Journal of Development Economics
Registered Report Stage 1: Proposal

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

Revision submitted: November 17, 2019

Abstract

Using a randomized experiment, this study will benchmark a youth employment training program, Huguka Dukore (HD), to cash grants in three Rwandan districts. The study population consists of underemployed youth, a high-risk demographic who may display powerful long-term benefits from additional knowledge and resources. The study enrolled individuals who meet HD eligibility criteria and express willingness to enroll in that program at baseline. Impacts will be measured 18 months after baseline. Primary outcomes are employment status, time use, beneficiary income, household consumption, and productive assets. The study randomizes cash transfer amounts around expected HD costs per beneficiary to estimate cost-equivalent impacts, equating expenditure per beneficiary across arms. In addition, a larger transfer was chosen to maximize the cost-effectiveness of cash given fixed costs. A final arm receives both HD and cash. Careful tracking of costs will allow cleanly benchmarking HD to an exactly cost-equivalent cash grant.

Keywords: Experimental Design, Cash Transfers, Employment
JEL Codes: O12, C93, I15
Study pre-registration: AEA Trial Registry Number AEARCTR-0004388

Contents

1 Introduction

The demographic dividend in Sub-Saharan Africa is a double-edged sword. A young population provides an opportunity to invest in the future, with a relatively limited burden of dependency from older generations—but not if young people are unable to find productive employment. 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). Hence, it is critical to understand the barriers in physical and human capital that prevent youth from being fully productive.

In spite of this pressing need, policymakers have limited access to evidence-based interventions with a track record of effectiveness. This is not to say that active labor-market interventions have not been studied; for example, a recent review discusses nine randomized evaluations from developing countries (McKenzie, 2017). Despite some signs of success in generating employment (Alfonsi et al., 2019; Diaz and Rosas, 2016), the impacts of programs aimed at lifting human capital have been variable and less impressive than hoped in terms of labor and income benefits. 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 (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. Given that the literature has long recognized both 'money and ideas' may serve as constraints to the productivity of young entrepreneurs (Giné and Mansuri, 2014), rigorous comparative cost effectiveness research across these different modalities is sorely needed.

This study addresses these challenges by undertaking an exercise in *cash benchmarking*: the direct comparison of in-kind- to cash-transfer programs in a single experimental setting. As an applied-science exercise, such a study is a form of comparative cost-effectiveness analysis; it compares the returns to alternative forms of programming on a pre-defined set of outcomes. And it does so subject to a distributional constraint, holding the value of programming per beneficiary constant across modalities. Such cash-benchmarking exercises also inform a basic-science question, by lifting distinct constraints to individual employment outcomes. Similar efforts include 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. One way of conceiving of the value of this benchmarking activity is that 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. We can not only determine the benefit generated by an increment of skills improvement, but we can calculate the counterfactual 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 capital constraints improve when human capital constraints have also been relaxed.

We study this question using an individually 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 to an additional arm that receives household grants of equal cost to the donor—a *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 (HD), which means 'work well done' in Kinyarwanda; it follows USAID's strategy on workforce readiness and skills training and was implemented by Education Development Center, Inc. (EDC). The benchmarking cash transfer program was implemented by GiveDirectly (GD), a US-based nonprofit that specializes in making unconditional household grants via mobile money. 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. Our 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, and the inclusion of larger cash transfer amounts allows us to learn about how the cost effectiveness of cash transfers varies with transfer value.

USAID's Huguka Dukore program is a particularly attractive candidate for a benchmarking evaluation. 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, youth with disabilities, HD will offer multiple program pathways including: i) employment preparation; ii) individual and cooperative youth microenterprise start-up; iii) business development for existing microenterprises. HD is based on a predecessor *Akazi Kanoze* program, which has been operating in the country for the past five years, and which was evaluated as successful in a recent RCT led by the implementers (Alcid, 2014). In its future plans, Government of Rwanda places a high priority on such 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). And training programs of this sort are widespread across the developing world: Blattman and Ralston (2015) estimate that the World Bank alone spends almost a billion dollars annually on skills training programs. But in spite of their prevalence, the cost-effectiveness of such programs is far from certain. In a recent review of evidence on active labor market programs that operate on the supply side of the labor market, McKenzie (2017) finds that employment and earnings impacts are modest, with costs averaging 50 times the monthly income gain. And indeed, in its *Future Drivers of Growth* report, produced jointly withe the World Bank, 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). Our study seeks to resolve this uncertainty by direct comparison.

In principle, one could consider 'benchmarking' a program in two ways. One would be to consider the amount of money that needs to be spent through the each modality in order to achieve a specific benefit, such as creating one newly employed youth. From this perspective evidence such as De Mel et al. (2012), showing that one-time transfers of $100-200 (smaller than the transfers in this paper) can having lasting effects on enterprise profitability, are encouraging. We pursued the opposite approach: for a equal expenditure per beneficiary household through each modality, how do the benefits compare? An advantage of our approach is that even if the two programs have a very different profile of impacts, we can compare these profiles in an apples-to-apples manner knowing that the resources required to generate these changes were identical.

The momentum for benchmarking has built as numerous studies 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, but some studies have found substantial impacts (Aizer et al., 2016; Barham et al., 2014; Fernald et al., 2009; Hoynes et al., 2016).[1] The largest extant literature on benchmarking 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), which has uncovered a fairly consistent result 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., 2017).

The randomized, controlled trial proceeds in four steps. First, EDC's local implementing partners were to identify at least 1800 eligible individuals who could be recruited into the study, from approximately 250 villages in the three districts of Rwamagana, Muhanga and Nyamagabe. The HD-imposed eligibility criteria for their training of vulnerable youth consist of (a) ages ranging from 16–30, and (b) under twelve years of basic education (below the sixth year of secondary education), but at least six (inclusive). Because of the conditions placed on GiveDirectly by the Rwandan government, we further strictly limited eligibility to (c) households registered in Ubudehe poverty status 1 or 2. In addition, in order to provide a study that has compliance rates with the HD training that are as high as possible, we further restricted eligibility to those who (d) expressed interest in participating in the employment and entrepreneurship readiness training. These individuals were recruited at a first 'orientation' meeting at which the local HD implementers and IPA

[1]For examples of studies that find dissipating long-term benefits, see Baird et al. (2016) and Araujo 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 been similarly uneven.

recorded sufficient information to enroll them and to subsequently perform baseline surveys at the household. Second, IPA collected baseline data, implementing survey instruments that collected information both at the household level and at the individual beneficiary level for study participants. The purpose of the baseline is to establish pre-program levels of incomes and other key dimensions of livelihoods. Third, we conducted a series of 13 public lotteries at the sector level, overseen by sector- and local-level officials, at which individuals were assigned to four arms, to be treated accordingly by implementers. Fourth and finally, the study will collect a range of post-intervention outcome indicators, approximately 18 months after the baseline survey, to evaluate impacts. [2] Combining these results on program impacts with detailed information on the costs of these activities will allow the study to provide rigorous evidence of both the benefits and costs associated with each approach.

