

**Pre-Analysis Plan**  
**Refugee and Host Integration through the Safety Net: Evidence from Ethiopia<sup>1</sup>**  
**Midline Survey – Individual Level Outcomes**

**Authors and institutional affiliation:**

Dennis Egger, Oxford University

Alfredo Manfredini Böhm,

Sandra V. Rozo, World Bank, Research Group

**Date of latest draft:** 3/19/2025

**Abstract**

This study investigates the integration of refugees and hosts into Ethiopia's urban safety net public works and livelihoods program, targeting 22,500 households in seven refugee camps and nearby areas. Among these, 12,000 households will be sampled in a randomized control trial to evaluate the program's impact. The Refugee and Host Integration through the Safety Net (RHISN) program offers paid public works on small-scale public goods and social infrastructure, training, coaching, and a \$600 livelihood grant to foster higher-income livelihoods and socio-economic integration. The project aims to measure the program's effect on beneficiaries' wellbeing (economic, social, and psychological), analyze the impact of mixing refugees and hosts in public works and training groups on productivity and social cohesion, and assess the economic and social spillover effects on the local economy. This initiative marks one of the first attempts in Sub-Saharan Africa to incorporate refugees and host communities into a national urban social protection program.

**This pre-analysis plan focuses on the data collected in the baseline and midline household surveys.**

**Keywords:** Forced Migration, Social Protection, Public Works

**JEL codes:** H4, J61, O1

**Study pre-registration:** AER RCT Registry Number AEARCTR-0012288.

---

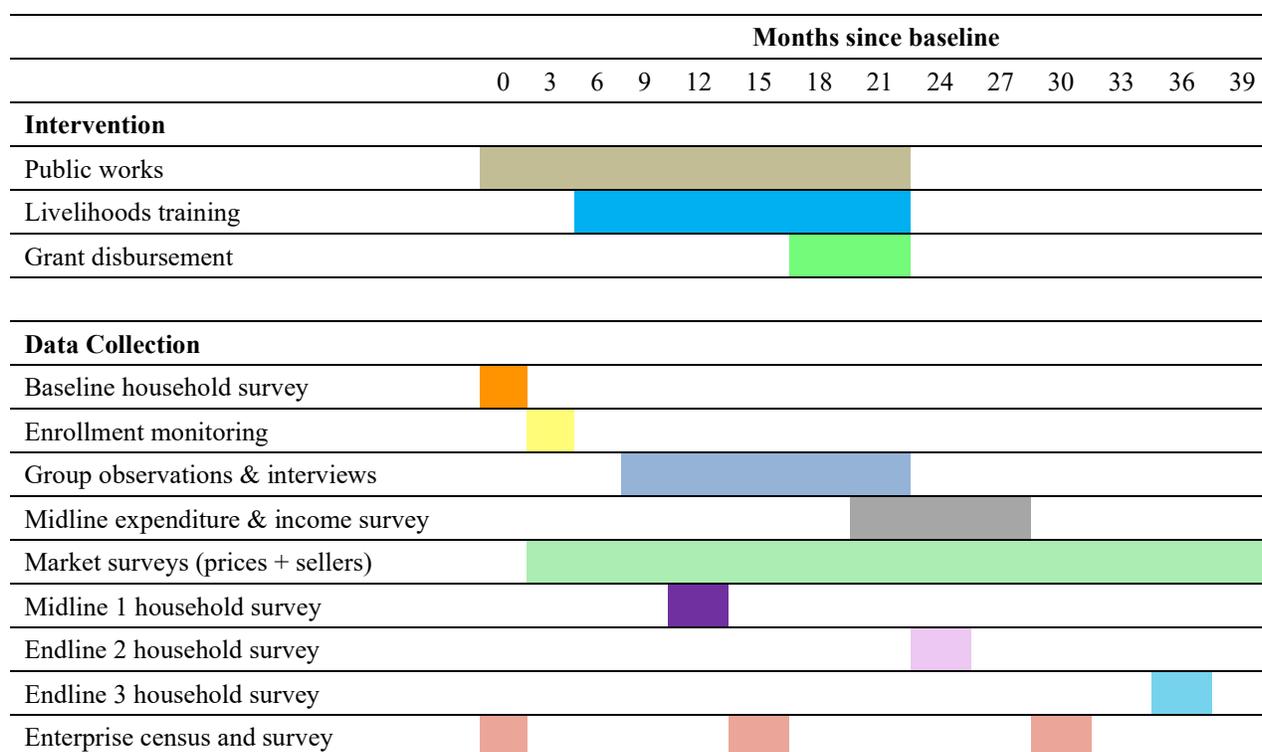
<sup>1</sup> We are grateful to Belay Mohammed Yimer, María Valentina Daza, and Thierry Hounsa for their support in preparing this document and as research assistants in this project.

