in cash or in direct training and materials costs. Column 5 shows the compliance rates, and since all costs are averted for non-compliers then the final column shows the final cost per study subject for each arm that are the basis of the cost-equivalent comparisons.

Table 3: Descriptive statistics and balance

		GiveDirectly					Control			
	HD	Lower	Middle	Upper	Large	Combined	Mean	Obs.	R^2	p-value
Ubudehe category I	0.01	0.00	0.07	0.01	0.01	−0.03	0.32	1720	0.07	0.73
	(0.03)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Beneficiary female	0.01	−0.02	0.03	−0.02	0.02	−0.05	0.60	1770	0.04	0.68
	(0.03)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Beneficiary age	−0.21	−0.41	−0.12	−0.66	0.43	−0.33	23.58	1770	0.03	0.12
	(0.23)	(0.31)	(0.34)	(0.32)	(0.32)	(0.31)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Beneficiary years of education	0.11	0.10	−0.03	0.07	0.03	−0.14	7.55	1770	0.07	0.91
	(0.15)	(0.22)	(0.21)	(0.20)	(0.21)	(0.19)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Household members	−0.32	−0.36	−0.03	0.00	−0.07	−0.32	4.98	1766	0.03	0.26
	(0.16)	(0.24)	(0.33)	(0.20)	(0.22)	(0.19)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Employed	0.04	−0.00	−0.03	0.01	0.04	0.03	0.33	1770	0.02	0.73
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Productive hours	0.45	−1.17	0.19	2.17	0.44	−0.13	10.81	1770	0.02	0.88
	(1.28)	(1.68)	(1.82)	(2.04)	(1.65)	(1.54)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Monthly income	0.07	−0.03	−0.24	0.08	0.11	0.18	4.37	1770	0.01	0.99
	(0.33)	(0.47)	(0.46)	(0.46)	(0.47)	(0.42)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Productive assets	−0.56	−0.47	−0.10	−0.30	−0.43	−0.13	2.49	1770	0.03	0.56
	(0.28)	(0.38)	(0.39)	(0.40)	(0.40)	(0.38)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
HH consumption per capita	−0.12	−0.11	−0.08	−0.10	−0.19	−0.01	9.46	1766	0.05	0.37
	(0.07)	(0.10)	(0.10)	(0.10)	(0.09)	(0.09)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Beneficiary-specific consumption	−0.07	0.07	−0.03	−0.17	0.11	0.01	7.53	1770	0.03	0.93
	(0.15)	(0.19)	(0.21)	(0.22)	(0.18)	(0.20)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
HH net non-land wealth	−0.05	0.35	0.15	−0.17	1.13	−0.22	10.53	1766	0.03	0.19
	(0.46)	(0.54)	(0.63)	(0.70)	(0.47)	(0.60)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Savings	−0.26	−0.48	−0.34	0.03	−0.09	0.13	8.01	1770	0.04	0.82
	(0.30)	(0.41)	(0.44)	(0.39)	(0.39)	(0.38)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Debt	0.12	−0.14	−0.30	0.03	0.15	0.75	7.84	1770	0.02	0.47
	(0.32)	(0.45)	(0.46)	(0.44)	(0.45)	(0.40)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
HH livestock wealth	0.29	−0.18	0.20	0.24	−0.26	−0.17	7.32	1766	0.03	0.93
	(0.40)	(0.57)	(0.58)	(0.55)	(0.56)	(0.52)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Business Knowledge	−0.01	0.10	−0.01	−0.08	−0.03	0.09	0.00	1770	0.03	0.57
	(0.07)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				

Notes: Table presents control means and standard deviations; regression coefficients and standard errors for associated comparisons, and p-value for a test of the hypothesis that all arms pool. Regression-based comparisons and associated hypothesis tests based on a regression with block indicators. ***, **, and * denote statistical significance at the 1, 5, and 10 percent levels, respectively. All continuous variables winsorized at top and bottom 1 percent. Inverse hyperbolic sine transformation taken for monthly income, household consumption, beneficiary expenditure, savings, debt, and wealth variables.

Table 4: ITT estimates, primary outcomes, separating GD transfer values

		GiveDirectly					Control			p-values		
	HD	Lower	Middle	Upper	Large	Combined	Mean	Obs.	R^2	(a)	(b)	(c)
Employed	0.02	0.03	0.05	0.00	0.01	0.01	0.48	1770	0.16	0.95	0.57	0.94
	(0.03)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)						
	[0.30]	[0.30]	[0.16]	[0.50]	[0.46]	[0.49]						
Productive hours	2.79*	2.76	6.54***	3.56	1.12	2.31	18.64	1770	0.19	0.82	0.33	0.63
	(1.57)	(2.34)	(2.40)	(2.52)	(2.06)	(2.03)						
	[0.07]	[0.16]	[0.01]	[0.12]	[0.33]	[0.16]						
Monthly income	0.31	0.76**	1.08***	1.14***	0.73**	1.04***	8.05	1770	0.21	0.31	0.24	0.41
	(0.26)	(0.36)	(0.34)	(0.35)	(0.35)	(0.32)						
	[0.16]	[0.04]	[0.00]	[0.00]	[0.04]	[0.00]						
Productive assets	1.54***	3.94***	3.80***	3.84***	4.02***	4.42***	5.61	1770	0.26	0.00	0.00	0.42
	(0.35)	(0.46)	(0.50)	(0.46)	(0.47)	(0.44)						
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]						
HH consumption per capita	0.05	0.20**	0.27***	0.23***	0.36***	0.27***	9.46	1737	0.33	0.12	0.67	0.31
	(0.06)	(0.08)	(0.09)	(0.07)	(0.07)	(0.07)						
	[0.21]	[0.01]	[0.00]	[0.00]	[0.00]	[0.00]						

Note: The six columns of the table provide the estimate on dummy variables for each of the treatment arms, compared to the control group. The five primary outcomes are in rows. Regressions include but do not report the lagged dependent variable, fixed effects for randomization blocks, and a set of LASSO-selected baseline covariates, and are weighted to reflect intensive tracking. Standard errors (in parentheses) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. Reported p-values in final three columns derived from F-tests of hypotheses that cost-benefit ratios are equal between: (a) GD Lower and HD; (b) GD Lower and GD Large; and (c) GD Large and Combined treatments. Employed is a dummy variable for spending more than 10 hours per week working for a wage or as primary operator of a microenterprise. Productive hours are measured over prior 7 days in all activities other than own-farm agriculture. Monthly income, productive assets, and household consumption are winsorized at 1% and 99% and analyzed in Inverse Hyperbolic Sine, meaning that treatment effects can be interpreted as percent changes.

Table 5: ITT estimates, secondary outcomes, separating GD transfer values

		GiveDirectly					Control			p-values		
	HD	Lower	Middle	Upper	Large	Combined	Mean	Obs.	R^2	(a)	(b)	(c)
Panel A. Beneficiary welfare												
Subjective well-being	0.19***	0.40***	0.53***	0.48***	0.55***	0.41***	0.00	1770	0.13	0.08	0.12	0.17
	(0.07)	(0.09)	(0.10)	(0.09)	(0.09)	(0.09)						
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]						
Mental health	−0.04	−0.06	0.07	0.03	0.11	0.12	0.00	1770	0.07	0.89	0.22	0.93
	(0.07)	(0.09)	(0.09)	(0.09)	(0.10)	(0.09)						
	[0.30]	[0.27]	[0.27]	[0.37]	[0.16]	[0.12]						
Beneficiary-specific consumption	0.15	0.51***	0.61***	0.62***	0.45***	0.69***	8.27	1770	0.23	0.01	0.01	0.11
	(0.12)	(0.12)	(0.13)	(0.12)	(0.15)	(0.12)						
	[0.12]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]						
Panel B. Household wealth												
HH net non-land wealth	−0.18	0.16	1.20***	1.33***	1.11***	0.89**	11.28	1770	0.21	0.56	0.53	0.67
	(0.40)	(0.59)	(0.44)	(0.42)	(0.41)	(0.48)						
	[0.36]	[0.42]	[0.01]	[0.00]	[0.01]	[0.05]						
HH livestock wealth	−0.01	1.76***	1.84***	2.64***	2.17***	2.21***	7.81	1770	0.25	0.00	0.12	0.92
	(0.37)	(0.49)	(0.52)	(0.45)	(0.47)	(0.45)						
	[0.42]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]						
Savings	1.03***	1.04***	1.29***	1.56***	1.43***	1.69***	9.24	1770	0.20	0.56	0.22	0.39
	(0.23)	(0.32)	(0.34)	(0.30)	(0.31)	(0.27)						
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]						
Debt	0.41	−0.10	−0.23	−0.56	−0.37	0.00	8.75	1770	0.20	0.19	0.86	0.45
	(0.28)	(0.41)	(0.43)	(0.45)	(0.42)	(0.38)						
	[0.10]	[0.42]	[0.34]	[0.13]	[0.24]	[0.42]						
Panel C. Beneficiary cognitive and non-cognitive skills												
Locus of control	0.06	0.13	0.02	0.00	0.08	0.23**	0.00	1770	0.28	0.59	0.30	0.12
	(0.06)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)						
	[0.65]	[0.38]	[1.00]	[1.00]	[0.65]	[0.03]						
Aspirations	−0.01	0.08	−0.05	0.13	0.03	0.14	0.00	1770	0.08	0.41	0.48	0.26
	(0.07)	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)						
	[1.00]	[0.69]	[1.00]	[0.38]	[1.00]	[0.33]						
Big Five index	0.12	0.08	0.11	0.02	−0.08	0.02	0.00	1770	0.10	0.55	0.25	0.35
	(0.07)	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)						
	[0.33]	[0.72]	[0.62]	[1.00]	[0.69]	[1.00]						
Business knowledge	0.65***	0.09	0.08	0.06	−0.03	0.63***	0.00	1770	0.23	0.00	0.29	0.00
	(0.07)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)						
	[0.00]	[0.69]	[0.69]	[0.83]	[1.00]	[0.00]						
Business attitudes	0.12	0.19	0.19	0.10	0.06	0.15	0.00	1770	0.09	0.63	0.08	0.40
	(0.07)	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)						
	[0.33]	[0.31]	[0.29]	[0.65]	[0.83]	[0.33]						

Notes: Regressions include but do not report the lagged dependent variable, fixed effects for randomization blocks, and a set of LASSO-selected baseline covariates, and are weighted to reflect intensive tracking. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. Reported p-values in final three columns derived from F-tests of hypotheses that cost-benefit ratios are equal between: (a) GD Lower and HD; (b) GD Lower and GD Large; and (c) GD Large and Combined treatments

Table 6: Cost-equivalent analysis: Primary outcomes

	Differential impact of HD	Cost-equivalent GD impact	Transfer Value	Control Mean	Obs.	R^2
Employed	-0.01	0.04	-0.00	0.48	1578	0.17
	(0.04)	(0.04)	(0.01)			
	[0.60]	[0.33]	[0.60]			
Productive hours	-2.25	5.05^{**}	-0.66	18.64	1578	0.19
	(2.19)	(2.14)	(0.56)			
	[0.29]	[0.05]	[0.27]			
Monthly income	-0.71^{*}	0.99^{**}	-0.02	8.05	1578	0.22
	(0.32)	(0.33)	(0.09)			
	[0.05]	[0.01]	[0.60]			
Productive assets	-2.27^{***}	3.83^{***}	0.03	5.61	1578	0.27
	(0.43)	(0.44)	(0.12)			
	[0.00]	[0.00]	[0.60]			
HH consumption per capita	-0.13	0.19^{**}	0.03	9.46	1548	0.31
	(0.08)	(0.08)	(0.02)			
	[0.12]	[0.04]	[0.16]			

Note: This table uses a linear adjustment of primary outcomes for program cost to compare HD and GD at exactly equivalent costs. The *Transfer value* column estimates the marginal effect of spending an extra $100 through cash transfers. The *Cost-equivalent GD impact* column is estimated as a dummy for either HD or GD treatment, and estimates the impact of cash at the exact cost of HD. The *Differential impact of HD* column then estimates the differential effect of HD above cash at this benchmarked cost. Regressions include but do not report the lagged dependent variable, fixed effects for randomization blocks, and a set of LASSO-selected baseline covariates, and are weighted to reflect intensive tracking. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. Employed is a dummy variable for spending more than 10 hours per week working for a wage or as primary operator of a microenterprise. Productive hours are measured over prior 7 days in all activities other than own-farm agriculture. Monthly income, productive assets, and household consumption are winsorized at 1% and 99% and analyzed in Inverse Hyperbolic Sine, meaning that treatment effects can be interpreted as percent changes.

Table 7: Cost-equivalent analysis: Secondary outcomes

	Differential impact of HD	Cost-equivalent GD impact	Transfer Value	Control Mean	Obs.	R^2
Panel A. Beneficiary welfare						
Subjective well-being	−0.23**	0.42***	0.03	0.00	1578	0.16
	(0.08)	(0.09)	(0.02)			
	[0.01]	[0.00]	[0.23]			
Mental health	−0.00	−0.04	0.03	0.00	1578	0.08
	(0.09)	(0.09)	(0.02)			
	[0.50]	[0.50]	[0.23]			
Beneficiary-specific consumption	−0.44***	0.61***	−0.02	8.27	1578	0.24
	(0.12)	(0.12)	(0.04)			
	[0.00]	[0.00]	[0.50]			
Panel B. Household wealth						
HH net non-land wealth	−0.80	0.62	0.15	11.28	1578	0.21
	(0.49)	(0.48)	(0.12)			
	[0.30]	[0.35]	[0.35]			
HH livestock wealth	−1.92***	1.90***	0.09	7.81	1578	0.27
	(0.45)	(0.46)	(0.12)			
	[0.00]	[0.00]	[0.52]			
Savings	−0.10	1.13***	0.08	9.24	1578	0.21
	(0.28)	(0.29)	(0.08)			
	[0.52]	[0.00]	[0.46]			
Debt	0.57	−0.16	−0.06	8.75	1578	0.21
	(0.39)	(0.39)	(0.11)			
	[0.35]	[0.52]	[0.52]			
Panel C. Beneficiary cognitive and non-cognitive skills						
Locus of control	−0.00	0.07	−0.00	0.00	1578	0.29
	(0.08)	(0.08)	(0.02)			
	[1.00]	[0.99]	[1.00]			
Aspirations	−0.06	0.06	−0.00	0.00	1578	0.08
	(0.09)	(0.08)	(0.02)			
	[0.99]	[0.99]	[1.00]			
Big Five index	−0.01	0.13	−0.04	0.00	1578	0.11
	(0.09)	(0.09)	(0.02)			
	[1.00]	[0.76]	[0.76]			
Business knowledge	0.54***	0.11	−0.03	0.00	1578	0.23
	(0.09)	(0.09)	(0.02)			
	[0.00]	[0.76]	[0.82]			
Business attitudes	−0.09	0.21*	−0.03	0.00	1578	0.09
	(0.08)	(0.08)	(0.02)			
	[0.82]	[0.09]	[0.76]			

Note: This table uses a linear adjustment of secondary outcomes for program cost to compare HD and GD at exactly equivalent costs. The *Transfer value* column estimates the marginal effect of spending an extra $100 through cash transfers. The *Cost-equivalent GD impact* column is estimated as a dummy for either HD or GD treatment, and estimates the impact of cash at the exact cost of HD. The *Differential impact of HD* column then estimates the differential effect of HD above cash at this benchmarked cost. Regressions include but do not report the lagged dependent variable, fixed effects for randomization blocks, and a set of LASSO-selected baseline covariates, and are weighted to reflect intensive tracking. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all outcomes within each family are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance.