We costed both programs in detail prior to, and after, the intervention period, following Levin and McEwan (2001). 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 $452.47. 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 vs cash, and small versus large cash transfer amounts. Incorporating GiveDirectly's operating costs, the amount actually received by households that generates the same expected cost to USAID as HD is $410.19. The comparison between these two arms therefore provides a straightforward window on expected cost-equivalent impacts. Because we anticipate the exact numbers from the ex-post costing exercise will differ somewhat from the ex-ante exercise, we have randomized two bracketing cash transfer arms which transfer $317.31 and $503.04 to households. A fourth and substantially larger transfer arm transfers $750 to beneficiaries; this amount was chosen by GiveDirectly as maximizing their own cost-effectiveness given the returns to transfers and the fixed costs in providing cash transfers via mobile money. The inclusion of this arm provides a statistically high-powered way of examining how benefit/cost ratios shift as the transfer amount rises. Using the final, ex-post costing exercise, we will arrive at an exact cost per eligible household for both implementers, and use a linear regression adjustment across all four GD transfer amounts to analyze comparative impacts at exactly equivalent costs to the donor, USAID.

Our design permits a number of interesting extensions. Because of the randomized variation in transfer amounts, we can ask 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? The combined arm provides an experimental (albeit less well-powered) window into the complementarity of these two interventions, namely the differential returns to human capital as resources increase. A series of incentivized discounting games conducted at baseline allow us to look for heterogeneity across time preferences, as well as

[2] We have received funding to conduct a second endline survey 36 months after baseline; this PAP covers the 18-month endline analysis

gender, expenditures, and employment rates in local labor markets. We will use random variation in the intensity of treatment at the village level to look for evidence of spillovers, a particularly important question given recent experimental evidence of spillovers from both job training programs (Crépon et al., 2013) and cash transfer programs (Angelucci and De Giorgi, 2009).

Results of this study will contribute to three literatures.

First, this study conducts a type of cost-effectiveness benchmarking increasingly called for in recent years: the comparison of a standard and widespread development intervention with the outcome that would occur if the cost of the 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). Just as index funds have helped to provide a reference rate of return against which fee-charging financial managers can be compared, cash transfers of equal cost to the implementer provide an important counterfactual, and establish a hurdle rate that places the burden of proof on complex, overhead-heavy development programs to show that they can justify their costs by generating benefits superior to what would have occurred if the expense of the program was simply disbursed directly to beneficiaries. The appeal of cash transfer programs as a benchmark lies in their simplicity and scaleability, their low overhead costs, and the extent to which they put aid beneficiaries in control of how resources are allocated. This study is the second benchmarking activity being run by this team in Rwanda, the first being the benchmarking of the Gikuriro child malnutrition intervention run by Catholic Relief Services (McIntosh and Zeitlin, 2019). The current study is unique in a) benchmarking against a labor market program, b) featuring a program roughly three times as expensive overall as Gikuriro, and c) testing for complementarities between cash and the in-kind intervention.

Second, we contribute to the evidence base on supply-side active labor market programs. Though the program we study, Huguka Dukore, represents a bundle of services, our study will shed light on its potential complementarities with cash transfers in a way that informs the design of future programming in this space. While the argument for complementarities may be less obvious in a standard job training context, where the objective of labor market programs is self-employed entrepreneurship the obstacles presented by a lack of capital are manifest. Because our study will provide estimates of the marginal returns to capital with and without training programs we hope to provide policy-relevant guidance as to the optimal mix of these two approaches.

And third, our results will speak to the design of cash transfer programs in this population. Variation in cash transfer values helps to establish the tradeoff between extensive and intensive margins in the design of such programs for larger populations. And analysis of heterogeneity in returns can inform the targeting of cash versus kind programs. To the extent that strong heterogeneity is present over dimensions such as gender or the strength of local labor markets, our results will be useful in more finely tailoring the blend of approaches that is most effective for any given sub-population.

In the remaining sections of this document, we provide details of the experimental design (Section 2) and planned analyses (Section 3). Section 4 describes the statistical power of the study, and in Section 5 we outline the conclusions that we anticipate may be drawn from the results.

2 Experimental Design

Here we outline the details of the experiment. Key features include the interventions studied; the enrollment criteria; the public lottery used for assignment to treatments; the data collected at baseline and follow-up; our approach to addressing any attrition at follow-up; and the outcomes studied. These are detailed in the corresponding sections below.

2.1 Interventions studied

Huguka Dukore: Employment and entrepreneurship readiness training

Huguka Dukore is a five-year activity that will provide 40,000 vulnerable youth with increased opportunities for wage and self-employment through a suite of interventions that, among other things, improve workforce readiness through education, training, and on-the-job training or internship experiences. This activity builds on lessons learned from USAIDs prior work in this area through Akazi Kanoze Youth Livelihoods Project (henceforth AK) implemented by EDC.

Over the life of the project, HD will prepare 21,000 new youth for employment with Rwandan employers, with an additional 2,000 alumni receiving middle management training. It will assist 13,000 new HD participants to start their own microenterprise, while supporting 4,000 youth (2,000 new and 2,000 AK alumni) with an existing microenterprise to grow their business, linking 15,000 youth to financial services. Finally, HD will provide support to 30 local Implementing Partners (IPs) to improve their job placement rates.

The evaluation described here will focus on the cost effectiveness of employment readiness training, entrepreneurship and market access support for self-employment enterprises, and links to financial services such as savings and loan associations among participants in the second year of HD implementation. This study sample does not include AK alumni, but rather looks at impacts among a population of new beneficiaries.

The primary purpose of the Huguka Dukore Activity is to increase stable employment opportunities, including self-employment, for male and female vulnerable youth, and to improve youth training and employment systems and increase investment in skills for vulnerable youth. Complementing this approach are secondary goals to provide (1) a higher quality, more coordinated workforce development service delivery system, and; (2) Improve linkages between program participants and employment opportunities.

To do so, EDC will build off the strong network of local training providers that it has brought together under the AK project, the lasting relationships developed with multiple Ministries and public agencies in Rwanda at the national and district levels, a network of over 130 private sector champions, and the ongoing AK2-funded project with the MasterCard Foundation to create integrated public-private partnerships for skills development and employment linkages for youth. EDC will leverage and strengthen this network by progressively positioning the Akazi Kanoze Access (AKA) organization to serve as the principal steward of relationships and information concerning youth labor market supply and demand as well as the technical focal point for building the capacity of local IPs to link youth to employment and self-employment opportunities.

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!', consisting 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 Enterpreneurship). This 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 sector of work in which they receive additional training, and the curriculum splits according to the nature of formal employment opportunities in local markets. In more urban areas students would then move on to a Technical and Vocational Training (TVET) module, Transition to Work programming, and Work Based Learning Services. Because our study areas are almost exclusively rural, HD instead encourages students to focus on self-employment, meaning that 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 a specific cohort of students, and lasts another 10 weeks. After this point HD students are typically placed in an internship or apprenticeship position with a local entrepreneur working in the selected sector. 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.

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.

Because the curriculum involves several components of choice (whether to pursue vocational or small business training, the sector in which to be trained), our experimental analysis will treat HD as a single intervention of which this choice is an integral component. Exploratory observational work, not described in this PAP, can be conducted to understand which types of students choose which sectors and how this correlates with outcomes.

GiveDirectly: Household grants program

To benchmark the impact of the HD program to cash, we worked with GiveDirectly, a US-based 501(c)3 Non-Profit 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 in fact be capable of providing transfers to the tens of thousands of households that are served by the HD program.

GiveDirectly staff were on hand for the lottery that assigned treatment (as described in Section 2.3 below), then summoned all of those assigned to the GD arm to a community meeting. At this meeting the GD process was explained to recipients, and they were enrolled in the system GD uses to verify eligibility and make transfers. Since eligibility did not condition on having a cellphone, during this 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 then made to beneficiaries in two installments two months apart, with the first payment comprising 40 percent 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.

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 be disclosed until the GD Treatment step below. GD Treatment (where transfer values will be 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.

Combined arm

The combined arm was notified after the lottery, not in public, that they were to receive both interventions. We did this out of concern that frustration over not having received the cash might otherwise drive immediate attrition from HD among those who were not in the combined arm. This arm received both treatments 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.

2.2 Enrollment criteria

The study recruits youth from 13 geographic 'sectors' in the districts of Rwamagana, Muhanga and Nyamagabe.[3] Study participants must be eligible for Huguka Dukore, must attend an informational session about Huguka Dukore, must enroll in a lottery to determine participation in that program following that informational setting, and must be traceable to a residence in a village in the Sector where they were recruited. Attendance in person at the public lottery is not required for program

[3] 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.

enrollment. The study enrolled in its sample all individuals who met criteria for treatment by Huguka Dukore in the study sectors.

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

- Under twelve years of basic education but at least six (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)
- Where possible EDC IPs are encouraging gender parity as well as consider youth with disabilities.
- Where youth have benefited from a similar NGO intervention they will not be considered.

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 introduces 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 will be 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 produces the final list of potential beneficiaries that is passed to local implementing partners (IPs).

4. **Invitation of potential beneficiaries to an orientation meeting**. The language of this invitation changes slightly relative to what is presumed typical of HD outside of the study, because potential beneficiaries are not guaranteed places in HD, and may be randomly allocated to the offer of another program. 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 HDs 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.

Table 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

Table 1 shows the process by which we moved from the original oversubscription universe to the final sample of 1848 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; who agreed at that information session to be included in the assignment lottery; who 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 2.6 below) of the study participants are detailed in Table 4. Consistent with Huguka Dukore's 'soft' targeting criteria, the sample is 59 percent 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 Huguka Dukore seeks to bolster employment opportunities for underemployed youth, it does not employ a hard criterion regarding employment for eligibility. Consequently, it is not unusual for individuals to report that they are employed: 33 percent of (control-group) respondents report that they are employed, using a definition that *excludes* agricultural work on a farm belonging to their own household (see Section 2.6 for more details).

Nonetheless, individuals in the study population are quite poor. 32 percent 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

The allocation of these study households to treatment was undertaken on a randomized basis across eligible, interested individuals using a public lottery. A public lottery was selected as the assignment mechanism given the very large sums of money being transferred and the desire by all parties to the research to ensure that the assignment was considered to be fair and impartial by the research subjects.

Lotteries were conducted at the sector level in each of the 13 sectors in the study, and the proportions assigned to each treatment were fixed at each lottery. This results in a fairly standard 'blocked' randomization structure across the 13 blocks in the study. 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.

The detailed protocol for the lottery is as follows:

1. Beneficiaries did not have to be physically present at the lottery to be included in the study. We explicitly recognized the right of EDC/HD to eliminate from eligibility any individuals who they feel, for whatever reason, was not serious about the program and that they did not believe will fully enroll in HD if selected.

2. Detailed information about GD was not provided prior to the lottery, but GD was be described in detail at the lottery and every effort was made to preserve the separate identity of HD and GD so as not to provoke confusion about the broader HD program. All information provided at the lottery was be given to everyone, and there was not an attempt to separate groups and give private information.

3. A representative of both GD and HD (or its local partner) were present at every lottery.

4. Individuals were notified whether they have been assigned to the GD, the combined arm or the HD arm at the time of the lottery.

5. Individuals assigned to GD received a variety of colors which correspond to different transfer amounts. This means that the random assignment to GiveDirectly simultaneously randomly assigned individuals to the different transfer size amounts. GD would not reveal financial amounts at the time of the lottery, but would reveal timing of transfers to all beneficiaries, and would describe their enrollment process. GD will explain that youth randomized to GD will be contacted soon after the lottery to orient them to the program, and that they will be visited at their place of residence to undertake the enrollment process by [month].

Table 2 shows the outcome of the lottery process, giving the number of individuals assigned to each of the treatment arms within each lottery, as well as overall.

The assignment of individuals to the main study arms was as follows:

1. HD beneficiaries (485 individuals);

2. Recipients of unconditional household grants (672 individuals);

3. Combined arm who received both HD and the household grants intervention (203 individuals)

4. A comparison group, in which no program was offered (488 individuals).

Household grants are randomized at the individual level over four transfer amounts. The value of the first transfer amount was made equivalent to the total cost of providing HD to each beneficiary, which is $452.47. Less GD's own associated costs of delivery, this means that an amount of $410.19 will actually be transferred to households in this arm to make them cost-equivalent to USAID. Because we do not know the true per-capita cost of HD with certainty beforehand, we randomize GD transfer amounts to two additional values that bracket this expected cost. The bracketing amounts are derived by supposing that the number of beneficiaries for the year two tranche of HD funding nationwide may vary between 8,000 and 12,000 beneficiaries, meaning that the per-capita cost will vary between $377.05 and $565.58. Again netting out GD's costs of making transfers, that means that households in these arms actually receive $317.34 and $503.04, respectively (note that because we costed each GD transfer amount separately and because many of GD's costs are fixed at the individual level, the fraction of total cost that is overhead declines as the transfer amount increases). The fourth transfer amount is designed to examine impact at the transfer amount that GD feels will maximize the benefit-cost ratio of household grants, and transfers $750 to beneficiaries. This variation is intended to shed light on optimal transfer size for purposes of the health and socio-economic outcomes that are the main objectives of this study.

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.

Table 2: Study Design

			GiveDirectly				Combined
Sector	Control	Huguka Dukore	317.34	410.19	503.04	750	HD + 410.19
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

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. To reflect this issue, for analysis of household outcomes we will analyze the fraction of individuals in the household treated, and for individual outcomes we will cluster standard errors at the household level. We will also report as a robustness check a replication of the comparative impact analysis for primary outcomes (equation (1)) that omits households from the first-phase lotteries that contain multiple, conflicting treatment assignments.

Given the public nature of the lottery assignment, the study was not blinded either to participants or to the survey firm. The study is not a pipeline design, and 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 Survey data collection and processing

2.4.1 Instruments

Given that the beneficiary population will in many cases be embedded within a household, and that consumption and assets may be pooled among or transferred between household members, the study will use two distinct instruments within each round of data collection. A household survey will be administered to the household head, and a beneficiary survey will be administered to the beneficiary. For beneficiaries who live on their own or who head their own household, these instruments will coincide.

We provide an overview of the contents of each instrument in Table 3. Construction of primary outcomes and hypothesized effect moderators are detailed in Section 2.6 below.

2.4.2 Tracking and follow-up

All households will be followed up at endline, roughly 18 months after baseline. This is a study fundamentally designed around an eligible individual (not household). Therefore all of the survey protocols and tracking protocols will be engaged in understanding the individual circumstances of this person, as well as the household in which he/she resides. Most of the modules of the study survey will be answered by the core respondent, with the exception of certain household-level outcomes such as consumption and spending, which may be answered by the household head.