Table 8: Complementarities: Primary outcomes

	HD	GiveDirectly Lower	Middle	Upper	Large	Combined	Complementarities tests (a)	(b)	(c)
Employed	0.02	0.03	0.05	0.00	0.01	0.01	−0.07	−0.00	0.00
	(0.03)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	[0.29]	[0.94]	[0.97]
Productive hours	2.79	2.76	6.54	3.56	1.12	2.31	−7.02	1.19	1.16
	(1.57)	(2.34)	(2.40)	(2.52)	(2.06)	(2.03)	[0.03]	[0.63]	[0.74]
Monthly income	0.31	0.76	1.08	1.14	0.73	1.04	−0.35	0.31	0.76
	(0.26)	(0.36)	(0.34)	(0.35)	(0.35)	(0.32)	[0.44]	[0.42]	[0.15]
Productive assets	1.54	3.94	3.80	3.84	4.02	4.42	−0.92	0.40	2.80
	(0.35)	(0.46)	(0.50)	(0.46)	(0.47)	(0.44)	[0.17]	[0.46]	[0.00]
HH consumption	0.05	0.20	0.27	0.23	0.36	0.27	−0.05	−0.09	0.06
per capita	(0.06)	(0.08)	(0.09)	(0.07)	(0.07)	(0.07)	[0.65]	[0.30]	[0.62]

Notes: Table replicates specification of ITT analysis of primary outcomes, as reported in Table 4, with each GD transfer value treated as a separate arm. The table derives two measures of complementarities and the associated p-values for the corresponding null of no complementary in brackets. Complementarity test (a) is the canonical 2×2 test: we estimate $\delta^{Combined} - \left(\delta^{HD} + \delta^{GD\,Middle}\right)$ and test the null that this difference is equal to zero. Complementarity test (b) reports the estimated difference between the Combined and GD Large arms, and a p-value associated with the null of equality. Complementarity test (c) tests the null that the difference between the coefficients on the Combined and GD-Large treatments is equal to the difference between the coefficients on HD and that on GD-Lower. We report the point estimate for this difference-in-differences, $\left(\delta^{Combined} - \delta^{GD\,Large}\right) - \left(\delta^{HD} - \delta^{GD\,Lower}\right)$ in the notation of Equation 1, and the p-value from the corresponding test of this null below it in brackets.

Table 9: Complementarities: Secondary outcomes

	HD	GiveDirectly Lower	GiveDirectly Middle	GiveDirectly Upper	GiveDirectly Large	Combined	Complementarities tests (a)	Complementarities tests (b)	Complementarities tests (c)
Panel A. Beneficiary welfare									
Subjective well-being	0.19	0.40	0.53	0.48	0.55	0.41	−0.31	−0.15	0.06
	(0.07)	(0.09)	(0.10)	(0.09)	(0.09)	(0.09)	[0.01]	[0.16]	[0.68]
Mental health	−0.04	−0.06	0.07	0.03	0.11	0.12	0.09	0.01	−0.01
	(0.07)	(0.09)	(0.09)	(0.09)	(0.10)	(0.09)	[0.50]	[0.93]	[0.93]
Beneficiary-specific consumption	0.15	0.51	0.61	0.62	0.45	0.69	−0.08	0.24	0.60
	(0.12)	(0.12)	(0.13)	(0.12)	(0.15)	(0.12)	[0.65]	[0.12]	[0.00]
Panel B. Household wealth									
HH net non-land wealth	−0.18	0.16	1.20	1.33	1.11	0.89	−0.13	−0.22	0.12
	(0.40)	(0.59)	(0.44)	(0.42)	(0.41)	(0.48)	[0.84]	[0.66]	[0.88]
HH livestock wealth	−0.01	1.76	1.84	2.64	2.17	2.21	0.38	0.04	1.80
	(0.37)	(0.49)	(0.52)	(0.45)	(0.47)	(0.45)	[0.58]	[0.94]	[0.01]
Savings	1.03	1.04	1.29	1.56	1.43	1.69	−0.62	0.26	0.27
	(0.23)	(0.32)	(0.34)	(0.30)	(0.31)	(0.27)	[0.14]	[0.41]	[0.54]
Debt	0.41	−0.10	−0.23	−0.56	−0.37	0.00	−0.18	0.37	−0.13
	(0.28)	(0.41)	(0.43)	(0.45)	(0.42)	(0.38)	[0.76]	[0.45]	[0.84]
Panel C. Beneficiary cognitive and non-cognitive skills									
Locus of control	0.06	0.13	0.02	0.00	0.08	0.23	0.14	0.15	0.21
	(0.06)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	[0.21]	[0.13]	[0.11]
Aspirations	−0.01	0.08	−0.05	0.13	0.03	0.14	0.19	0.11	0.19
	(0.07)	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	[0.13]	[0.26]	[0.16]
Big Five index	0.12	0.08	0.11	0.02	−0.08	0.02	−0.20	0.10	0.06
	(0.07)	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	[0.11]	[0.35]	[0.68]
Business knowledge	0.65	0.09	0.08	0.06	−0.03	0.63	−0.11	0.65	0.09
	(0.07)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	[0.43]	[0.00]	[0.57]
Business attitudes	0.12	0.19	0.19	0.10	0.06	0.15	−0.16	0.09	0.16
	(0.07)	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	[0.19]	[0.40]	[0.26]

Notes: Table replicates specification of ITT analysis of secondary outcomes, as reported in Table 5, with each GD transfer value treated as a separate arm. The table derives two measures of complementarities and the associated p-values for the corresponding null of no complementary in brackets. Complementarity test (a) is the canonical 2×2 test: we estimate $\delta^{Combined} - \left(\delta^{HD} + \delta^{GD\ Middle}\right)$ and test the null that this difference is equal to zero. Complementarity test (b) reports the estimated difference between the Combined and GD Large arms, and a p-value associated with the null of equality. Complementarity test (c) tests the null that the difference between the coefficients on the Combined and GD-Large treatments is equal to the difference between the coefficients on HD and that on GD-Lower. We report the point estimate for this difference-in-differences, $\left(\delta^{Combined} - \delta^{GD\ Large}\right) - \left(\delta^{HD} - \delta^{GD\ Lower}\right)$ in the notation of Equation 1, and the p-value from the corresponding test of this null below it in brackets.

Table 10: Spillovers on Program Compliance

	HD	Combined
HD Saturation	0.27**	0.09
	(0.09)	(0.17)
	[0.01]	[0.85]
GD Main Saturation	0.13	−0.01
	(0.08)	(0.14)
	[0.45]	[1.00]
GD Large Saturation	0.12	0.15
	(0.14)	(0.26)
	[0.85]	[0.85]
Average compliance	0.86	0.90
Observations	466	192
R^2	0.32	0.55
p-value	0.01	0.85

Notes: Table uses a Linear Probability Model to examine the likelihood that an individual assigned to that arm participates in Huguka Dukore (Column 1) or the HD component of the Combined arm (Column 2), as a function of the saturation of each of the three treatments among other members of the same village. Regressions include but do not report the lagged dependent variable, fixed effects for randomization blocks, and a set of LASSO-selected baseline covariates, and are weighted to reflect intensive tracking. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. Bottom row is the p-value on an F-test of the joint significance of the three saturation terms.

Table 11: Spillover effects: levels model

	Employed	Productive Hours	Monthly Income	Productive Assets	Consumption
HD	0.01	0.91	0.22	1.16***	0.05
	(0.03)	(1.25)	(0.20)	(0.30)	(0.05)
	[1.00]	[1.00]	[0.93]	[0.00]	[1.00]
GD main	0.01	2.73	0.90***	3.49***	0.23***
	(0.03)	(1.28)	(0.19)	(0.26)	(0.04)
	[1.00]	[0.13]	[0.00]	[0.00]	[0.00]
GD Huge treatment	0.00	0.18	0.70	3.81***	0.35***
	(0.04)	(1.92)	(0.34)	(0.46)	(0.07)
	[1.00]	[1.00]	[0.13]	[0.00]	[0.00]
HD Saturation	0.01	−1.80	−0.08	−0.01	0.02
	(0.06)	(2.75)	(0.45)	(0.66)	(0.10)
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
GD Main Saturation	−0.00	0.41	−0.39	0.75	−0.01
	(0.06)	(2.77)	(0.42)	(0.59)	(0.09)
	[1.00]	[1.00]	[1.00]	[0.77]	[1.00]
GD Large Saturation	−0.11	−9.89	0.15	0.12	−0.07
	(0.10)	(4.85)	(0.72)	(1.04)	(0.16)
	[0.93]	[0.13]	[1.00]	[1.00]	[1.00]
Control mean	0.48	18.64	8.05	5.61	9.46
Observations	1770	1770	1770	1770	1737
R^2	0.17	0.19	0.21	0.26	0.33
p-value	0.69	0.23	0.79	0.64	0.96

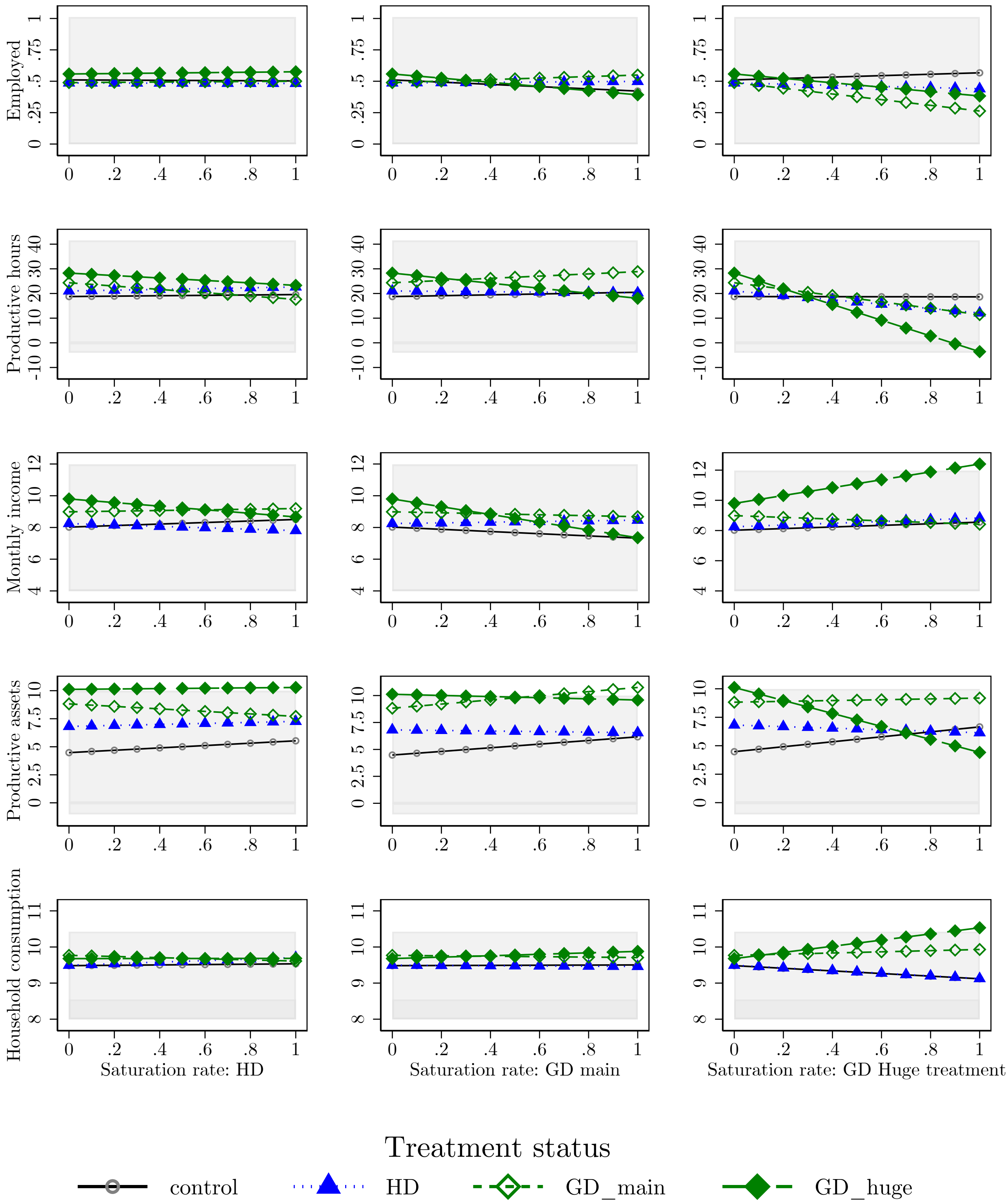
Notes: Table analyzes spillover effects of the three main treatments (HD, GD Main, and GD Large) on the five primary outcomes. The first three rows are dummy variables for own treatment status, and the next three are the saturation rates for the three treatments among others in the village, so measure the marginal effect of going from no one else treated to everyone else treated. Regressions include but do not report the lagged dependent variable, fixed effects for randomization blocks, and a set of LASSO-selected baseline covariates, and are weighted to reflect intensive tracking. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. Bottom row is the p-value on an F-test of the joint significance of the three saturation terms.

Table 12: Accounting for cash transfers

	Control mean	Treatment effect				
		Lower	Middle	Upper	Large	Combined
Panel A. Inflows						
Cash received	0.00	317.16	410.65	502.96	750.30	410.65
Beneficiary income	209.36	158.39	226.36	239.48	152.48	218.09
Transfers received	23.38	40.93	62.50	61.16	72.74	57.63
Total inflows	.	516.47	699.51	803.60	975.53	686.37
Panel B. Outflows						
Household consumption	625.85	124.50	166.53	146.10	223.08	169.13
Livestock	118.64	208.40	218.03	313.31	257.79	262.49
Productive assets	49.89	196.35	189.78	191.37	200.70	220.60
Savings	51.99	54.11	66.88	81.22	74.34	88.11
Debt	61.93	−5.93	−14.04	−34.77	−22.79	0.14
Loans made	3.83	3.46	1.74	5.41	2.61	4.74
Transfers made	4.53	3.08	8.48	2.56	1.46	3.65
Total outflows	.	595.82	665.47	774.73	782.78	748.60
Panel C. Totals						
Share accounted	.	115%	95%	96%	80%	109%