The interventions studied in this trial have the possibility of inducing geographic movement of respondents. For this reason, it is particularly important to have a strategy to address attrition in place. Our core protocol is to track all individuals who continue to reside either in the district in which they were at baseline, another study district, or in Kigali. In addition to this, we will randomly sample (as budget allows) a subset of any remaining attriters, and intensively track them irrespective of their new place of residence (including to Uganda or elsewhere). The aim of this exercise is to provide complete coverage of a representative sample of those who could not be found

Table 3: 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

by standard tracking protocols. As described in Section 2.5, we will up-weight these individuals in our analyses to ensure that estimated impacts reported are representative of the study sample as a whole.

In addition to this endline, we also intend to return to the field 36 months after baseline to conduct a longer-term follow-up survey, providing an eventual window into longer-term impacts. Data collected then will be used to, among other things, revisit impacts on the primary outcomes set out in this Pre-Analysis Plan.

2.4.3 Data Management

Data are collected by IPA digitally using tablets and the ODK platform and all data is maintained in a password-protected environment. Survey responses will be coded using a unique alphanumeric identifier so that no personal identifiers are entered into the database. Raw data is stored in password-protected files on a password-protected computer. Any records linking names to unique identifiers are stored in a locked office in Kigali, Rwanda and will be destroyed upon completion of data collection.

All research data will be transmitted in encrypted form to a secure server. Data will be de-identified in analytical datasets. No personally identifiable information will be shared with those outside the research team, and results will be presented publicly only in aggregate form (summary statistics, regression results, etc.)

2.5 Variations from the intended sample size

Our core tracking protocol will attempt to follow up on all study participants from the baseline sample. The beneficiary survey is administered to the individual beneficiary, and the household survey instrument is administered to the head of the household in which the beneficiary is resident at the time of the survey. The tracking protocol will attempt to locate every individual who continues to reside in the baseline districts, or who has move to Kigali. At the end of the standard survey exercise we will then randomly sample a subset of the attriters and enroll them in the 'intensive tracking' exercise. We will work very hard to push the followup rate within the intensive tracking sample to 100 percent, and then we we will up-weight the individuals from the intensive tracking exercise by the share of attriters assigned to intensive tracking. All core analysis will include these intensive tracking weights. The size of the intensive-tracking sample will be determined based on budget and attrition rates. To the extent that tracking rates remain less than 100 percent, even in this intensive tracking arm, we will estimate Lee bounds (Lee, 2002) as a robustness check.

2.6 Outcomes studied

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.[4]

Descriptive statistics and tests of balance across the primary treatment arms are provided in Table 4. This information is provided for all primary and secondary outcomes. We note that the study appears to be well balanced; for all primary outcomes we fail to reject the null of equal means across the full set of study arms. Small imbalances in baseline values for some outcomes (specifically, household consumption) will be addressed by conditioning on baseline values and through the selection of covariates described in Section 3.

2.6.1 Primary outcomes

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 a wage job or as primary operator of a microenterprise. 1 week recall, 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`)

[4] 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.

Table 4: Descriptive statistics and balance

	Control mean (SD)	HD vs Control (SE)	GD vs Control (SE)	Combined vs Control (SE)	GD vs HD (SE)	Combined vs HD (SE)	Combined vs GD (SE)	p-value
Ubudehe category I	0.32 (0.47)	0.01 (0.01)	0.02 (0.02)	-0.03 (-0.03)	0.00 (0.03)	-0.04 (0.04)	0.05 (0.04)	0.58
Beneficiary female	0.59 (0.49)	0.03 (0.03)	0.02 (0.02)	-0.03 (-0.03)	-0.01 (0.03)	-0.06 (0.04)	0.05 (0.04)	0.43
Beneficiary age	23.53 (3.56)	-0.10 (-0.10)	-0.24 (-0.24)	-0.42 (-0.42)	-0.15 (0.20)	-0.32 (0.29)	0.18 (0.28)	0.44
Beneficiary years of education	7.55 (2.11)	0.12 (0.12)	0.05 (0.05)	-0.21 (-0.21)	-0.06 (0.13)	-0.33* (0.17)	0.26 (0.16)	0.27
Household members	4.99 (2.25)	-0.32** (-0.32)	-0.10 (-0.10)	-0.25 (-0.25)	0.23* (0.14)	0.07 (0.18)	0.16 (0.17)	0.11
Employed	0.33 (0.47)	0.04 (0.04)	0.00 (0.00)	0.02 (0.02)	-0.04 (0.03)	-0.02 (0.04)	-0.02 (0.04)	0.47
Productive time use, hrs	22.82 (22.30)	0.55 (0.55)	1.74 (1.74)	-0.92 (-0.92)	1.19 (1.82)	-1.47 (2.23)	2.66 (2.22)	0.65
Monthly income	9.26 (2.38)	0.04 (0.04)	0.03 (0.03)	-0.12 (-0.12)	-0.02 (0.22)	-0.16 (0.31)	0.15 (0.30)	0.96
Productive assets	8.46 (3.94)	-0.27 (-0.27)	-0.72 (-0.72)	0.19 (0.19)	-0.45 (0.51)	0.46 (0.64)	-0.91 (0.59)	0.33
HH consumption per capita	9.46 (1.02)	-0.11* (-0.11)	-0.11* (-0.11)	-0.00 (-0.00)	0.00 (0.06)	0.11 (0.08)	-0.11 (0.08)	0.13
Beneficiary-specific expenditure	8.00 (1.16)	-0.03 (-0.03)	-0.03 (-0.03)	0.10 (0.10)	0.01 (0.07)	0.13 (0.09)	-0.12 (0.09)	0.51
HH net non-land wealth	9.37 (7.32)	0.16 (0.16)	0.36 (0.36)	-0.06 (-0.06)	0.19 (0.43)	-0.22 (0.62)	0.42 (0.59)	0.82
Savings	10.16 (1.40)	-0.00 (-0.00)	-0.08 (-0.08)	0.12 (0.12)	-0.08 (0.09)	0.12 (0.14)	-0.20 (0.13)	0.46
Debt	10.45 (1.57)	-0.05 (-0.05)	-0.03 (-0.03)	0.03 (0.03)	0.03 (0.10)	0.08 (0.14)	-0.05 (0.14)	0.94
HH livestock wealth	11.66 (1.50)	0.16 (0.16)	0.10 (0.10)	0.11 (0.11)	-0.06 (0.11)	-0.05 (0.15)	-0.01 (0.15)	0.62
Business Knowledge	0.01 (1.03)	-0.04 (-0.04)	-0.01 (-0.01)	0.06 (0.06)	0.02 (0.06)	0.10 (0.08)	-0.07 (0.08)	0.68

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.

- Non-agricultural wage income (B02qnoagricwage)
- Microenterprise profits (B02qenterpwage + B02qsemploywage);
- 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.

2.6.2 Secondary outcomes and mechanisms

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,[5] 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.

2.7 Multiple outcomes and hypothesis testing

To mitigate risks of false discovery across multiple outcomes and treatments, we will 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 2.6.2), ensuring that the false discovery rate at the family level is controlled at five percent.

3 Empirical Analysis

3.1 Cost Equivalence, Before and After the Fact

The costing exercise in the study utilized the 'ingredients method' which specifies all the ingredients (resources and inputs) used in performing the activities that produce the key outcomes of interest.