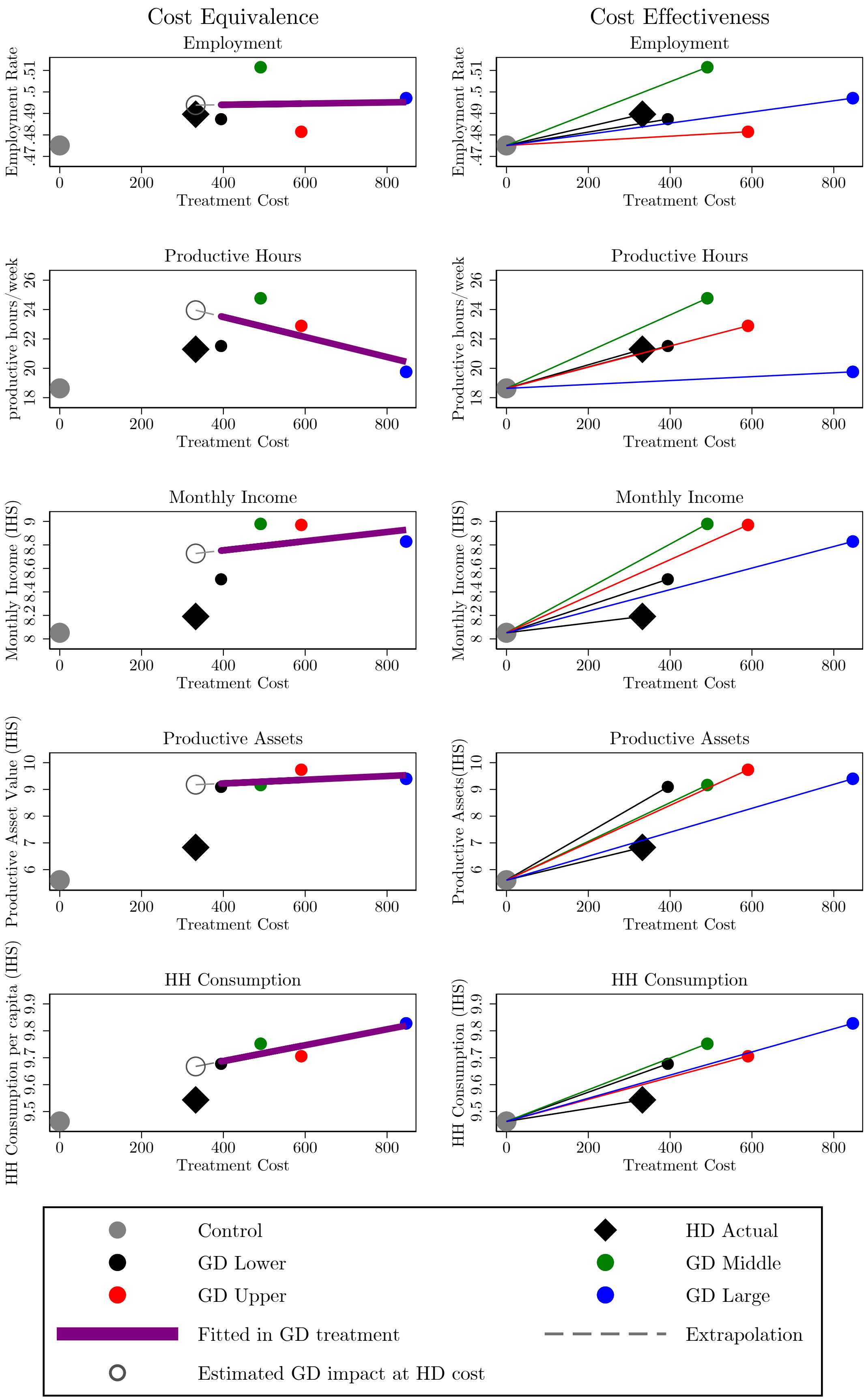
Note. Table presents control means and estimated impacts on financial values, in dollars. Beneficiary income and household consumption are estimated 12-month totals, assuming constant flows over the period between delivery of cash transfers and follow-up. Inter-household transfers and loans are 12-month recall variables. All other variables are stocks measured at follow-up. *Total inflows* are the sum of cash received, beneficiary income, and transfers received. *Total outflows* are the sum of household consumption, livestock values, other productive asset values, savings values, the negative of debt values, loans made, and transfers made. *Share accounted* is the ratio between total outflows and total inflows.

Figure 1: Expected outcomes by treatment arm, under alternative saturation rates



Notes: Each panel presents predicted outcomes under each of the four main treatment arms (Control, HD, GD Main, and GD Large), as the saturation level of a specific active treatment arm changes. Rows correspond to the outcomes of employment, productive hours, and the inverse hyperbolic sine of monthly incomes, productive assets, and household consumption per adult-equivalent, respectively. Horizontal shaded bands highlight one standard deviation above and below the control mean. Columns illustrate effects of variation in saturation rates in HD, GD-main, and GD-large, respectively. All predicted outcomes evaluated at means of covariates used in Equation 5.

Figure 2: Cost Equivalence versus Cost Effectiveness



Appendix A Supplementary tables and Figures

Table A.1: Process of identifying the eligible sample

Sector	Orientation sign-ups	Verified eligible	Baseline completed
Kaduha	273	261	235
Kibumbwe	144	139	127
Kigabiro	66	52	49
Kiyumba	102	70	66
Mugano	244	198	196
Muhazi	192	170	159
Munyaga	157	137	124
Munyiginya	115	102	94
Musange	170	115	110
Mushishiro	88	87	82
Nyakariro	227	200	190
Nyarusange	245	226	214
Shyogwe	252	210	202
Total	2275	1967	1848

Notes: Table gives the number of individuals participating in each of three phases of study recruitment, for each of the 13 sectors in which recruitment took place.

Table A.2: Survey modules by instrument and round

Module	Baseline instrument	Endline instrument
Identification	Both	Both
Social network	Beneficiary	—
Firm creation and employment history	Beneficiary	—
Wage employment	Beneficiary	Beneficiary
Microenterprise activities and assets	Both	Both
Time use	Beneficiary	Beneficiary
Income	Both	Beneficiary
Savings	Both	Both
Borrowing	Both	Both
Lending	Both	Both
Business contacts	Beneficiary	—
Private consumption	Beneficiary	Beneficiary
Private assets	Beneficiary	Beneficiary
Psychometrics	Beneficiary	Beneficiary
Raven's test	Beneficiary	—
Digit-span recall	Beneficiary	—
Numeracy	Beneficiary	—
Lottery choice	Beneficiary	—
Convex time budget	Beneficiary	—
Locus of control	Beneficiary	Beneficiary
Big Five	—	Beneficiary
Aspirations	—	Beneficiary
Mental health	—	Beneficiary
Business knowledge	—	Beneficiary
Business attitudes	—	Beneficiary
Program participation	—	Beneficiary
Gender empowerment	—	Beneficiary
Household roster	Household	Household
Dwelling characteristics	Household	Household
Land use and ownership	Household	Household
Inter-household transfers	Household	Household
Consumption	Household	Household
Dietary diversity	Household	Household
Household assets	Household	Household

Table A.3: Correlates of HD Participation

	Huguka Dukore stage completed…			
	Complier	Work Ready Now	Be Your Own Boss	Technical Training
Ubudehe category I	0.0106	0.0146	0.0128	−0.0199
	(0.0281)	(0.0319)	(0.0356)	(0.0327)
Beneficiary female	0.0142	0.0146	0.0284	−0.0187
	(0.0285)	(0.0317)	(0.0363)	(0.0336)
Beneficiary age	0.0120***	0.0100**	0.0085*	0.0067
	(0.0041)	(0.0048)	(0.0052)	(0.0042)
Beneficiary years of education	0.0037	0.0051	0.0050	0.0073
	(0.0061)	(0.0071)	(0.0083)	(0.0073)
Household members	0.0025	0.0061	0.0083	0.0146*
	(0.0074)	(0.0078)	(0.0085)	(0.0079)
Employed	0.0930*	0.0757	0.0846	0.0450
	(0.0504)	(0.0555)	(0.0615)	(0.0590)
Productive hours	−0.0038***	−0.0041***	−0.0047***	−0.0044***
	(0.0014)	(0.0015)	(0.0015)	(0.0014)
Monthly income	−0.0019	−0.0008	0.0004	−0.0026
	(0.0045)	(0.0051)	(0.0055)	(0.0050)
Productive assets	−0.0001	0.0004	−0.0016	−0.0027
	(0.0030)	(0.0037)	(0.0041)	(0.0040)
HH consumption per capita	0.0027	0.0042	0.0103	0.0157
	(0.0134)	(0.0152)	(0.0179)	(0.0154)
Beneficiary-specific consumption	−0.0092*	−0.0139**	−0.0109*	−0.0021
	(0.0053)	(0.0059)	(0.0064)	(0.0058)
HH net non-land wealth	0.0030	0.0040*	0.0050*	0.0026
	(0.0023)	(0.0024)	(0.0028)	(0.0024)
Savings	−0.0018	−0.0028	−0.0023	−0.0030
	(0.0033)	(0.0036)	(0.0041)	(0.0036)
Debt	0.0080**	0.0099***	0.0096**	0.0086**
	(0.0034)	(0.0038)	(0.0042)	(0.0038)
HH livestock wealth	0.0025	0.0025	−0.0003	−0.0034
	(0.0026)	(0.0030)	(0.0033)	(0.0030)
Business Knowledge	−0.0129	0.0056	0.0019	−0.0151
	(0.0135)	(0.0154)	(0.0179)	(0.0156)
Average completion rate	0.86	0.79	0.69	0.48
Observations	668	668	668	668
R^2	0.12	0.16	0.18	0.47
p-value	0.00	0.00	0.01	0.00

Notes: Table estimates Linear Probability Model for four measures of progression through the Huguka Dukore program: attending the first week of the WRN coursework and hence triggering payment to the implementer, then completing each of the three subsequent components of the program. Rows are the baseline covariates over which we look for heterogeneity in these compliance rates. p-value in the final row is from F-test on the joint significance of all of the covariates. Standard errors in parentheses; *=10%, **=5%, and ***=1% significance

Table A.4: Sampling of Attritors for Intensive Tracking Exercise

	Intensive tracking	Control mean	Observations	R^2
Ubudehe category I	−0.10 (0.09) [1.00]	0.38	120	0.01
Beneficiary female	−0.00 (0.09) [1.00]	0.62	122	0.00
Beneficiary age	1.66 (0.61) [0.13]	22.64	122	0.06
Beneficiary years of education	0.07 (0.45) [1.00]	8.10	122	0.00
Household members	−0.25 (0.37) [1.00]	4.66	122	0.00
Employed	−0.00 (0.09) [1.00]	0.31	122	0.00
Productive hours	2.48 (4.09) [1.00]	11.75	122	0.00
Monthly income	−0.41 (0.91) [1.00]	4.15	122	0.00
Productive assets	0.65 (0.85) [1.00]	2.59	122	0.00
HH consumption per capita	0.14 (0.20) [1.00]	9.44	122	0.00
Beneficiary-specific consumption	0.08 (0.36) [1.00]	7.61	122	0.00
HH net non-land wealth	0.95 (1.02) [1.00]	10.48	122	0.01
Savings	0.42 (0.83) [1.00]	7.67	122	0.00
Debt	−0.21 (0.86) [1.00]	7.60	122	0.00
HH livestock wealth	2.22 (1.09) [0.37]	5.26	122	0.04
Business Knowledge	0.34 (0.18) [0.37]	−0.02	122	0.03

Notes: Table regresses a sequence of covariates on a dummy variable for having been sampled for intensive tracking, within the sample of original attritors. Standard errors in parentheses, p-values corrected for False Discovery Rate across whole table in hard brackets; stars are based on FDR-adjusted values, *=10%, **=5%, and ***=1% significance.

Table A.5: Intent-to-treat analysis: Primary outcomes, aggregated specification

		GiveDirectly			Control			p-values		
	HD	Main	Large	Combined	Mean	Obs.	R^2	(a)	(b)	(c)
Employed	0.02	0.03	0.01	0.01	0.48	1770	0.16	0.89	0.52	0.46
	(0.03)	(0.03)	(0.05)	(0.04)						
	[0.29]	[0.24]	[0.51]	[0.51]						
Productive hours	2.80*	4.34***	1.12	2.33	18.64	1770	0.18	0.93	0.03	0.07
	(1.57)	(1.65)	(2.06)	(2.03)						
	[0.07]	[0.01]	[0.35]	[0.17]						
Monthly income	0.32	1.00***	0.73**	1.04***	8.05	1770	0.21	0.11	0.02	0.10
	(0.26)	(0.25)	(0.35)	(0.32)						
	[0.16]	[0.00]	[0.03]	[0.00]						
Productive assets	1.54***	3.86***	4.02***	4.42***	5.61	1770	0.26	0.00	0.00	0.00
	(0.35)	(0.34)	(0.47)	(0.44)						
	[0.00]	[0.00]	[0.00]	[0.00]						
HH consumption per capita	0.05	0.23***	0.36***	0.27***	9.46	1737	0.33	0.05	0.64	0.17
	(0.06)	(0.06)	(0.07)	(0.07)						
	[0.23]	[0.00]	[0.00]	[0.00]						

Note: Intention to treat pooling the three smaller cash transfer amounts into a single arm, GD Main. Regressions include but do not report the lagged dependent variable, fixed effects for randomization blocks, and a set of LASSO-selected baseline covariates, and are weighted to reflect intensive tracking. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. p-values in final three columns derived from F-tests of hypotheses that benefit-cost ratios are equal between (a) GD Main and HD; (b) GD Main and GD Large; and (c) GD Main and Combined. Employed is a dummy variable for spending more than 10 hours per week working for a wage or as primary operator of a microenterprise. Productive hours are measured over prior 7 days in all activities other than own-farm agriculture. Monthly income, productive assets, and household consumption are winsorized at 1% and 99% and analyzed in Inverse Hyperbolic Sine, meaning that treatment effects can be interpreted as percent changes.

Table A.6: Intent-to-treat analysis: Secondary outcomes, aggregated specification

		GiveDirectly			Control			p-values		
	HD	Main	Large	Combined	Mean	Obs.	R^2	(a)	(b)	(c)
Panel A. Beneficiary welfare										
Subjective well-being	0.19***	0.47***	0.55***	0.41***	0.00	1770	0.13	0.03	0.03	0.00
	(0.07)	(0.07)	(0.09)	(0.09)						
	[0.00]	[0.00]	[0.00]	[0.00]						
Mental health	−0.04	0.01	0.11	0.12	0.00	1770	0.07	0.43	0.48	0.40
	(0.07)	(0.07)	(0.10)	(0.09)						
	[0.34]	[0.34]	[0.16]	[0.12]						
Beneficiary-specific consumption	0.15	0.58***	0.45***	0.69***	8.27	1770	0.23	0.01	0.00	0.02
	(0.11)	(0.10)	(0.15)	(0.12)						
	[0.12]	[0.00]	[0.00]	[0.00]						
Panel B. Household wealth										
HH net non-land wealth	−0.17	0.91***	1.11***	0.90**	11.28	1770	0.20	0.02	0.41	0.28
	(0.40)	(0.35)	(0.41)	(0.48)						
	[0.31]	[0.01]	[0.01]	[0.05]						
HH livestock wealth	−0.00	2.08***	2.17***	2.22***	7.81	1770	0.25	0.00	0.02	0.02
	(0.37)	(0.35)	(0.47)	(0.45)						
	[0.46]	[0.00]	[0.00]	[0.00]						
Savings	1.04***	1.30***	1.43***	1.70***	9.24	1770	0.20	0.40	0.04	0.15
	(0.23)	(0.23)	(0.31)	(0.27)						
	[0.00]	[0.00]	[0.00]	[0.00]						
Debt	0.40	−0.29	−0.37	0.00	8.75	1770	0.20	0.02	0.80	0.34
	(0.28)	(0.30)	(0.42)	(0.38)						
	[0.10]	[0.19]	[0.22]	[0.46]						
Panel C. Beneficiary cognitive and non-cognitive skills										
Locus of control	0.06	0.05	0.08	0.23**	0.00	1770	0.27	0.58	0.98	0.14
	(0.06)	(0.06)	(0.08)	(0.08)						
	[0.45]	[0.53]	[0.45]	[0.02]						
Aspirations	−0.01	0.05	0.03	0.14	0.00	1770	0.08	0.52	0.60	0.63
	(0.07)	(0.07)	(0.09)	(0.08)						
	[1.00]	[0.54]	[1.00]	[0.20]						
Big Five index	0.12	0.07	−0.08	0.02	0.00	1770	0.10	0.23	0.10	0.40
	(0.07)	(0.07)	(0.09)	(0.09)						
	[0.20]	[0.45]	[0.53]	[1.00]						
Business knowledge	0.65***	0.08	−0.03	0.63***	0.00	1770	0.23	0.00	0.19	0.00
	(0.07)	(0.07)	(0.09)	(0.09)						
	[0.00]	[0.45]	[1.00]	[0.00]						
Business attitudes	0.12	0.16*	0.06	0.15	0.00	1770	0.09	0.81	0.06	0.26
	(0.07)	(0.06)	(0.09)	(0.09)						
	[0.20]	[0.06]	[0.65]	[0.20]						

Notes: Regressions include but do not report the lagged dependent variable, fixed effects for randomization blocks, and a set of LASSO-selected baseline covariates, and are weighted to reflect intensive tracking. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance.

Table A.7: Breakdown of employment impacts

		GiveDirectly					Control			
	HD	Lower	Middle	Upper	Large	Combined	Mean	Obs.	R^2	p-value
Panel A. Employment composition										
Non-agricultural microenterprise	0.05	0.09*	0.08*	0.14**	0.17***	0.12**	0.22	1770	0.12	0.00
	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)				
	[0.11]	[0.06]	[0.08]	[0.02]	[0.00]	[0.02]				
Other microenterprise or self-employment	0.04*	0.04	0.11**	0.07*	0.05*	0.04	0.07	1770	0.09	0.02
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)				
	[0.06]	[0.11]	[0.02]	[0.06]	[0.09]	[0.11]				
Agricultural processing or trading	0.01	0.13**	0.01	0.07*	0.06	0.02	0.17	1770	0.08	0.05
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)				
	[0.24]	[0.02]	[0.24]	[0.08]	[0.11]	[0.24]				
Agricultural wage labor	−0.02	−0.08**	−0.07*	−0.08**	−0.11**	−0.09**	0.22	1770	0.15	0.00
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)				
	[0.24]	[0.03]	[0.06]	[0.03]	[0.02]	[0.03]				
Non-agricultural wage labor	0.06*	−0.04	0.01	−0.06	−0.07*	0.01	0.30	1770	0.20	0.00
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)				
	[0.06]	[0.16]	[0.24]	[0.11]	[0.09]	[0.24]				
Panel B. Alternative hours thresholds										
Employed (0 hr)	0.05	0.05	0.07	0.02	0.06	0.06	0.70	1770	0.12	0.46
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Employed (10 hr)	0.02	0.03	0.05	0.00	0.01	0.01	0.48	1770	0.16	0.95
	(0.03)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Employed (20 hr)	0.04	0.04	0.08	0.02	0.01	0.04	0.29	1770	0.17	0.62
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Employed (30 hr)	0.02	0.03	0.09	0.04	0.00	0.06	0.19	1770	0.17	0.37
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)				
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
Employed (40 hr)	0.03	0.03	0.09	−0.01	0.02	0.04	0.13	1770	0.17	0.26
	(0.02)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)				
	[1.00]	[1.00]	[0.70]	[1.00]	[1.00]	[1.00]				

Notes: Panel A presents impacts on indicators for employment of any hours in the corresponding activity type in the preceding week. Panel B presents impacts on an indicator for overall employment, using the reported threshold for minimum hours. Regressions include but do not report an indicator for lagged employment status, fixed effects for randomization blocks, and a set of LASSO-selected baseline covariates, and are weighted to reflect intensive tracking. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across outcomes in each panel are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance.

Table A.8: Robustness of Linearity in Primary Cost Equivalence Adjustment

	Base Linear	Quad-ratic	Cubic	Drop lower	Drop mid	Drop upper	Drop huge
Employed	−0.012	−0.022	0.072	−0.020	−0.002	−0.024	−0.022
	(0.041)	(0.067)	(0.143)	(0.058)	(0.046)	(0.043)	(0.055)
Productive hours	−2.249	0.706	8.098	−4.722	−0.980	−2.460	−0.398
	(2.189)	(3.616)	(7.736)	(3.054)	(2.473)	(2.225)	(2.928)
Monthly income	-0.709^{**}	−0.204	−0.070	-1.005^{**}	-0.677^{*}	-0.654^{*}	−0.358
	(0.322)	(0.534)	(1.102)	(0.419)	(0.370)	(0.335)	(0.439)
Productive assets	-2.272^{***}	-2.499^{***}	-2.567^{*}	-2.130^{***}	-2.271^{***}	-2.292^{***}	-2.492^{***}
	(0.434)	(0.678)	(1.524)	(0.603)	(0.480)	(0.449)	(0.563)
HH consumption per capita	-0.133^{*}	−0.151	0.003	−0.158	−0.110	-0.147^{*}	−0.157
	(0.077)	(0.116)	(0.267)	(0.106)	(0.084)	(0.081)	(0.097)

Notes: Table reports the coefficient on the differential effect of HD over cost-equivalent cash using seven different specifications. Column 1 is the linear adjustment reported elsewhere. Column 2 includes a quadratic, and column 3 a quadratic and cubic term in the cost deviations from Gikuriro. Columns 4-7 leave out one of the cash treatment arms and repeat the linear cost adjustment. Asterices denote significance at the 10, 5, and 1 percent levels, and are based on household-clustered standard errors, in parentheses.

Table A.9: Robustness of Linearity in Secondary Cost Equivalence Adjustment

	Base Linear	Quad-ratic	Cubic	Drop lower	Drop mid	Drop upper	Drop huge
Panel A. Beneficiary welfare							
Subjective well-being	-0.229^{***}	−0.193	0.011	-0.288^{**}	-0.199^{**}	-0.225^{**}	-0.225^{**}
	(0.085)	(0.135)	(0.300)	(0.119)	(0.094)	(0.087)	(0.111)
Mental health	−0.001	0.059	0.268	−0.074	0.032	−0.009	0.048
	(0.090)	(0.142)	(0.298)	(0.122)	(0.099)	(0.092)	(0.117)
Beneficiary-specific consumption	-0.444^{***}	−0.265	−0.213	-0.573^{***}	-0.419^{***}	-0.429^{***}	-0.317^{**}
	(0.120)	(0.169)	(0.359)	(0.166)	(0.128)	(0.124)	(0.146)
Panel B. Household wealth							
HH net non-land wealth	−0.803	0.284	0.879	-1.515^{***}	−0.683	−0.634	−0.169
	(0.491)	(0.877)	(1.641)	(0.549)	(0.583)	(0.524)	(0.713)
HH livestock wealth	-1.918^{***}	-1.286^{*}	−2.520	-2.215^{***}	-2.109^{***}	-1.832^{***}	-1.412^{**}
	(0.453)	(0.706)	(1.559)	(0.620)	(0.498)	(0.475)	(0.593)
Savings	−0.097	0.278	0.063	−0.308	−0.086	−0.044	0.183
	(0.280)	(0.446)	(1.005)	(0.390)	(0.314)	(0.292)	(0.364)
Debt	0.567	0.233	0.675	0.743	0.660	0.480	0.337
	(0.394)	(0.642)	(1.374)	(0.556)	(0.442)	(0.401)	(0.521)
Panel C. Beneficiary cognitive and non-cognitive skills							
Locus of control	−0.003	−0.129	−0.168	0.097	−0.027	−0.027	−0.084
	(0.078)	(0.126)	(0.269)	(0.104)	(0.088)	(0.081)	(0.103)
Aspirations	−0.061	−0.024	−0.420	−0.035	−0.121	−0.026	−0.025
	(0.089)	(0.139)	(0.296)	(0.117)	(0.100)	(0.093)	(0.117)
Big Five index	−0.011	0.012	0.139	−0.040	0.014	−0.008	−0.003
	(0.087)	(0.145)	(0.300)	(0.114)	(0.099)	(0.091)	(0.118)
Business knowledge	0.536^{***}	0.548^{***}	0.539^{*}	0.558^{***}	0.538^{***}	0.527^{***}	0.543^{***}
	(0.090)	(0.142)	(0.292)	(0.119)	(0.101)	(0.093)	(0.118)
Business attitudes	−0.092	−0.125	−0.004	−0.070	−0.072	−0.103	−0.129
	(0.085)	(0.143)	(0.290)	(0.113)	(0.097)	(0.088)	(0.116)

Notes: Table reports the coefficient on the differential effect of HD over cost-equivalent cash using seven different specifications. Column 1 is the linear adjustment reported elsewhere. Column 2 includes a quadratic, and column 3 a quadratic and cubic term in the cost deviations from Gikuriro. Columns 4-7 leave out one of the cash treatment arms and repeat the linear cost adjustment. Asterices denote significance at the 10, 5, and 1 percent levels, and are based on household-clustered standard errors, in parentheses.

Table A.10: Heterogeneity: Gender

	Employed	Productive Hours	Monthly Income	Productive Assets	Consumption
HD	0.09	1.43	0.14	1.03	0.01
	(0.05)	(2.78)	(0.37)	(0.61)	(0.10)
	[0.30]	[0.96]	[0.96]	[0.30]	[1.00]
GD main	0.05	4.44	0.79^{*}	2.97^{***}	0.29^{**}
	(0.05)	(2.98)	(0.33)	(0.59)	(0.10)
	[0.71]	[0.41]	[0.06]	[0.00]	[0.02]
GD large	0.04	−0.08	0.62	3.18^{***}	0.36^{**}
	(0.07)	(3.62)	(0.51)	(0.77)	(0.12)
	[0.88]	[1.00]	[0.52]	[0.00]	[0.01]
Combined	0.05	3.55	0.59	3.50^{***}	0.37^{**}
	(0.06)	(3.67)	(0.43)	(0.71)	(0.13)
	[0.75]	[0.71]	[0.44]	[0.00]	[0.02]
HD × Female	−0.12	2.16	0.02	0.32	0.11
	(0.07)	(3.35)	(0.50)	(0.75)	(0.13)
	[0.30]	[0.88]	[1.00]	[0.96]	[0.71]
GD main × Female	−0.05	−0.32	−0.07	1.17	−0.06
	(0.07)	(3.53)	(0.48)	(0.73)	(0.12)
	[0.75]	[1.00]	[1.00]	[0.33]	[0.96]
GD large × Female	−0.04	1.95	0.28	1.02	−0.01
	(0.09)	(4.29)	(0.68)	(0.99)	(0.15)
	[0.96]	[0.96]	[0.96]	[0.70]	[1.00]
Combined × Female	−0.08	−2.65	0.61	1.25	−0.21
	(0.09)	(4.30)	(0.60)	(0.89)	(0.16)
	[0.71]	[0.88]	[0.70]	[0.44]	[0.49]
Female	-0.16^{***}	-13.87^{***}	-1.80^{***}	-1.32^{*}	0.03
	(0.05)	(2.26)	(0.35)	(0.53)	(0.09)
	[0.01]	[0.00]	[0.00]	[0.05]	[0.96]
Control mean	0.48	18.64	8.05	5.61	9.46
Observations	1770	1770	1770	1770	1737
R^2	0.06	0.10	0.07	0.11	0.10
p-value	0.53	0.84	0.81	0.43	0.38

Notes: Table presents tests for heterogeneity of treatment effects by Gender. Uninteracted coefficients in the first four rows give the treatment effect of the program on men, and the next four rows test for the differential effect between women and men. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. p-value in the last row from an F-test on whether treatments have a jointly differential effect by gender.

Table A.11: Heterogeneity: Risk aversion

	Employed	Productive Hours	Monthly Income	Productive Assets	Consumption
HD	0.01	2.64	0.12	1.21***	0.08
	(0.03)	(1.62)	(0.27)	(0.37)	(0.06)
	[0.84]	[0.23]	[0.84]	[0.01]	[0.40]
GD main	0.01	4.25**	0.74**	3.68***	0.25***
	(0.03)	(1.70)	(0.26)	(0.36)	(0.06)
	[0.84]	[0.05]	[0.02]	[0.00]	[0.00]
GD large	0.02	0.78	0.74	3.83***	0.35***
	(0.05)	(2.10)	(0.37)	(0.49)	(0.08)
	[0.84]	[0.84]	[0.13]	[0.00]	[0.00]
Combined	0.01	2.66	1.01***	4.26***	0.26***
	(0.04)	(2.11)	(0.31)	(0.43)	(0.08)
	[0.84]	[0.38]	[0.01]	[0.00]	[0.01]
HD × Baseline risk aversion	0.03	1.68	0.50	−0.11	0.04
	(0.03)	(1.61)	(0.26)	(0.37)	(0.06)
	[0.67]	[0.50]	[0.15]	[0.84]	[0.84]
GD main × Baseline risk aversion	0.01	−0.51	0.34	−0.06	0.00
	(0.03)	(1.71)	(0.25)	(0.35)	(0.06)
	[0.84]	[0.84]	[0.38]	[0.84]	[0.94]
GD large × Baseline risk aversion	0.05	1.96	0.55	−0.68	−0.01
	(0.05)	(2.07)	(0.37)	(0.49)	(0.08)
	[0.40]	[0.56]	[0.30]	[0.36]	[0.84]
Combined × Baseline risk aversion	0.07	3.75	0.75*	0.12	−0.03
	(0.04)	(2.07)	(0.31)	(0.43)	(0.08)
	[0.23]	[0.18]	[0.05]	[0.84]	[0.84]
Baseline risk aversion	−0.01	−0.38	−0.43*	0.10	−0.05
	(0.02)	(1.10)	(0.19)	(0.26)	(0.04)
	[0.84]	[0.84]	[0.06]	[0.84]	[0.50]
Control mean	0.48	18.64	8.05	5.61	9.46
Observations	1770	1770	1770	1770	1737
R^2	0.02	0.03	0.03	0.11	0.10
p-value	0.42	0.27	0.13	0.65	0.95

Notes: Table presents tests for heterogeneity of treatment effects by Risk Aversion. Risk Aversion demeaned before interaction so first four rows give effect of treatment at average value, and next four rows test for differential treatment effect by risk aversion. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. p-value in the last row from an F-test on whether treatments have a jointly differential effect by gender.

Table A.12: Heterogeneity: Baseline household consumption

	Employed	Productive Hours	Monthly Income	Productive Assets	Consumption
HD	0.01	2.60	0.14	1.27***	0.13*
	(0.03)	(1.62)	(0.27)	(0.37)	(0.06)
	[0.64]	[0.18]	[0.63]	[0.00]	[0.09]
GD main	0.02	4.31**	0.76**	3.72***	0.28***
	(0.03)	(1.70)	(0.26)	(0.36)	(0.06)
	[0.63]	[0.04]	[0.01]	[0.00]	[0.00]
GD large	0.03	1.17	0.84*	3.93***	0.40***
	(0.05)	(2.08)	(0.36)	(0.48)	(0.08)
	[0.63]	[0.63]	[0.06]	[0.00]	[0.00]
Combined	0.02	2.90	1.08***	4.32***	0.26***
	(0.04)	(2.14)	(0.30)	(0.42)	(0.08)
	[0.64]	[0.27]	[0.00]	[0.00]	[0.00]
HD × Baseline HH consumption per AE	−0.04	−1.09	0.03	−0.19	−0.02
	(0.03)	(1.72)	(0.25)	(0.38)	(0.07)
	[0.31]	[0.63]	[0.80]	[0.63]	[0.64]
GD main × Baseline HH consumption per AE	0.01	0.83	0.29	−0.07	−0.11
	(0.03)	(1.63)	(0.27)	(0.34)	(0.06)
	[0.64]	[0.63]	[0.37]	[0.76]	[0.13]
GD large × Baseline HH consumption per AE	0.09*	2.51	0.70	0.46	−0.13
	(0.04)	(1.95)	(0.39)	(0.48)	(0.08)
	[0.09]	[0.29]	[0.14]	[0.45]	[0.18]
Combined × Baseline HH consumption per AE	−0.02	−1.44	−0.69*	−0.68	0.00
	(0.04)	(2.15)	(0.30)	(0.41)	(0.11)
	[0.63]	[0.63]	[0.06]	[0.17]	[0.82]
Baseline HH consumption per AE	0.02	0.43	0.10	0.51	0.32***
	(0.02)	(1.05)	(0.18)	(0.27)	(0.04)
	[0.63]	[0.64]	[0.63]	[0.12]	[0.00]
Control mean	0.48	18.64	8.05	5.61	9.46
Observations	1770	1770	1770	1770	1737
R^2	0.02	0.03	0.04	0.12	0.18
p-value	0.05	0.43	0.01	0.21	0.24

Notes: Table presents tests for heterogeneity of treatment effects by baseline Household Consumption. Consumption demeaned before interaction so first four rows give effect of treatment at average value, and next four rows test for differential treatment effect by consumption. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. p-value in the last row from an F-test on whether treatments have a jointly differential effect by gender.