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

In this costing, cost is defined as opportunity cost: the value of a good or service in its best alternative use. When a good or service is used for a specific purpose, the user "gives up" the possibility of employing it in another application, (for more discussion, see Dhaliwal and Tulloch, 2012; Levin and McEwan, 2001) for more discussion.

The policy question is asked from the perspective of the donor (in this case, USAID): the policy objective is to achieve the highest benefit-cost ratio per intended beneficiary for each dollar that is spent on a program. Overhead expenditures in the implementation chain are an inherent part of these costs, and so the lower transactions costs in getting mobile money to the beneficiary play an important role in their potential attractiveness. 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 the lower GiveDirectly transfer amounts would be randomized. Then, a rigorous ex-post costing exercise was conducted for both programs after the fact, using actual budgets and expenditures.

Since the HD program covers eight 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 overhead costs entailed in managing the HD program. Beneficiary identification costs, incurred by the survey firm and identical across all arms of the study, are excluded from the cost-benefit calculation. 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 56,127 beneficiary households like HD. This is the relevant scale for a USAID program officer contemplating commissioning a program to move the outcomes studied.

We costed each GD arm separately, asking what the overhead rate would have been if GD had run a national program at the scale of HD giving only transfers of that amount. Overhead costs as a percentage of the amount transferred decline sharply with transfer amount for GD because fixed costs represent a large share of their total overhead. This 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.

The headline costs used in the comparative cost effectiveness exercise will reflect the optimized costs if both implementers were running efficient programs at scale. Given that this number does not reflect the costs of identifying and treating the specific population actually used in the experiment however, we will conduct a robustness check that uses incurred costs per eligible individual in the study. This number will be derived by a bottom-up exercise measuring what was actually spent to register, lottery, and treat the true study sample. Given differences in scale of implementation, we use these costs—and an analysis of their fixed and variable components—to provide cost estimates based on these ingredients, for a scenario in which the full district eligible population (here, equivalent to the study sample) were treated by each partner, operating alone.[6] Because this number is sensitive to modeling assumptions, we consider it to be a robustness check that can corroborate the informativeness of HD's costs at eight-district scale for the three districts covered in the study.

3.2 Overall Comparative Impact Analysis

The data from the study are analyzed consistent with the design being a three-armed, individually 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, fixed effects for the sector-level assignment blocks within which the randomization was conducted μ_b, as well as a set of baseline control variables selected from the baseline data on the basis of their ability to predict the primary outcomes, denoted by X_{ihb0}. 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}:

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

Block-level fixed effects,μ_b, are be included to account for the block-randomization of the study. Standard errors will be clustered at the household level because the second tranche of treatment was assigned at the household level. Following the 'post-double-LASSO' procedure of Belloni et al. (2014b), a set of covariates will be selected using a LASSO algorithm on the control data; further details of this procedure are provided in Appendix B. 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).

The core question for the combined arm is whether there is a complementarity between the receipt of HD and GD treatment that leads the impact of receiving both to be larger than the sum of the impacts of receiving each. Given that the value of the cash transfer in the combined arm is equal to the value of the cash transfer in the intermediate GD-small arm ($GDS2$), a test for complementarities is dervied from equation (1) above by an F test of the null hypothesis that $\delta^{HD} + \delta^{GDS2} = \delta^{COMB}$. In addition, while not a strict cost-equivalent benchmarking question, equation (1) can be used to answer the question of which intervention achieves the greatest average benefit in a fixed population at a fixed budget, by dividing each arm's estimated benefits by its costs. We will present the F-test as to whether the impact differential between HD and GD Large is different than the cost differential.

[6]Note that in the presence of fixed operating costs and different numbers of treated beneficiaries by each implementer within the study, unadjusted cost-per-beneficiary numbers from the study sample are not strictly comparable.

We pre-specify 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 eligible within each transfer amount arm, denoted by t_c. Subtract from this number the benchmarked HD cost per household C described above, 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. Because τ absorbs the deviation of the GD arm from the benchmarked HD cost, the dummy coefficient on HD treatment will serve as an intercept measuring the impact of HD benchmarked an exactly donor cost-equivalent cash transfer. So, we have:

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

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 amount effects, the slope coefficient τ_c captures impacts arising from deviations in GD cost from HD cost, the coefficient δ effectively 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.

3.3 Differential Compliance

We anticipate a substantial difference in compliance rates across the two interventions. Preliminary estimates suggest that GD compliance within the study sample will be approximately 98%, while compliance with HD will be closer to 75%. This suggests that the simplest comparative analysis of the two arms, which is the Intention to Treat (comparing those *assigned* to the HD versus GD arms, ignoring whether they actually took the program) may be misleading.

In terms of costing, we divide costs into two types; *averted* costs are those that are not incurred by EDC if the individual does not comply with the treatment, while *un-averted* costs are those that EDC incurs whether or not the individual complies with treatment. EDC pays its sub-implementers based on enrollment numbers as of the third meeting, meaning that all non-compliance that occurs prior to this (expected to be 15%) allows EDC to avoid the costs of treating these individuals, while the subsequent non-compliance (expected to be 10%) imposes un-avoided costs. We therefore measure costs relative to the study sample by considering attrition prior to week three as averting costs to EDC, while subsequent attrition does not decrease their costs. We will calculate the cost per beneficiary using the overall national program data from the HD program, not the cost per study beneficiary which may have been distorted by the experiment. Un-averted costs are spread across the whole targeted population while averted costs are first multiplied times the compliance rate for each treatment.

The cost-equivalent analysis is the core way of adjusting for differential non-compliance, because to the extent that costs are *avoided* (see above), then the comparison of benefit/cost ratios conducted in the cost-equivalent exercise automatically adjusts for non-compliance.

If we want to recover the 'Local Average Treatment Effect' of HD (the impact of actually receiving treatment), we can pursue the simple course of dividing the ITT in an arm by the compliance rate in that arm. It is important to recognize that this approach assumes, first, that there are no spillover effects of the interventions on each other or on non-compliers within an arm, so if the spillover analysis discussed below shows significant evidence of contamination we will not attempt to estimate a LATE in this study. Moreover, this approach assumes that there is no benefit of attending a number of HD classes fewer than the averted-cost threshold, but LATE assumptions would be violated if students are impacted by attendance at the first two classes.[7]

3.4 Analysis of Heterogeneity

We will use a standard interaction analysis (between treatment indicators and baseline characteristics) to study heterogeneity across the outcomes listed in the Moderators section above, namely:

1. Beneficiary gender. (`B1q3`).

2. Household baseline consumption per capita. (Defined as baseline analogue of primary midline outcome.)

3. Risk aversion (Baseline question `B13q_beg_lottery1`). We will define an indicator of individuals who are more risk averse as those choosing lotteries A, B, or C in this Binswanger-Eckel-Grossman lottery over payouts the following day.

4. Local labor market conditions, proxied for by the fraction of individuals in that sector at baseline who were employed.

In addition, we will examine whether the number of treated individuals in the beneficiary's social network affect primary outcomes (spillovers). This analysis is substantially more complicated and so is described in detail in the next section.