Table A.13: Heterogeneity: Baseline local employment rates

	Employed	Productive Hours	Monthly Income	Productive Assets	Consumption
HD	0.01	2.53	0.10	1.21***	0.08
	(0.03)	(1.61)	(0.27)	(0.37)	(0.06)
	[1.00]	[0.37]	[1.00]	[0.01]	[0.60]
GD main	0.01	4.29**	0.73**	3.68***	0.25***
	(0.03)	(1.71)	(0.26)	(0.36)	(0.06)
	[1.00]	[0.05]	[0.02]	[0.00]	[0.00]
GD large	0.02	0.83	0.77	3.84***	0.34***
	(0.05)	(2.11)	(0.37)	(0.49)	(0.08)
	[1.00]	[1.00]	[0.13]	[0.00]	[0.00]
Combined	0.02	2.80	1.04***	4.30***	0.26***
	(0.04)	(2.13)	(0.31)	(0.42)	(0.08)
	[1.00]	[0.60]	[0.01]	[0.00]	[0.01]
HD × Baseline cell share employed	0.35	19.81	2.94	4.69	0.15
	(0.32)	(16.21)	(2.56)	(3.64)	(0.62)
	[0.60]	[0.60]	[0.60]	[0.60]	[1.00]
GD main × Baseline cell share employed	0.17	−1.97	3.00	1.72	0.26
	(0.32)	(17.05)	(2.47)	(3.50)	(0.59)
	[1.00]	[1.00]	[0.60]	[1.00]	[1.00]
GD large × Baseline cell share employed	0.34	0.26	2.38	5.25	−0.88
	(0.46)	(22.84)	(3.23)	(4.42)	(0.79)
	[1.00]	[1.00]	[1.00]	[0.60]	[0.60]
Combined × Baseline cell share employed	0.27	−1.71	2.68	7.59	1.51
	(0.43)	(21.35)	(3.11)	(4.23)	(0.76)
	[1.00]	[1.00]	[1.00]	[0.23]	[0.16]
Baseline cell share employed	0.04	0.29	1.02	−1.41	−0.24
	(0.25)	(12.35)	(1.98)	(2.87)	(0.50)
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
Control mean	0.48	18.64	8.05	5.61	9.46
Observations	1770	1770	1770	1770	1737
R^2	0.02	0.03	0.03	0.11	0.10
p-value	0.85	0.65	0.77	0.35	0.10

Notes: Table presents tests for heterogeneity of treatment effects by baseline Employment Rates. Employment demeaned before interaction so first four rows give effect of treatment at average value, and next four rows test for differential treatment effect by employment rates. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. p-value in the last row from an F-test on whether treatments have a jointly differential effect by gender.

Table A.14: Heterogeneity: Age 23 and over

	Employed	Productive Hours	Monthly Income	Productive Assets	Consumption
HD	0.06	4.32	0.26	1.17	0.09
	(0.05)	(2.52)	(0.43)	(0.56)	(0.10)
	[0.64]	[0.32]	[1.00]	[0.15]	[0.82]
GD main	0.02	2.39	0.72	4.26^{***}	0.28^{**}
	(0.05)	(2.46)	(0.42)	(0.53)	(0.09)
	[1.00]	[0.82]	[0.32]	[0.00]	[0.02]
GD large	0.00	0.03	0.88	4.27^{***}	0.31^{*}
	(0.07)	(3.23)	(0.62)	(0.76)	(0.12)
	[1.00]	[1.00]	[0.47]	[0.00]	[0.05]
Combined	0.08	2.02	1.53^{**}	3.97^{***}	0.21
	(0.06)	(2.98)	(0.47)	(0.63)	(0.11)
	[0.64]	[1.00]	[0.01]	[0.00]	[0.30]
HD × Older than 22	−0.08	−3.01	−0.23	0.08	−0.01
	(0.07)	(3.25)	(0.55)	(0.74)	(0.12)
	[0.64]	[0.82]	[1.00]	[1.00]	[1.00]
GD main × Older than 22	0.00	3.64	0.12	−1.06	−0.05
	(0.07)	(3.41)	(0.53)	(0.72)	(0.12)
	[1.00]	[0.71]	[1.00]	[0.47]	[1.00]
GD large × Older than 22	0.01	1.04	−0.29	−0.76	0.04
	(0.10)	(4.23)	(0.74)	(0.99)	(0.15)
	[1.00]	[1.00]	[1.00]	[0.97]	[1.00]
Combined × Older than 22	−0.11	1.55	−0.88	0.57	0.10
	(0.09)	(4.19)	(0.63)	(0.86)	(0.16)
	[0.64]	[1.00]	[0.47]	[1.00]	[1.00]
Older than 22	0.12^{*}	2.47	1.04^{*}	0.28	0.22^{*}
	(0.05)	(2.18)	(0.39)	(0.53)	(0.09)
	[0.06]	[0.64]	[0.05]	[1.00]	[0.06]
Control mean	0.48	18.64	8.05	5.61	9.46
Observations	1770	1770	1770	1770	1737
R^2	0.02	0.04	0.04	0.11	0.11
p-value	0.51	0.46	0.57	0.26	0.90

Notes: Table presents tests for heterogeneity of treatment effects by age. First four rows give effect of treatment among young, and next four rows test for differential treatment effect for those 23 and over. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. p-value in the last row from an F-test on whether treatments have a jointly differential effect by gender.

Table A.15: Spillover effects: full model, employment outcome

	Treatment		
	HD	GD Main	GD Huge
Direct effects of treatment at saturation level of zero			
Direct effect	−0.02	−0.02	0.05
	(0.07)	(0.07)	(0.13)
	[1.00]	[1.00]	[1.00]
Spillover effects of treatment onto control individuals			
Spillover to control	−0.01	−0.09	0.06
	(0.09)	(0.09)	(0.16)
	[1.00]	[1.00]	[1.00]
Additional effect of treatment onto individuals assigned to...			
HD	0.00	0.10	−0.10
	(0.11)	(0.11)	(0.19)
	[1.00]	[1.00]	[1.00]
GD main	0.02	0.15	−0.28
	(0.11)	(0.12)	(0.20)
	[1.00]	[1.00]	[1.00]
GD large	0.03	−0.08	−0.23
	(0.22)	(0.17)	(0.42)
	[1.00]	[1.00]	[1.00]
Saturation mean	0.36	0.36	0.09
Saturation SD	0.23	0.23	0.13
p-value	1.00	0.55	0.49

Notes: Each column describes the direct and spillover effects of a specific treatment on Employment; all results in the table are from a single estimation. Saturation mean and standard deviation correspond to the distribution of saturation rates for the treatment in question. Regressions include but do not report the lagged dependent variable, fixed effects for randomization blocks, and a set of LASSO-selected baseline covariates, and are weighted to reflect intensive tracking. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. p-value in the last row corresponds to a test for whether the treatment in question has interference effects on any arm, including control.

Table A.16: Spillover effects: full model, productive hours outcome

	Treatment		
	HD	GD Main	GD Huge
Direct effects of treatment at saturation level of zero			
Direct effect	2.26	5.60	9.52
	(3.71)	(3.67)	(6.27)
	[1.00]	[0.76]	[0.76]
Spillover effects of treatment onto control individuals			
Spillover to control	0.81	1.74	−0.12
	(4.39)	(4.68)	(6.56)
	[1.00]	[1.00]	[1.00]
Additional effect of treatment onto individuals assigned to...			
HD	0.81	−2.50	−8.79
	(5.72)	(5.82)	(9.63)
	[1.00]	[1.00]	[0.87]
GD main	−7.55	2.76	−12.81
	(5.60)	(6.39)	(8.73)
	[0.76]	[1.00]	[0.76]
GD large	−5.79	−11.98	−31.75
	(10.27)	(7.70)	(17.74)
	[1.00]	[0.76]	[0.76]
Saturation mean	0.36	0.36	0.09
Saturation SD	0.23	0.23	0.13
p-value	0.65	0.49	0.09

Notes: Each column describes the direct and spillover effects of a specific treatment on Productive Hours; all results in the table are from a single estimation. Saturation mean and standard deviation correspond to the distribution of saturation rates for the treatment in question. Regressions include but do not report the lagged dependent variable, fixed effects for randomization blocks, and a set of LASSO-selected baseline covariates, and are weighted to reflect intensive tracking. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. p-value in the last corresponds to a test for whether the treatment in question has interference effects on any arm, including control.

Table A.17: Spillover effects: full model, monthly income outcome

	Treatment		
	HD	GD Main	GD Huge
Direct effects of treatment at saturation level of zero			
Direct effect	0.23	0.96	1.77
	(0.55)	(0.58)	(0.88)
	[1.00]	[1.00]	[1.00]
Spillover effects of treatment onto control individuals			
Spillover to control	0.48	−0.69	0.53
	(0.82)	(0.79)	(1.13)
	[1.00]	[1.00]	[1.00]
Additional effect of treatment onto individuals assigned to...			
HD	−0.92	0.91	0.05
	(0.87)	(0.82)	(1.57)
	[1.00]	[1.00]	[1.00]
GD main	−0.28	0.39	−1.10
	(0.91)	(0.92)	(1.40)
	[1.00]	[1.00]	[1.00]
GD large	−1.61	−1.75	2.08
	(1.41)	(1.40)	(3.37)
	[1.00]	[1.00]	[1.00]
Saturation mean	0.36	0.36	0.09
Saturation SD	0.23	0.23	0.13
p-value	0.66	0.28	0.88

Notes: Each column describes the direct and spillover effects of a specific treatment on Monthly Income (IHS); all results in the table are from a single estimation. Saturation mean and standard deviation correspond to the distribution of saturation rates for the treatment in question. Regressions include but do not report the lagged dependent variable, fixed effects for randomization blocks, and a set of LASSO-selected baseline covariates, and are weighted to reflect intensive tracking. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. *p*-value in the last row corresponds to a test for whether the treatment in question has interference effects on any arm, including control.

Table A.18: Spillover effects: productive assets outcome

	Treatment		
	HD	GD Main	GD Huge
Direct effects of treatment at saturation level of zero			
Direct effect	2.34***	4.35***	5.64***
	(0.74)	(0.73)	(1.09)
	[0.01]	[0.00]	[0.00]
Spillover effects of treatment onto control individuals			
Spillover to control	1.05	1.69	2.18
	(1.00)	(1.10)	(1.72)
	[0.36]	[0.26]	[0.34]
Additional effect of treatment onto individuals assigned to...			
HD	−0.61	−1.95	−2.85
	(1.32)	(1.17)	(2.10)
	[0.49]	[0.26]	[0.32]
GD main	−2.17	0.23	−1.82
	(1.34)	(1.11)	(1.88)
	[0.26]	[0.52]	[0.36]
GD large	−0.88	−2.22	−7.88
	(1.90)	(1.87)	(4.10)
	[0.49]	[0.35]	[0.20]
Saturation mean	0.36	0.36	0.09
Saturation SD	0.23	0.23	0.13
p-value	0.62	0.21	0.27

Notes: Each column describes the direct and spillover effects of a specific treatment on Productive Assets (IHS); all results in the table are from a single estimation. Saturation mean and standard deviation correspond to the distribution of saturation rates for the treatment in question. Regressions include but do not report the lagged dependent variable, fixed effects for randomization blocks, and a set of LASSO-selected baseline covariates, and are weighted to reflect intensive tracking. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. p-value in the last row corresponds to a test for whether the treatment in question has interference effects on any arm, including control.

Table A.19: Spillover effects: consumption

	Treatment		
	HD	GD Main	GD Huge
Direct effects of treatment at saturation level of zero			
Direct effect	0.01	0.29	0.20
	(0.12)	(0.11)	(0.20)
	[1.00]	[0.23]	[1.00]
Spillover effects of treatment onto control individuals			
Spillover to control	0.05	0.02	−0.36
	(0.16)	(0.18)	(0.25)
	[1.00]	[1.00]	[0.75]
Additional effect of treatment onto individuals assigned to...			
HD	0.16	−0.05	−0.01
	(0.20)	(0.19)	(0.32)
	[1.00]	[1.00]	[1.00]
GD main	−0.21	−0.08	0.53
	(0.17)	(0.19)	(0.28)
	[1.00]	[1.00]	[0.40]
GD large	−0.04	0.18	1.22
	(0.33)	(0.30)	(0.59)
	[1.00]	[1.00]	[0.37]
Saturation mean	0.36	0.36	0.09
Saturation SD	0.23	0.23	0.13
p-value	0.68	0.90	0.11

Notes: Each column describes the direct and spillover effects of a specific treatment on Household Consumption (IHS); all results in the table are from a single estimation. Saturation mean and standard deviation correspond to the distribution of saturation rates for the treatment in question. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. p-value in the last row corresponds to a test for whether the treatment in question has interference effects on any arm, including control.

Table A.20: Impact on Business Outcomes

			GiveDirectly						
	HD	Lower	Middle	Upper	Large	Combined	Control Mean	Obs.	R^2
Number of businesses	-0.01	0.43***	0.58***	0.59***	0.58***	0.62***	1.40	1770	0.09
	(0.09)	(0.13)	(0.13)	(0.13)	(0.14)	(0.12)			
Household workers	0.16**	0.21**	0.21**	0.45***	0.41***	0.24**	0.27	1770	0.03
	(0.06)	(0.09)	(0.08)	(0.12)	(0.10)	(0.10)			
Non-household workers	0.20	0.12	0.09	0.50	0.24*	0.34*	0.22	1770	0.02
	(0.15)	(0.13)	(0.11)	(0.31)	(0.13)	(0.19)			
Own work days per month	2.71**	8.01***	6.65***	11.59***	11.37***	8.63***	9.11	1770	0.07
	(1.07)	(1.51)	(1.77)	(1.70)	(1.76)	(1.50)			
Monthly customers	12.61	75.78**	54.03**	34.68**	41.11***	38.22**	31.13	1770	0.02
	(12.44)	(30.91)	(25.10)	(14.59)	(15.19)	(14.96)			
Daily sales	5.57	9.81***	13.79***	14.42***	10.22***	18.53***	6.64	1770	0.02
	(4.38)	(3.46)	(3.21)	(4.15)	(2.87)	(4.97)			
Monthly profits	6.22***	9.81***	9.86***	12.08***	10.90***	8.81***	6.11	1770	0.03
	(2.32)	(2.58)	(2.68)	(2.65)	(2.37)	(1.72)			

Note: Table uses data from the beneficiary enterprise survey, summing values across all businesses reported. Variables such as number of employees and customers may be subject to double-counting across businesses. Outcomes are overwritten with zeros for beneficiaries who do not operate any business. All monetary values are in US Dollars. Standard errors are in parentheses, clustered at the household level.