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) finding potential *positive* spillovers, while a more recent

[7]Such impacts could include both the direct effects of training, and the effects of networks formed.

controvery over the long-run impacts of GiveDirectly's programs in Kenya and Rwanda suggesting that the price increases caused by the surge in consumption may lead to harm to non-beneficiaries (Haushofer and Shapiro, 2016; McIntosh and Zeitlin, 2019). Because our study uses an individually randomized design these spillovers are also a direct threat to internal validity, and so we provide substantial detail here on how we plan to test for the existence of, and correct for the potential presence of, spillover effects.

The study will look for spillovers on both 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. 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, and provides no variation in treatment saturation at that or higher levels.

The goal of this spillover analysis is to be able to extrapolate from estimands identified with minimal assumptions within the study to those of broader interest, as outlined below. As such, evidence of the the presence of spillovers will not change our primary tests for the presence of differences across arms, though their interpretation is affected.

3.5.1 Outcomes for spillover analysis

We will test for the presence of spillovers on primary outcomes and on program take-up. We have two distinct purposes in estimating spillover effects; the first is the standard one of assessing whether outcomes in our control group are contaminated. The second, given the emphasis on costing in this study, is to recognize that if compliance rates are also driven local treatment intensity then both benefits and costs must be adjusted to account for this.

For this latter purpose we define take-up of HD in the manner relevant for understanding costs. In particular, we focus on spillovers onto registration status as of Class 3, since the variable costs of HD are not considered to be averted for those who drop out after this point.

3.5.2 Definition of local network and local saturation rates

We estimate a model that allows for local spillovers, which we define as spillovers between individuals who reside in the same village (cell) at baseline.

We use the following notation to denote this. Consider a graph G whose entry (i, j) is an indicator for a connection between individuals i and j, and define $G_{i,i} = 0$. Further letting $d_i = \sum_j G_{ij}$ denote the 'degree' of individual i—the number of study participants in her village. Let T^w_{ivb} denote the assignment of individual i in *village* v to treatment $w \in \{\text{GDS}, \text{GDL}, \text{HD}\}$, where we pool the three smaller cash-transfer values into a single arm, $w = \text{GDS}$, as distinct from the larger transfer value, $w = \text{GDL}$. Individuals in the combined arm have $T^{GDS}_{ivb} = T^{HD}_{ivb} = 1$. Define $T_{ivb} = [T^{GDS}_{ivb}, T^{GDL}_{ivb}, T^{HD}_{ivb}]$ as the vector denoting individual i's treatment status. Finally, we let $\bar{T}_{-i,vb}$ denote the average treatment status of study individuals *other than* individual i in village v, and we adopt the convention that $\bar{T}_{-i,vb} = 0$ if individual i is the only study participant in village v.

3.5.3 Estimating spillover effects

For outcome Y_{ivb1}, we modify the specification of equation used to estimate ITT effects (equation 1) to allow the treatment status of individuals who are socially connected to the respondent to have direct effects on the respondent, and to modify the impacts of treatment on that respondent. We will estimate the following linear-in-means model:

$$Y_{ivb1} = \delta_1 T_{ivb} + \delta_2 \bar{T}_{-i,vb} + \delta_3 T_{ivb} \bar{T}_{-i,vb} + \beta X_{ivb0} + \rho Y_{ivb0} + \mu_b + \varepsilon_{ivb1}. \tag{3}$$

In equation (3), the null hypothesis that $\delta_2 = 0$ corresponds to an absence of direct spillovers from connected individuals; the null hypothesis that $\delta_3 = 0$ implies that treatment assignments of village co-residents do not modify the impact of a given treatment. Note this latter null might be violated, such that $\delta_3 \neq 0$, if neighbors' treatment assignments affect program take-up rates, but it will also be violated by other forms of interference.

3.5.4 Testing for spillovers on specific outcomes

To conduct inference about these coefficients, we note that our assignment protocol generates variation in village-level exposures, $\bar{T}_{-i,vb}$, and that our model of spillovers assumes an absence of spillovers across villages. Moreover, each of the coefficients δ_1, δ_2, and δ_3 represent substantively different hypotheses about the nature of direct effects and interference across the treatments. We therefore 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 (16 at a minimum) we will also correct the p-values in these regressions using Anderson's (2008) False Discovery Rate correction across all coefficients within each regression.

3.5.5 Spillover effects on costs

In analyzing GD or HD takeup, we will do so using the set of individuals assigned to GD only—pooling small and large treatments—or HD only—omitting the combined arm—in separate regressions. Consequently, we omit the interaction between T_{ivb} and the treatment saturation rate in individual i's neighborhood and estimate the linear probability model

$$\mathrm{E}[P^w_{ivb} | \bar{T}_{-i,vb}] = \mu^w_b + \phi^w \bar{T}_{-i,vb} \tag{4}$$

where P^W_{ivb} is a measure of individual i's participation in treatment $w \in \{GDS, GDL, HD, Combined\}$ to which they have been assigned.

3.5.6 Calibrating cost-effectiveness for uniform treatment regimes

If interference is present, then policymakers may be interested in a counterfactual question that requires extrapolation beyond our experiment: what would the cost effectiveness of each intervention have been if *all* eligible individuals had been assigned to that treatment?

In this thought experiment, the estimand of interest is that which compares outcomes in a completely untreated network with outcomes in a network in which all applicants are treated with the same treatment. Answering this question requires adjustments to both costs and benefits.

For example, estimates of HD participation from equation (4) allow us to estimate the take-up rate of HD under a scenario in which all eligible applicants in a local network are assigned to HD. This is estimated from those regression coefficients as $\mathrm{E}[P^{HD}|HD] = \mu_b^{HD} + \phi^{HD}$. Following the terminology of Section 3.3, average costs per beneficiary of treatment T are then the avoidable costs multiplied by the predicted participation rate at full saturation, $\mu_b^T + \phi^T$, plus the unavoidable costs. Should spillover effects of HD on GD be found, this issue will be handled in a symmetric way.

Equation (3) provides a basis for extrapolation to each program's impacts under uniform assignment. For a given treatment, w, this corresponds to $\delta_1(w) + \delta_2(w) + \delta_3(w)$: the treatment effect in this counterfactual exercises comprises a direct effect of treatment $\delta_1(w)$, an externality that applies irrespective of treatment status $\delta_2(w)$, and an externality that operates by modifying the effect of the beneficiary's own treatment $\delta_3(w)$.

4 Statistical power

We present power for analyses of pairwise comparisons between any two of the major arms of the study: Control, Huguka Dukore, or the pooled GiveDirectly 'small' arm. The power we present applies either to the comparison of the HD arm to the control, or the F-test on the pooled effect of the three smaller against the control from the main estimating equation. These planned comparisons are ITT analyses, with randomization-block fixed effects as well as baseline values of the outcome and lasso-selected covariates included to reduce residual variation in the corresponding outcome. Following presentation of these MDEs, we discuss magnitudes of detectable effects in relation to both policy-relevant treatment magnitudes and the range of treatment effects found in the relevant literatures on training and cash-transfer programs in developing-country settings.

As reported in Table 2, we note that each of these arms consists of approximately 485 beneficiaries, individually randomized. We therefore compute minimum detectable effect sizes for a two-sided test as

$$MDES = \frac{t_{\alpha/2} + t_{1-\kappa}}{\sqrt{P(1-P)N}} \left(1 - R^2\right), \tag{5}$$

with size $\alpha = 0.05$, power $\kappa = 0.8$, the proportion in any given arm as $P = 0.5$, and the effective number of observations as $N = 970$. Here, R^2 corresponds to a regression of the outcome on the full set of controls, including block fixed effects and covariates. Because this value is unknown a priori, we consider the MDES that corresponds for a range of potential values of 0, 0.1, 0.2, and 0.3. This yields minimum detectable effects of 0.18, 0.16, 0.14, and 0.13 standard deviations, respectively.

In Table 5, we map these effect sizes into their corresponding economic magnitudes for primary outcomes, by using the baseline standard deviations previously reported in Table 4. (These are reproduced for convenience in Table 5 as well.) We note that compliance with assignment to the cash-transfer arms of the study is universal, but that take-up of the HD program is 85 percent even among this population, so for comparisons involving that arm, the MDE should be scaled up by dividing it by this compliance rate. Given the financial values of the programs evaluated here, and the relative baseline poverty of the population under study, we believe that this individually randomized trial leaves us adequately powered to detect clinically relevant effect sizes—that is, effects large enough to justify either the expense (relative to control) or a change in programmatic approach (for cash-vs-kind comparisons), as we discuss below.

Table 5: Minimum detectable effects for pairwise comparison of primary outcomes across arms

	Control mean	Minimum detectable effect			
	(SD)	$R^2 = 0$	$R^2 = 0.1$	$R^2 = 0.2$	$R^2 = 0.3$
Employed	0.33 (0.47)	0.08	0.08	0.07	0.06
Productive time use, hrs	22.82 (22.30)	4.02	3.61	3.21	2.81
Monthly income	9.26 (2.38)	0.43	0.39	0.34	0.30
Productive assets	8.46 (3.94)	0.71	0.64	0.57	0.50
HH consumption per capita	9.46 (1.02)	0.18	0.17	0.15	0.13

Note: Table reports minimum detectable effects for primary outcomes, as defined in Section 2.6. Employment is an indicator for non-farm employment. Productive time use is in hours. Monthly income, productive assets, and household consumption per capita are transformed using the inverse hyperbolic sine. MDEs reported are for corresponding values of R^2, where this is defined as the share of the endline outcome variance explained by included baseline covariates.

How large are the effects detectable by this study?

One way to gauge effect sizes is relative to those found elsewhere in prior studies. As a starting point, a randomized evaluation of Akaze Kanoze, the predecessor program of Huguka Dukore, working in a similar context and with a similar target population, estimated a positive impact of 13 percentage points on the probability of being employed (Alcid, 2014).[8] This finding was used to justify the decision to fund the Huguka Dukore program by USAID. In the scenario in which controls for strata, baseline levels of outcomes, and lasso-selected covariates (as discussed in Appendix B) explain 20 percent of the variation in the employment outcome, we would be powered to detect an impact approximately half that size with 80 percent probability.

That said, EDC's estimates of the positive employment impacts of the Akaze Kanoze program make it something of an outlier in relation to the literature on vocational training programs. As reviewed in McKenzie (2017), randomized evaluations of vocational training programs have typically found only very small effects on employment, on the order of two percentage points, with modest impacts on earnings, with a mean estimated increase of 17 percent over control-group outcomes. It should be noted that McKenzie (2017) reports ITT esitmates, for which he suggests typical compliance rates imply LATE estimates should be scaled by 1.2 to 1.4; thus, the literature he reviews suggests earnings impacts between 20 and 24 percentage points for those induced to take up vocational training programs. More recent experimental evidence from neighboring Uganda shows much larger impacts of vocational training—their ATT estimates suggest a gain in the probability of employment of 13.5 percentage points, an impact of 7.12 on the number of hours worked, and an impact of 42 percent on earnings, on average across that study's follow-up period (Alfonsi et al., 2019); that study suggests that the formal certifiability of skills created in the vocational training program—a feature of the Huguka Dukore program as well—is a potentially important feature in its success. Impacts of cash-transfer programs on time use, earnings, and productive assets have been larger in the literature, suggesting that our study is relatively well powered to detect comparative impacts of this magnitude.

An alternative perspective altogether on statistical power focuses on power to detect rates of return sufficiently large to justify interventions. Analogous to 'clinical relevance' as a standard for judging effect sizes in medical trials, this perspective holds that it is not worth powering studies to detect impacts, e.g., on employment of two percentage points, because impacts that small would not be sufficient to justify the costs of intervention. So how big would impacts need to be in order to justify intervention costs in this setting?

On the cost side of the ledger, Huguka Dukore's ex ante benchmarked costs are modest relative to vocational training programs noted in the literature. For example, Blattman and Ralston (2015) cite typical program costs as in the range of USD 1,000 to 2,000 per person in developing-country settings, and exclusive of "many fixed and administrative costs". Our ex-ante estimate of HD Program costs, at approximately $450, puts this on the low end of costs for vocational training programs that have been the subject of randomized evaluations; program costs are lower in only three of the nine studies reviewed by McKenzie (2017).

In spite of these modest costs, given the low levels of beneficiary incomes at baseline, the study is well powered to detect impacts large enough to provide a reasonable rate of return on the

[8]Note that relatively high survey attrition rates in this study—although balanced by treatment arms—permit a range of possible treatment effects.

intervention. Median beneficiary income at baseline is RWF 8,400 per month. For a beneficiary with median income, the conservative MDE of 0.43 implies an impact of RWF 3,612, or approximately USD 4.47, using nominal exchange rates at the outset of the intervention. Implied internal rates of return will depend on how long this differential is assumed to persist: for a program impact that persists for T months, the *minimum detectable internal rate of return* solves the monthly discount factor, $\delta = 1/(1+r)$, such that

$$Costs = MDE \frac{1-\delta^{T-1}}{1-\delta}. \tag{6}$$

The implied annualized internal rate of return is then $(1+r)^{12}-1$. The existing literature suggests fade-out over the long-term (McKenzie, 2017), but to be conservative, we consider benefits that last for periods of five or ten years. The internal rate of return for a five-year impact at this MDE is actually *negative* 16.9 percent. If we assume instead that benefits of a magnitude corresponding to this MDE persist for 10 years, then the implied rate of return is 3.8 percent. Put differently, in order to meet the threshold of a five percent internal rate of return, the programs studied would have to deliver an impact of USD 14.01 per month over five years, or USD per month 11.44 for 10 years. These are more than 3 and 2 times our MDE, respectively.

Apart from the comparative impact of the primary treatments, the cost-equivalent analysis in Equation (2) allows us to estimate the return to a marginal dollar invested in the size of cash transfers. To get an approximate sense of the power of this estimate, we consider a regression that compares the pooled effect of the two larger GiveDirectly transfer amounts against the pooled effect of the two smaller GiveDirectly transfer amounts. These have sample sizes of 329 and 343, respectively, as reported in Table 2. Taking an intermediate case for the explanatory power of our baseline covariates ($R^2 = 0.1$), such a comparison has a minimum detectable effect size of 0.19.

To get a sense of the plausibility of such a difference between large and small transfers, it may be useful to focus on the outcomes monthly income and household consumption. An MDES of 0.19 translates into an effect size of 0.49 on IHS-transformed monthly income Taking again an individual at median beneficiary income, this effect size implies a gain of USD 4.85 in exchange for an additional transfer of USD 262.76 (derived as the difference in transfer values between the average of the two larger GD arms and the average of the two smaller GD arms). An increment of at least this much is almost surely required to justify the increment in spending.