Table A.21: Cost-Effectiveness (benefit per $100 spent), Primary outcomes

		GiveDirectly					p-values			
	HD	Lower	Middle	Upper	Large	Combined	(a)	(b)	(c)	(d)
Employed	0.007	0.008	0.010	0.000	0.001	0.001	0.85	0.95	0.57	0.94
	(0.010)	(0.012)	(0.009)	(0.008)	(0.005)	(0.005)				
Productive hours	0.838	0.699	1.332	0.603	0.132	0.275	0.14	0.82	0.33	0.63
	(0.471)	(0.593)	(0.489)	(0.427)	(0.244)	(0.241)				
Monthly income	0.094	0.192	0.220	0.194	0.086	0.124	0.16	0.31	0.24	0.41
	(0.077)	(0.092)	(0.069)	(0.059)	(0.041)	(0.038)				
Productive assets	0.463	0.998	0.775	0.650	0.475	0.526	0.00	0.00	0.00	0.42
	(0.107)	(0.118)	(0.102)	(0.078)	(0.055)	(0.052)				
HH consumption per capita	0.016	0.050	0.054	0.040	0.042	0.032	0.37	0.12	0.67	0.31
	(0.018)	(0.020)	(0.018)	(0.013)	(0.009)	(0.008)				

Note: Table gives the impact per $100 spent, which is calculated by dividing the estimated ITT impacts by the cost per arm in hundreds of dollars. The standard errors in the table are similarly the ITT SEs divided by costs. Reported p-values in final three columns derived from F-tests of hypotheses that cost-benefit ratios are equal between: (a) joint test across all arms, (b) GD Lower and HD; (c) GD Lower and GD Large; and (d) GD Large and Combined arms.

Table A.22: Cost-Effectiveness (benefit per $100 spent), Secondary outcomes

		GiveDirectly					p-values			
	HD	Lower	Middle	Upper	Large	Combined	(a)	(b)	(c)	(d)
Panel A. Beneficiary welfare										
Subjective well-being	0.058	0.101	0.107	0.081	0.066	0.048	0.09	0.08	0.12	0.17
	(0.020)	(0.024)	(0.020)	(0.016)	(0.011)	(0.010)				
Mental health	−0.013	−0.016	0.014	0.005	0.013	0.014	0.60	0.89	0.22	0.93
	(0.022)	(0.023)	(0.019)	(0.016)	(0.011)	(0.011)				
Beneficiary-specific consumption	0.045	0.129	0.125	0.105	0.053	0.082	0.00	0.01	0.01	0.11
	(0.035)	(0.030)	(0.027)	(0.020)	(0.018)	(0.014)				
Panel B. Household wealth										
HH net non-land wealth	−0.054	0.041	0.244	0.226	0.131	0.106	0.05	0.56	0.53	0.67
	(0.120)	(0.150)	(0.089)	(0.070)	(0.049)	(0.057)				
HH livestock wealth	−0.002	0.445	0.374	0.447	0.257	0.263	0.00	0.00	0.12	0.92
	(0.110)	(0.123)	(0.105)	(0.075)	(0.056)	(0.053)				
Savings	0.311	0.264	0.262	0.265	0.169	0.202	0.14	0.56	0.22	0.39
	(0.070)	(0.080)	(0.068)	(0.051)	(0.036)	(0.032)				
Debt	0.122	−0.024	−0.046	−0.095	−0.043	0.000	0.18	0.19	0.86	0.45
	(0.084)	(0.105)	(0.088)	(0.076)	(0.050)	(0.045)				
Panel C. Beneficiary cognitive and non-cognitive skills										
Locus of control	0.019	0.032	0.005	0.000	0.010	0.027	0.65	0.59	0.30	0.12
	(0.018)	(0.021)	(0.017)	(0.014)	(0.010)	(0.009)				
Aspirations	−0.002	0.020	−0.009	0.022	0.003	0.016	0.52	0.41	0.48	0.26
	(0.022)	(0.024)	(0.019)	(0.015)	(0.010)	(0.009)				
Big Five index	0.035	0.019	0.022	0.003	−0.009	0.003	0.17	0.55	0.25	0.35
	(0.020)	(0.025)	(0.019)	(0.015)	(0.011)	(0.010)				
Business knowledge	0.197	0.022	0.016	0.011	−0.003	0.074	0.00	0.00	0.29	0.00
	(0.022)	(0.024)	(0.019)	(0.016)	(0.011)	(0.011)				
Business attitudes	0.037	0.049	0.038	0.017	0.007	0.018	0.25	0.63	0.08	0.40
	(0.020)	(0.024)	(0.018)	(0.015)	(0.011)	(0.010)				

Note: Table gives the impact per $100 spent, which is calculated by dividing the estimated ITT impacts by the cost per arm in hundreds of dollars. The standard errors in the table are similarly the ITT SEs divided by costs. Reported p-values in final three columns derived from F-tests of hypotheses that cost-benefit ratios are equal between: (a) joint test across all arms, (b) GD Lower and HD; (c) GD Lower and GD Large; and (d) GD Large and Combined arms.

Table A.23: Program impacts on household-to-household transfers

	HD	GiveDirectly Lower	GiveDirectly Middle	GiveDirectly Upper	GiveDirectly Large	Combined	Control Mean	Obs.	R^2	p-values (a)	p-values (b)	p-values (c)
HH loans made	0.11	0.90**	0.45	1.41***	0.68*	1.24***	2.24	1705	0.15	0.10	0.18	0.27
	(0.30)	(0.44)	(0.44)	(0.45)	(0.42)	(0.43)						
	[0.31]	[0.05]	[0.19]	[0.00]	[0.09]	[0.01]						
HH gifts made	−0.33	1.75***	2.67***	2.62***	3.11***	2.46***	4.90	1704	0.17	0.00	0.60	0.33
	(0.36)	(0.57)	(0.55)	(0.59)	(0.53)	(0.50)						
	[0.19]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]						
HH gifts made	0.39	0.68*	1.87***	0.56	0.32	0.81*	3.41	1675	0.15	0.63	0.20	0.36
	(0.32)	(0.42)	(0.48)	(0.47)	(0.44)	(0.45)						
	[0.15]	[0.09]	[0.00]	[0.15]	[0.20]	[0.08]						

Note: Regressions include but do not report the lagged dependent variable, fixed effects for randomization blocks, and a set of LASSO-selected baseline covariates, and are weighted to reflect intensive tracking. Standard errors are (in soft brackets) are clustered at the household level to reflect the design effect, and p-values corrected for False Discovery Rates across all the outcomes in the table are presented in hard brackets. Stars on coefficient estimates are derived from the FDR-corrected p-values, *=10%, **=5%, and ***=1% significance. Reported p-values in final three columns derived from F-tests of hypotheses that cost-benefit ratios are equal between: (a) GD Lower and HD; (b) GD Lower and GD Large; and (c) GD Large and Combined treatments.

Figure A.1: Project timeline

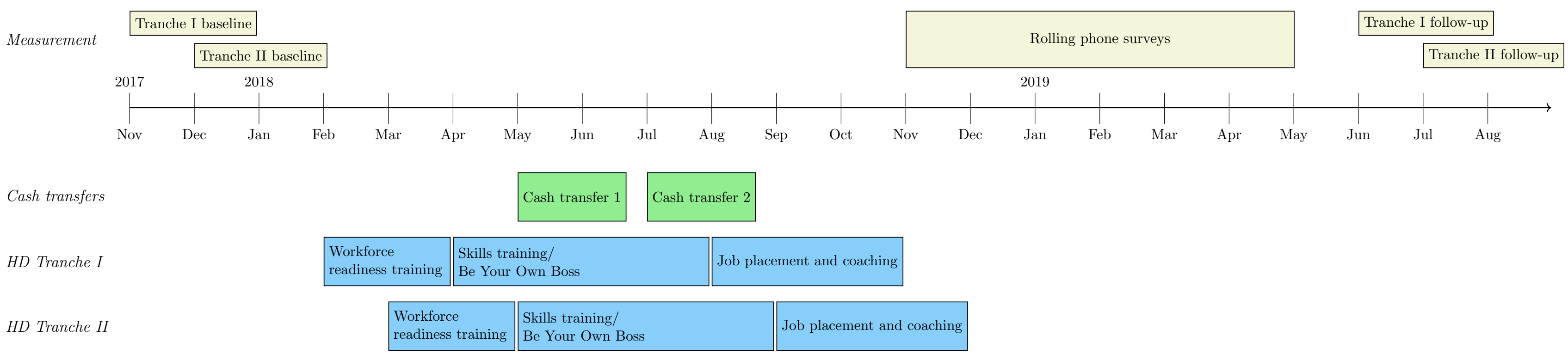


Figure A.2: CDF of Savings Stocks (IHS)

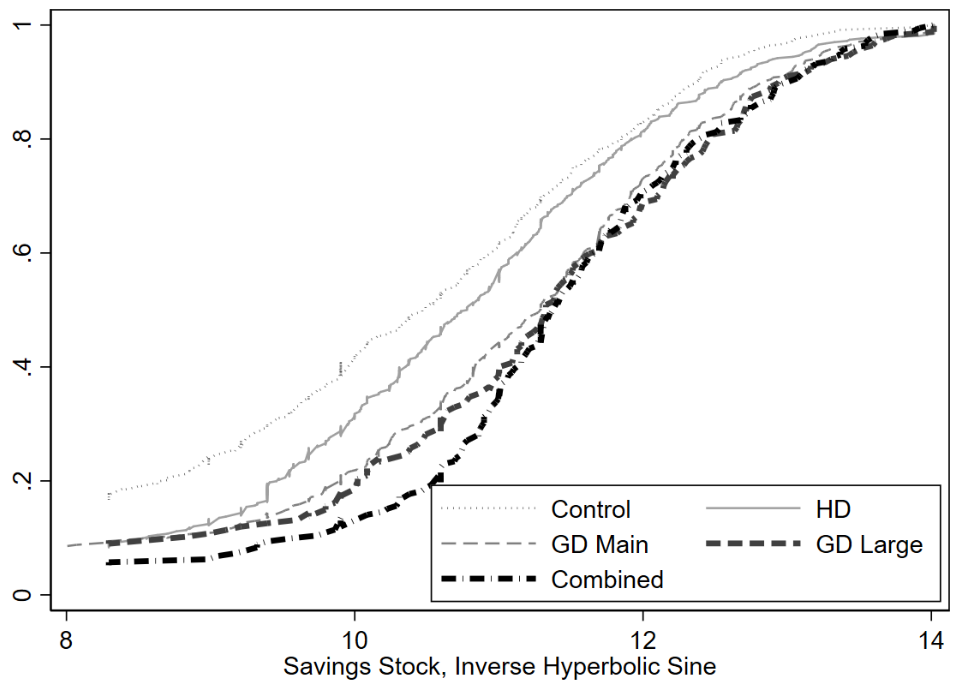


Figure A.3: CDF of Productive Hours

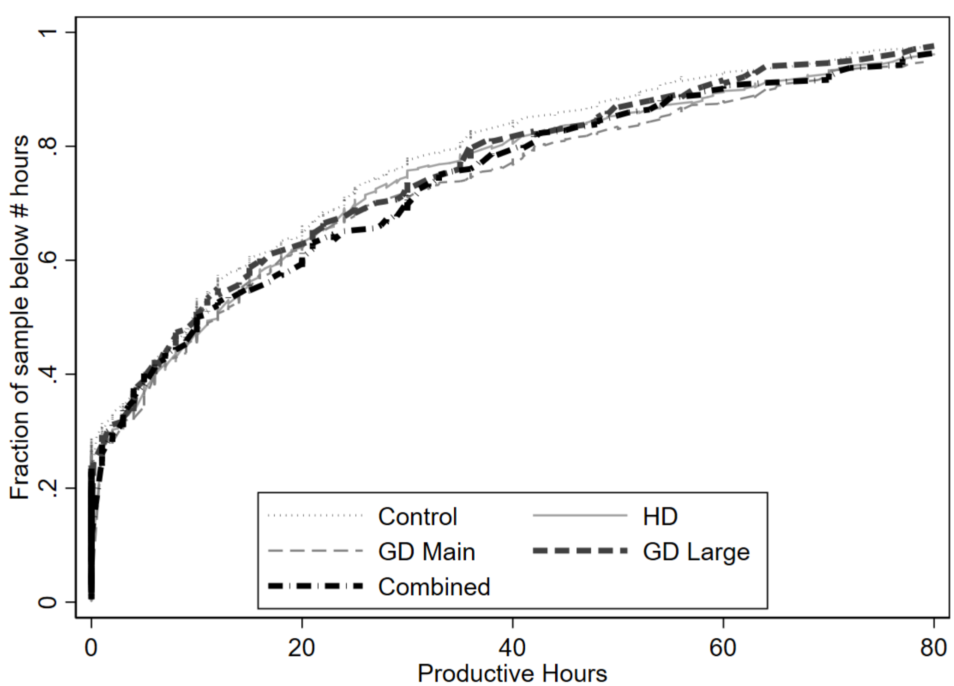


Figure A.4: Non-Ag Wage Employment, varying hours thresholds

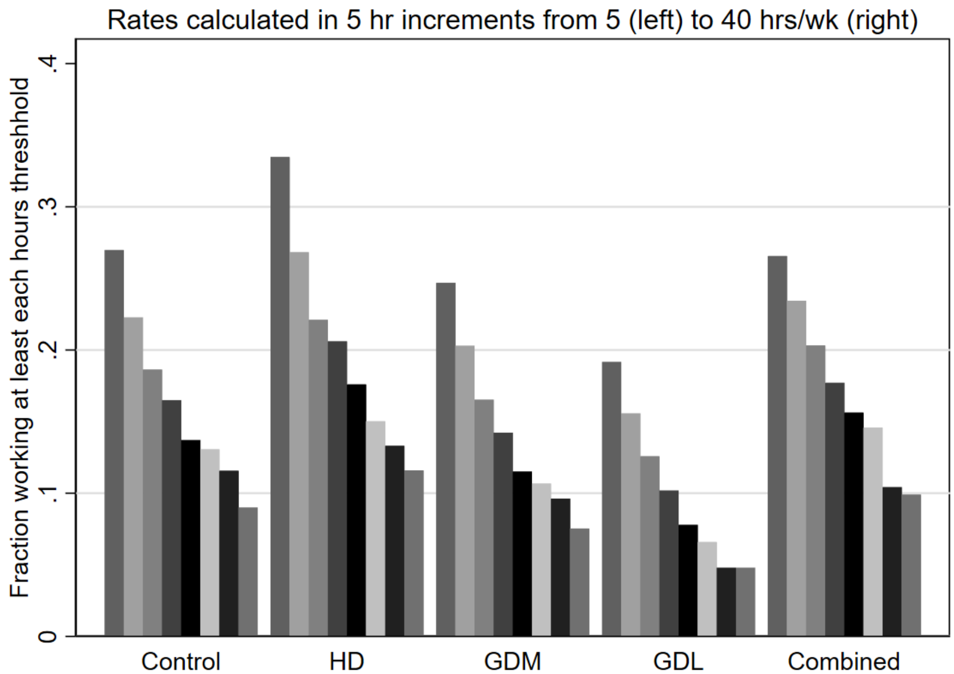


Figure A.5: Non-Ag Self Employment, varying hours thresholds

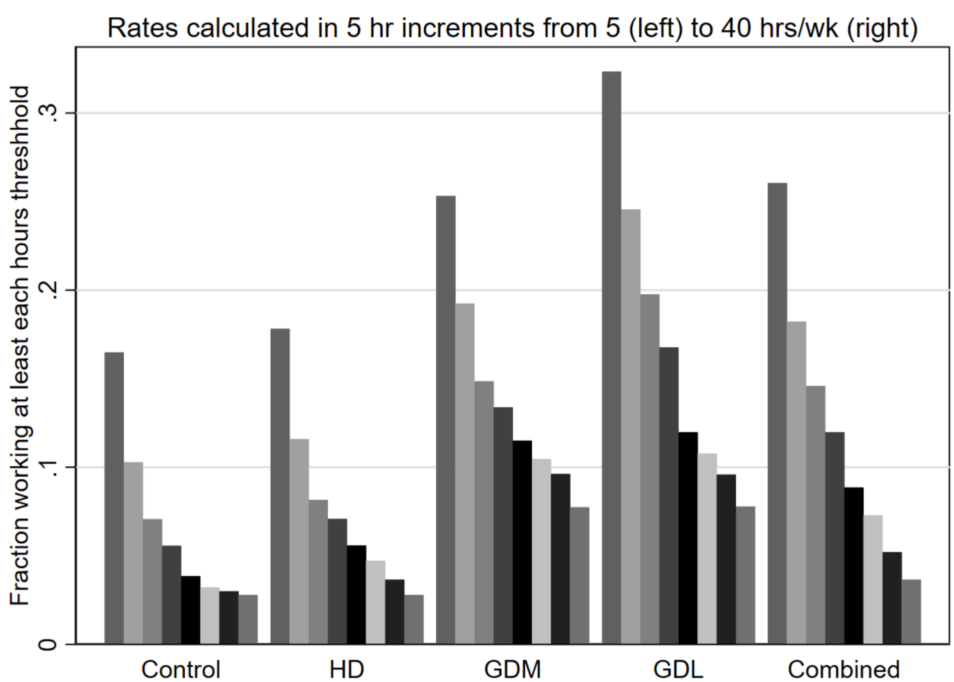


Figure A.6: Ag Wage Employment, varying hours thresholds

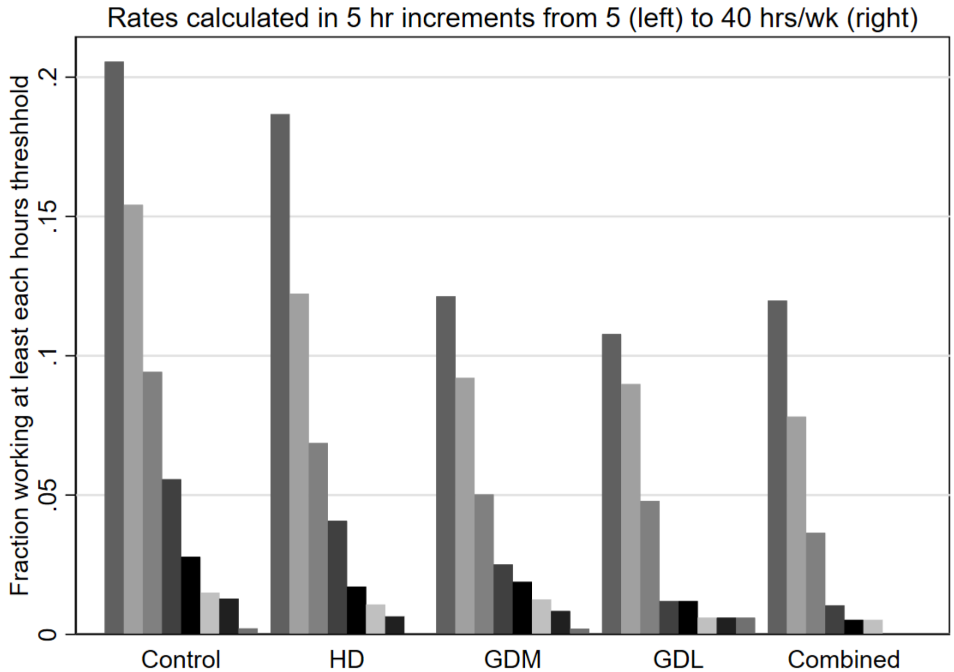


Figure A.7: Ag Self Employment, varying hours thresholds

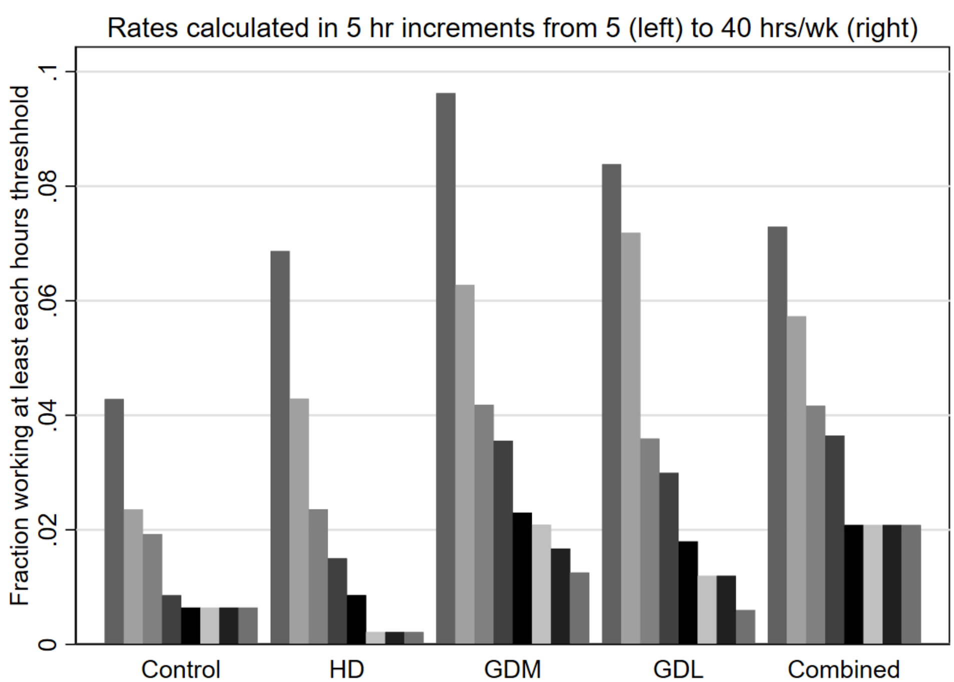


Figure A.8: Ag processing or trading Employment, varying hours thresholds

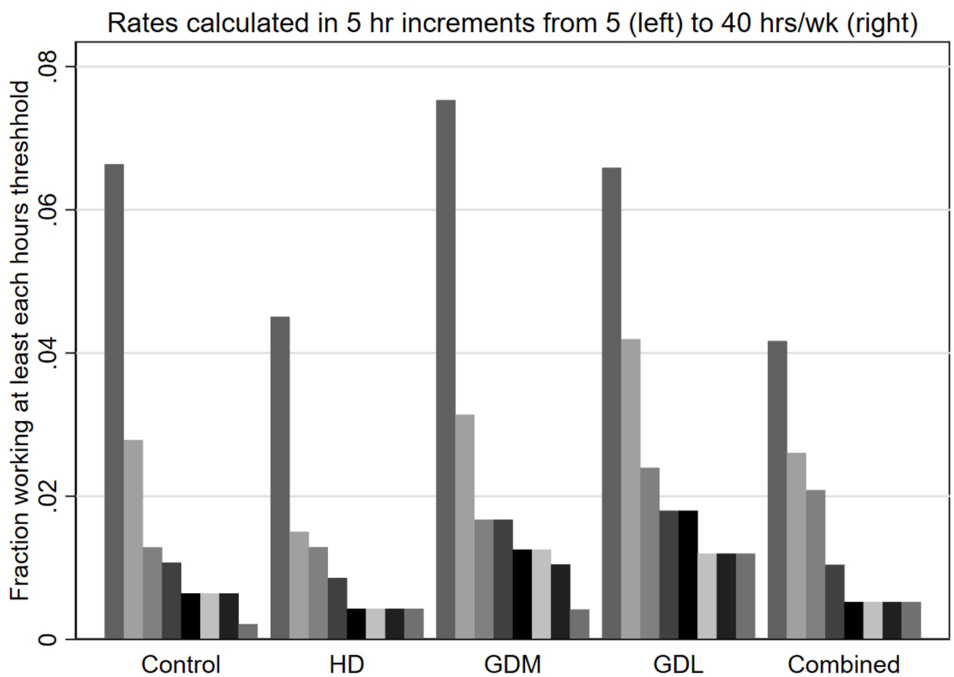


Figure A.9: Rolling Phone Survey: Monthly impacts on Productive Hours

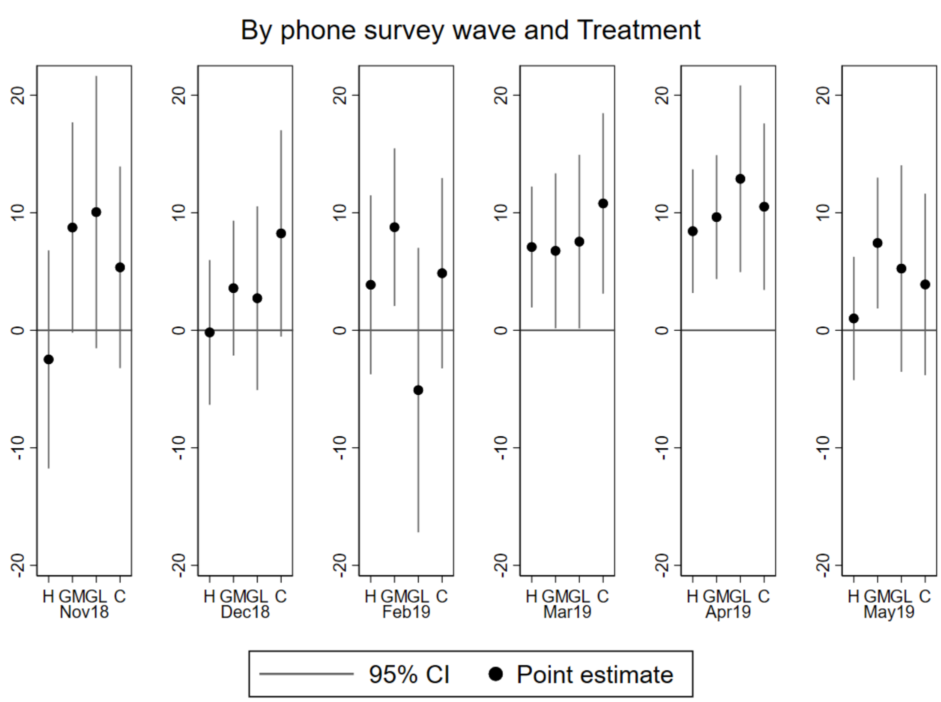


Figure A.10: Rolling Phone Survey: Monthly impacts on Apprenticeship Hours

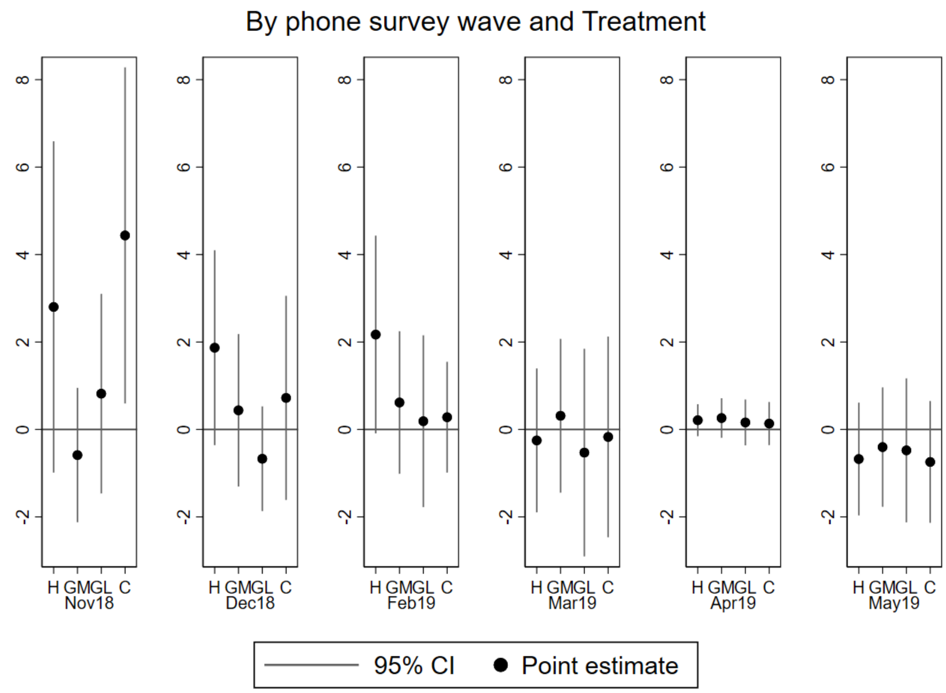


Figure A.11: Cost Equivalence on Secondary Outcomes

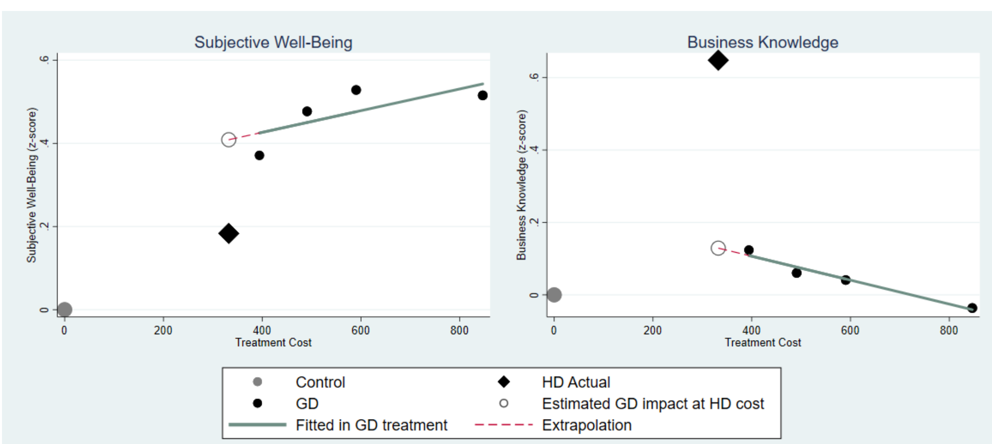
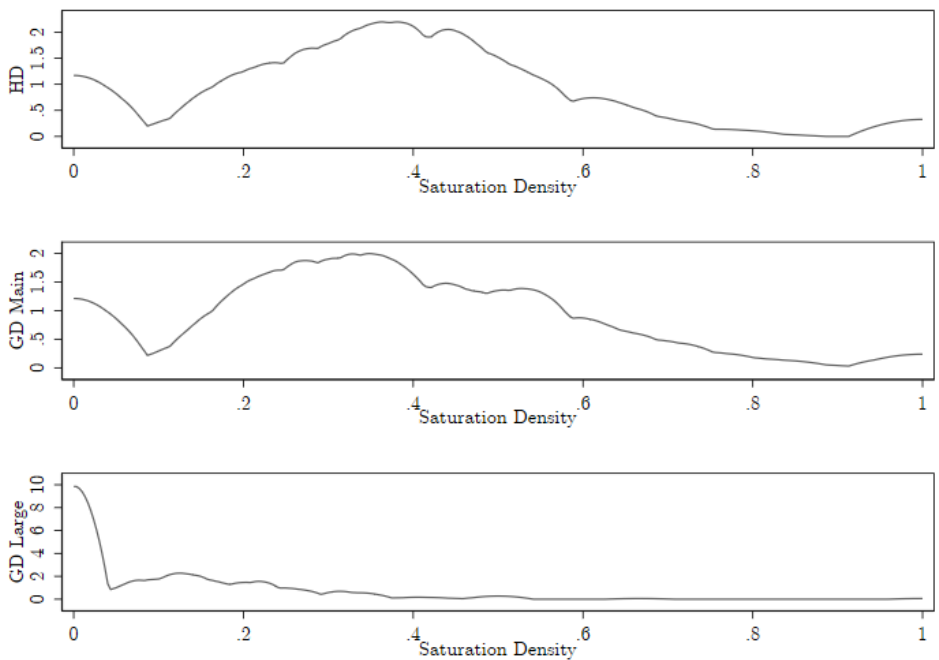


Figure A.12: Village-Level Treatment Saturations



Online Appendices

Appendix B Outcome definitions

Appendix B.1 Defining Primary outcomes

For each of the outcomes defined below, we provide a definition, followed by an explanation of how that measure will be constructed from survey data. Survey questions either begin with a 'B-' for the beneficiary instrument or a 'H-' for the household instrument, followed by the two-digit section number, followed by 'q' and the question number. These refer to the beneficiary and household instruments, respectively.[30]

There are five primary outcomes:

1. *Employment status.* A binary indicator variable taking a value of one if the beneficiary spent 10 hours or more in the prior week working in a wage job or as primary operator of a microenterprise. The 1 week recall is per ILO definition. Defined as 'Yes' if beneficiary spent 10 hours or more on any of the following activities:
   - Processing or trading of agricultural goods (`B02qagroprocesshrs`)
   - Agricultural (off farm) wage labor (`B02qfarmhours`)
   - Non-agricultural wage labor (`B02qnoagrichrs`)
   - Non-agricultural microenterprise (`B02qenterphrs`)
   - Microenterprise or other self employment (`B02qsemployhrs`).
2. *Off-own-farm productive time use.* Defined as the number of productive hours over the past 7 days. Sum of hours from questions:
   - Processing or trading of agricultural goods (`B02qagroprocesshrs`)
   - Agricultural (off farm) wage labor (`B02qfarmhours`)
   - Non-agricultural wage labor (`B02qnoagrichrs`)
   - Non-agricultural microenterprise (`B02qenterphrs`)
   - Microenterprise or other self employment (`B02qsemployhrs`)
   - Apprenticeship (`B02qapprenticehrs`)
3. *Beneficiary's (monthly) income.* Defined as the sum of the following monthly recall questions:
   - Agricultural own-farm income (`B02qagricearn`)
   - Agricultural wage income (`B02qfarmwage`)
   - Non-agricultural wage income (`B02qnoagricwage`)
   - Microenterprise profits (`B02qenterpwage` + `B02qsemploywage`);

[30] In the electronic survey instrument, all variables begin with an 'm' prefix, but this notation does not guarantee uniqueness across instruments. Consequently for the purposes of this PAP we adopt the 'B-' and 'H-' convention above.