Finally, power to reject the absence of complementarities—operationalized as a test of whether $\delta^{HD} + \delta^{GDS2} = \delta^{COMB}$, based on estimates in equation (1)—will be smaller than that for the comparison of the direct effects of cash and training, given the smaller sample size allocated to the combined arm.[9] Consequently, results will be interpreted with appropriate caution, and we will be particularly careful not to pass our interpretation through a statistical significance filter: we appreciate that the extent to which estimated differences *that are statistically significant* tend

[9] A ballpark estimate of the minimum detectable complementarity, in effect size terms, can be obtained by assuming the variance-covariance matrix for the estimated coefficients is diagonal. This yields an MDES of 0.599 standard deviations. Translated into monetary values, for an individual with median income, this would imply a complementarity of USD 14.03 in the impact on mothly beneficiary income.

to overstate the underlying data-generating process is inversely related to power. Nonetheless, we maintain that this is an important question. As Blattman and Ralston (2015) write in their review:

> Despite the growing number of studies, there are almost none comparing capital transfers with and without training. Even so, programs where both capital and skills were provided suggests they can be complements.

Given the substantial fixed costs involved in shifting programming from cash-only (or for that matter, from training-only) to cash-plus-training modalities, large complementarities would be required in order to justify shifting single-purpose organizations and programs into delivering both types of programming simultaneously.

5 Study implications and potential conclusions

This study is intended to achieve a research result increasingly called for in recent years, namely a direct cost-equivalent 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. Through careful experimental design we hope to be able to present very transparent and readily interpretable results for policymakers wanting to spend aid money in the most cost effective manner to assist underemployed youth in their transition to a productive adult work life.

Cost-equivalent comparisons of impacts on specific outcomes should not be equated with welfare maximization. The use of a cost-equivalent comparison allows us to abstract from the strong stance on distributional preferences required to evaluate programs that differ in their reach, though it leaves open the possibility that, for any *given* social welfare function, the cost-effectiveness-maximizing level of intensity for a given intervention may not be the cost-equivalent one. Moreover, beneficiaries may place weights on the outcomes studied that differ from donors' implicit or explicit objects—or may value other outcomes altogether. A salient feature of cash transfers is the extent to which, in well functioning markets, they are fungible across outcomes in accordance with beneficiary preferences. Consequently, it is particularly when markets fail for reasons of inter- or externalities that the approach we undertake in this study can be at its most useful to guide governments' programmatic decisions.

6 Bibliography

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.

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.

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.

_ , _ , and _ , "When the Money Runs Out: Do Cash Transfers Have Sustained Effects on Human Capital Accumulation?," *World Bank Policy Research Working Paper*, 2016, *7901.*

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

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.

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.

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.

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

Brudevold-Newman, Andrew Peter, Maddalena Honorati, Pamela Jakiela, and Owen W Ozier, "A firm of one's own: experimental evidence on credit constraints and occupational choice," 2017.

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.

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.

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.

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.

Lee, David, "Trimming for Bounds on Treatment Effects with Missing Outcomes," NBER Technical Working Paper No. 277 June 2002.

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.

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.

McKenzie, David and Christopher Woodruff, "What are we learning from business training and entrepreneurship evaluations around the developing world?," *The World Bank Research Observer*, 2013, *29* (1), 48–82.

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, Munshi Sulaiman et al., "Cash-Plus: Cash Transfer Extensions and Recipient Agency," Technical Report, Centre for the Study of African Economies, University of Oxford 2017.

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.

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

Appendix A Supplementary tables

Table A.1: Adult-equivalence scales

Age range	Male	Female
Less than 1 year	0.41	0.41
1-3 years	0.56	0.56
4-6 years	0.76	0.76
7-9 years	0.91	0.91
10-12 years	0.97	1.08
13-15 years	0.97	1.13
16-19 years	1.02	1.05
20-39 years	1.00	1.00
40-49 years	0.95	0.95
50-59 years	0.90	0.90
60-69 years	0.80	0.80
More than 70 years	0.70	0.70

Note: Adult-equivalence scale is used in Rwanda Poverty Profile Report 2013/2014, Results of Integrated Household Living Conditions Survey (EICV), Table B2.

Appendix B 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 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.[10]

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 2.6;
- Baseline values of all secondary outcomes,
- Baseline values of all dimensions of heterogeneity pre-specified in Section 3.4.

[10]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.

- 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 3.5.2.

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.

Appendix C Timeline

The key milestones of this study are the following:

Activities	**Period**
Beneficiary Targeting & Recruitment	September 2017 – December 2017
Baseline	December 2017 – February 2018
Treatment Randomization	January 2018
HD Program Implementation	January 2018 – November 2018
GiveDirectly Payment	May 2018 – July 2018
Endline	July 2019 – August 2019
Long-term follow-up	November 2020 – February 2021

The PIs have not received endline data at the time of this submission, and will remain blind to outcomes if and as required during the review period for this submission.

Appendix D Administrative information

Appendix D.1 Funding

All research funding for this project was provided by USAID.

Appendix D.2 Institutional Board Review (ethics approval)

Details of the procedures and of the consent process were read aloud, in Kinyarwanda, to each respondent prior to each measurement activity.

In addition to acquiring ethical approval from RNEC, the research team has acquired approval from the IRB at Innovations for Poverty Action, and from Georgetown University and the University of California, San Diego

Informed consent

Participant consent for inclusion in the study can be divided into two separate components:

1. Consent for inclusion in the identification of beneficiaries. Eligible applicants to EDC's HD program were informed that, given oversubscription, there was a chance they would not receive this program, but that they might receive an alternative benefit instead. Determination of eligibility was undertaken by EDC and its HD partner organizations, and required the collection of data regarding the socio-economic status (Ubudehe) of households, and the ages and education levels of youth in the household. The collection of these details was required for enrollment in the study sample.

2. Consent for the collection of socio-economic data, via questionnaire, which included details of savings, consumption, and nutritional outcomes. Households were informed that participation in this questionnaire was not required for inclusion in any of the programs under study. Households were also informed of the opportunity to decline to answer any specific question within the questionnaire.

Appendix D.3 Declaration of interest

The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

Appendix D.4 Acknowledgements

We are grateful to DIV, Google.org, and USAID Rwanda for funding, and to USAID, EDC, GiveDirectly, and IPA for their close collaboration. We thank Leodimir Mfura, Marius Chabi, and Phillip Okull for overseeing the fieldwork, and Sarait Cardenas-Rodriguez, Aruj Shukla, and Diana Martinez for research assistance.

UNIVERSITY OF CALIFORNIA, SAN DIEGO UCSD

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO SANTA BARBARA • SANTA CRUZ

Craig McIntosh
Professor of Economics
School of Global Policy and Strategy
ctmcintosh@ucsd.edu

9500 Gilman Drive
La Jolla, California. 92093-0519
Ph: 858 822 1125
URL: www-gps.ucsd.edu

To Dean Karlan, Editor, *Journal of Development Economics*, November 19, 2019

We hereby attest that both authors of this paper (Craig McIntosh and Andrew Zeitlin) are equally responsible for all stages of the project, including experimental design, survey design, oversight of field logistics, analysis of data, and the writing of the eventual paper. We have no conflicts of interest to declare with respect to this paper.

Sincerely,

Craig McIntosh and Andrew Zeitlin