- Livestock rearing income (`B02qlivestockwage`)
- Agricultural processing and trading income (`B02qagroprocessearn`)
- Apprenticeship income (`B02qapprenticewage`)

  This outcome will be winsorized at the 1st and 99th percentile, and we will take the inverse hyperbolic sine transformation of this as the primary measure.

4. *Productive assets* under beneficiary control. (Sum of asset values from beneficiary enterprise module that are reported as used in the beneficiary's business, Section `B05`: tools, machinery, furniture, inventories, and other physical assets.) This outcome will be winsorized at the 1st and 99th percentile, and we will take the inverse hyperbolic sine transformation of this as the primary measure.

5. *Household consumption* per capita. Sum of monthly purchase values of Section `H10`, divided by adult-equivalent household members. This outcome will be winsorized at the 1st and 99th percentile, and we will take the inverse hyperbolic sine transformation of this as the primary measure.

The first three of these primary outcomes provide direct measures of the extent to which a study participant is productively employed: their formal (non-farm) employment categorization, their productive time use, and their earnings. To the extent that these measures are potentially seasonal in nature, one might worry that interventions could differentially affect the sectoral composition of employment, and that differential seasonality across these would tip the scales in favor of one or the other mode of intervention. More broadly, income may be more fully measured in one sector relative to another. Such concerns are partly addressed by the inclusion of household consumption as a primary outcome: to the extent that beneficiaries smooth consumption, household consumption will be less susceptible to such concerns. In addition, we will include as a robustness check an analysis of impacts on a rolling panel of employment status measures, collected over the six months prior to the endline.

One potential challenge for the analysis of monetary outcomes (income, assets, and consumption) is that, if treatments induce migration, they may cause subjects to face different prices. Such differences in prices could cause the study to over- (or under-)state the the real value of estimated impacts. On the other hand, deflating values to control-group prices is not straightforward, for at least two reasons: study subjects may alter the *quality* of products purchased in ways not captured by the study, therefore giving the appearance of price impacts; and study subjects may earn incomes in more expensive locations but intend for part of that income may be consumed—by the subject themselves, or by family members to which they remit income—in their place of origin. To address these concerns, we will report as a robustness check an analysis of primary outcomes (3)–(5) that uses control-group prices to deflate these values. This will be particularly important to the interpretation of the study results if treatments have effects on migration.

Appendix B.2 Defining Secondary outcomes

We propose to analyze three families of secondary outcome: one which speaks to alternative measures of beneficiary welfare; a second that speaks to wealth effects that may indicate likely long-term benefits; and a third family that highlights key mechanisms of interest.

1. **Alternative measures of beneficiary welfare**

   Within this family, we consider the following alternative measures of beneficiary well-being:

   (a) Subjective well being: Index of responses to `B10_swb_happiness` and `B10_swb_lifesatisfaction`, constructed as the average of z-scores.

   (b) Mental health: Index of section `B11` responses. Z-score of the simple average across all questions for each beneficiary.

   (c) Beneficiary-specific consumption expenditures (sum of values from Section B08). This outcome will be winsorized at the 1st and 99th percentile, and we will take the inverse hyperbolic sine transformation of this as the primary measure.

2. **Household net wealth, and its components**

   Like productive assets, the accumulation and protection of household wealth. Conditional on this, households' access to borrowing opportunities—viewed as a measure of their financial access—may be a mechanism through which the interventions studied are multiplied. Given this welfare ambiguity, we propose to analyze both total household net (non-land) wealth, as well as stocks of savings and debt, taken individually.

   (a) Household net non-land wealth. Sum of values of household assets (`H12`), plus savings value (`H06`), value of loans outstanding that are expected to be repaid (`H08`), less debt value (`H07`). This outcome will be winsorized at the 1st and 99th percentile, and we will take the inverse hyperbolic sine transformation of this as the primary measure.

   (b) Total value of all livestock wealth. Sum of values of household livestock assets (`H12`). Specifically, summing over values derived from `H12_oxen` through `H12_ducks` in the household instrument. This outcome will be winsorized at the 1st and 99th percentiles, and we will take the inverse hyperbolic sine transformation of this as the primary measure.

   (c) Stock of savings. Beneficiary stock of savings, sum of values in `B06`. Plus household stock of savings from analogous questions (`H06`). This outcome will be winsorized at the 1st and 99th percentile, and we will take the inverse hyperbolic sine transformation of this as the primary measure.

   (d) Stock of debt. Beneficiary sum of borrowed amounts from all (formal and informal) sources (`B07`), plus household borrowings from analogous questions (`H07`). This outcome will be winsorized at the 1st and 99th percentile, and we will take the inverse hyperbolic sine transformation of this as the primary measure.

3. **Cognitive and non-cognitive skills**

   A specific feature of the theory of change that motivates EDC's curriculum is that a focus not just on specific skills, but on non-cognitive attitudes and attitudes, may make that intervention more likely to have persistent effects. At the same time, cash transfers may also change, inter alia, beneficiaries' sense of control and aspirations. To test these mechanisms, we define the following family of secondary outcomes:

   (a) Locus of control: Index of responses to `B09`. Z-score of the simple average across all questions for each beneficiary.

   (b) Aspirations: Index of responses to `B13`. Z-score of the simple average across all questions for each beneficiary.

   (c) Conscientiousness, agreeableness, and emotional stability from BFI (Section `B12`). Each index is the Z-score of the simple average of the questions related to the corresponding dimension. Following EDC's analysis of Akaze Kanoze employers,[31] we will examine program impacts on the three most highly-rated components of the Big-Five Index from employers' perspective: conscientiousness, agreeableness, and emotional stability.

   (d) Business knowledge. Index of `B14`. Z-score of the simple average across all questions for each beneficiary.

   (e) Business attitudes. Index of `B15`. Z-score of the simple average across all questions for each beneficiary.

Appendix C Selection of Control Variables

In our pre-analysis plan, we state that control variables for the primary specification "will be selected on the basis of their ability to predict the primary outcomes". In doing so, we seek to build on recent developments that balance the challenge of using baseline data to select variables that will reduce residual variance in equation (1) with the danger that researcher freedom in the selection of control variables can lead to p-hacking, in which right-hand-side variables are selected specifically on the basis of the statistical significance of the coefficient of interest (Card and Krueger, 1995; Casey et al., 2012), thereby invalidating inference.

To balance these concerns, we adapt the *post-double-selection* approach set forth in Belloni et al. (2014b, henceforth BCH). BCH advocate a two-step procedure in which, first, Lasso is used to automate the selection of control variables, and second, the post-Lasso estimator (Belloni et al., 2012) is used to estimate the coefficients of primary interest in Equation (1), effectively using Lasso as a model selection device but *not* imposing the shrunken coefficients that results from the Lasso estimates directly. Belloni et al. (2014b) demonstrate that this approach not only reduces bias in estimated treatment effects better than alternative approaches—less a concern given the successful

[31] Povec Pagel, Olaru, Alcid, and Beauvy-Sany, 2017, "Identifying cross-cutting non-cognitive skills for positive youth development", Final report, Education Development Center, Inc.

randomization in our experiment—but that it may improve power while retaining uniformly valid inference.

In the first stage, model selection is undertaken by retaining control variables from the union of those chosen either as predictive of the treatment assignment or of the outcome. This model selection stage can be undertaken after residualizing to account for a set of control variables that the authors have a priori determined below in the model, as in Belloni et al. (2014a). In our case, we retain block fixed effects, lagged values of the outcome, and lagged values of (the inverse hyperbolic sine of) household wealth in all specifications, per our pre-analysis plan.

We modify the BCH approach for application to a randomized experiment in three ways. First, again following (Jones et al., 2019), for each outcome we choose the Lasso penalty parameter that minimizes the 10-fold cross-validated mean squared error. Second, to ensure that chance differences in the leverage of observations across different covariate sets do not lead to different conclusions about the (relative) impacts of treatment across different outcomes (Young, 2019), we take the union of covariate sets selected to be predictive of the five primary outcomes of the study, and use these as controls for all outcomes. And third, we modify the heteroskedasticity-robust Lasso estimator of Belloni et al. (2012) to incorporate sampling weights consistent with our design.[32]

The set of *potential* covariates is determined as follows:

- Baseline values of all primary outcomes, including the individual components of the employment status, productive time use, monthly income variables outlined in Section 2.8;
- Baseline values of all secondary outcomes,
- Baseline values of all dimensions of heterogeneity pre-specified in Section 3.4.
- The number of study participants (in any arm of the study) in an individual's village, which is defined as the measure of network 'degree' for each individual in the spillover analysis of Section 3.5.

All variables are normalized prior to inclusion in the selection routine, to have mean zero and variance of one in the baseline sample. We include squares of all continuous variables and all pairwise interactions among the potential covariates above, and between the potential covariates above and the set of variables that force the routine to include without penalty To ensure that sample size is not affected by the choice of covariates, we impute values of zero for all variables in the *potential* covariate list, and for each potential covariate we include an indicator for whether such an imputation was undertaken among the list of potential covariates to be fed into the BCH first-stage selection procedure.

[32]Specifically, we up-weight observations in our 'intensive tracking' endline sample by the inverse of the fraction of not-initially-reached individuals in the follow-up survey who were then assigned to intensive tracking.

Appendix D Administrative Information

Appendix D.1 Funding

All research funding for this project was provided by USAID.

Appendix D.2 Details of Study Participant Selection

To meet the Huguka Dukore eligiblity criteria, participating youth must meet the following criteria:

- 6-12 years of basic education (inclusive).
- Age 16-30 at enrollment.
- Drawn from Ubudehe poverty groups 1 and 2, per GiveDirectly's remit from the Rwandan Government to treat only the poorest households with cash transfers.

Additionally, HD in its outreach specifically targeted the following criteria for inclusion, meaning that such youth will be specially recruited to participate:

- Out of school for three consecutive years
- Income of less than $1.75 per day
- Youth exhibiting some form of disability (that can be accommodated in HD programming)
- Women.
- Youth who have not benefited from related interventions in the past.

Hard eligibility criteria and targeted characteristics were provided to local government leaders, who provided lists of potential candidates to EDC. Those candidates were then invited to the information session and formally screened for eligibility.

All listing and determination of eligibility were conducted by EDC via an 'over-subscription' process. Under this protocol, EDC enrolled more eligible individuals than they were able to treat with HD, in order to generate the samples for the alternate (household grants) arm and the control. In the end we recruited 1848 study youth from approximately 250 villages in our 13 sectors, for an average of roughly 7.4 study individuals per village.

Below, we characterize the process for (over)subscription, which delivered the sample of individuals for the baseline.

1. **Sector-level meeting** to discuss HD with local leaders that introduced the study. In this meeting, sector officials were fully informed about the scope of the study, emphasizing the separateness of the two interventions and implementers.

2. **Announcement to the community** in public places (churches, community halls) or a meeting to engage potential beneficiaries. At this point only the HD program wsa described to beneficiaries, and with only general language about the household grants arm. Guiding language: *"We are pleased to be able to bring programming to this community that seeks to improve the livelihoods of vulnerable youth. To this end, we are requesting the names and contact details of youth meeting the following criteria: ¡insert eligibility criteria here¿. Participating youth should be willing and interested to join an employment skills program, called HD, that will provide training and work experience to participants."*

3. **Screening of youth** by the selection committee which produced the final list of potential beneficiaries that was passed to local implementing partners (IPs).

4. **Invitation of potential beneficiaries to an orientation meeting**. The language of this invitation reflected the fact that potential beneficiaries were not guaranteed places in HD, and might be randomly allocated to a different program or the control. Guiding language for official communication: *"We have determined that you are eligible for the Huguka Dukore program. There may be more eligible individuals than Huguka Dukore will treat this year, so you are not yet guaranteed a place, though some of those not treated by Huguka Dukore will be supported by another NGO. To find out more about the Huguka Dukore program and to take the next step toward this opportunity, please attend an orientation meeting at XXX on YYY date."*

5. **Orientation and awareness meeting** with selected youth by local IPs at which they are given further explanation about the program. In HD's other districts, these orientation meetings convey information about the scope of that program, under a presumption that those who participate in the orientation meeting can have a place in HD should they choose to take it up.

6. **Description of the lottery for program assignment.** The lottery is described during this meeting with reference to another intervention providing livelihoods assistance that will also be determined by the lottery. Guiding language: *"Today you have learned more about the Huguka Dukore program. This is one of two programs that are being delivered by distinct NGOs, in coordination with Sector and District officials, both of which seek to improve livelihoods for vulnerable youth. If you decide that you are interested in participating in one of these programs, there is one more step in the selection process. To participate, you must attest that you have the time and interest required to participate in Huguka Dukore. Your name will then be entered into a pool of applicants. There will be a public meeting in which a lottery will be used to determine which of these applicants receives a place in HD. You may attend this meeting if you wish, but you do not have to do so in order to gain a place. Not all whose names are entered into the lottery will be placed in HD. Some of those who participate in the lottery will be passed to a second NGO, which provides assistance to individuals seeking*

to improve their livelihoods. Those who receive a place in either program will be contacted directly by the relevant organization after the lottery. To gain access to either program, you must participate in this lottery. If you are willing to participate, please provide your name and contact details in writing. Prior to the lottery, you may be contacted by an independent research organization called Innovations for Poverty Action, who are conducting a survey of potential beneficiaries. You do not have to participate in this survey in order to gain access to our program, and participation will not affect your chances of enrollment. However, we would be grateful for your willingness to participate in an interview with IPA, which will help us to understand the design and impacts of our work.

7. **Registration for the lottery assignment.** To correctly reflect the lottery process to participants, they were told when asked to enroll in the study that it is "a lottery in which you will have a chance of receiving HD, a chance of receiving assistance from a different organization that gives household grants, and a chance that you do not receive either program." Individuals who do not choose to register for the study will not be excluded from receiving HD if they are eligible & choose to participate.