## Proposed timeline and relation to other analyses.

This pre-analysis plan covers the data collected in the baseline and midline household surveys collected for the stage one of a research agenda planned for two stages:

1. **STAGE 1 - Public works and training impact evaluation:** In the first 12 months, beneficiary households will be assigned to homogenous or mixed refugee-host groups and will start participating in remunerated public works, attending initial trainings and coaching in financial management and life-skills.
2. **STAGE 2 - Livelihood development impact evaluation:** Between months 12-24th, beneficiaries continue to attend trainings and coaching sessions, focusing increasingly on developing a business plan to improve their livelihood. Based on their business plans, beneficiaries receive a USD 600 livelihood grant, disbursed between the 18<sup>th</sup> and 24<sup>th</sup> month.

**Figure 1. Timeline of the RHISN Program**



**This pre-analysis plan will evaluate only the first phase of the program and the individual level outcomes** (see Figure 1 for an illustration of the program’s timeline).<sup>2</sup> That is, it will cover effects up to 12 months after enrollment in the program, drawing on data collected from baseline and midline household surveys. While our goal is to explore the interaction and combined effects of both stages of the program, in the first stage we will only be studying public works and training independently from the large injection of cash that is the livelihood grant's disbursement. This initial stage may have very distinct, and different

<sup>2</sup> We envisage conducting longer-term follow-up data collection to study the impacts of the second phase of the program, and longer-term impact of the RHISN program on beneficiaries, non-beneficiaries, local firms, and economies. We will file additional pre-analysis plans for this future work.

effects on recipient households and the local economy. By examining the first stage separately, we can therefore shed light on the mechanisms driving the program's overall impact. Specifically, this approach allows us to isolate the effects of diversity in the workplace and trainings, distinguishing them from influences arising from the disbursement of the livelihood grants and the large spillover effects they are expected to create. Moreover, public works and training programs are fundamental to social protection strategies in many developing countries, making our findings particularly relevant for policymaking.

We envisage collecting longer-term follow-up data to study the impacts of the second stage of the program, and the longer-term impact of the RHISN program on beneficiaries, non-beneficiaries, local firms, markets, prices, and economies. We will file additional pre-analysis plans for this future work.

### **Analyses and progress to date**

As of now, no principal researcher has accessed the data. The only processing performed was by research assistants and the survey firm for quality monitoring, without any analysis or regressions related to treatment status.

## 1. INTRODUCTION

**Motivation.** Forced displacement, a critical global development issue, has increased significantly. By 2023, over 108 million individuals have been uprooted worldwide, with refugees constituting approximately a third (UNHCR, 2023). This situation is exacerbated by climate change threats and emerging conflicts. Notably, most displaced people seek refuge in low- and middle-income countries, straining regions already grappling with scarce resources and vulnerable populations. Their displacement is often protracted: over 67% of refugees reside in host countries for five years or more (UNHCR, 2023). Given this backdrop, it is imperative to explore and understand the effects of measures aimed at promoting refugee integration and fostering social harmony between refugees and host communities.

**Research Questions.** This research project will assess the impacts of the Refugee and Host Integration through the Safety Net program (RHISN), a pioneering initiative by the Ethiopian government to integrate camp-based refugees into the national social safety net. RHISN, implemented by the Ministry of Urban and Infrastructure and the Refugees and Returnees Service, is a notable effort in a low-to-middle income country (LMIC) context. In its initial year, what we call Phase 1 and focus of this pre-analysis plan, both refugees and hosts will engage in paid public works (such as street cleaning, greening, and watershed management) for up to 20 hours weekly per household. Participants will also receive residence and work permits for legal employment in the local labor market and business and livelihoods trainings. This program offers a rare chance to analyze the effects of integrating refugees into a national program on a large scale using experimental methods. Such research is scarce, particularly in humanitarian settings where the high vulnerability and mobility of displaced populations complicate randomized controlled trials (RCTs). Studies that apply randomization on a large scale are even less common, yet they are essential for understanding the challenges of implementing programs at scale, the broader economic and social integration dynamics, and for extrapolating results to other contexts.

Concretely, this research project will assess three main research questions: i) the public works program's effect on beneficiaries' well-being (economic, social, and psychological) in the short term (12 months after implementation), ii) the impacts of mixed refugee-host public works groups on productivity and social cohesion, and iii) the economic and social spillover effects on the local economy (to be further specified in future pre-analysis plans).

**Methods.** In collaboration with the Ethiopian government, we set up a two-stage clustered Randomized Control Trial (RCT) across four of Ethiopia's refugee-hosting regions. These areas exhibit varying levels of baseline social integration, with some sharing common languages, religions, and tribal affiliations, while others lack such commonalities and experience intermittent violent conflicts. Our study involves over 12,000 individuals from eligible refugee and host community households. These participants are first randomly selected to participate in the program, and then randomly assigned to either mixed or non-mixed work groups. This design creates exogenous variation in inter-group contact, fostering prolonged, intensive co-working relationships over many months. Additionally, a randomly chosen subset of these work groups will participate in additional after-work social activities, monthly teas and dinners, to complement the social interaction initiated in the workplace. Our approach also aims to measure the community-wide economic and social impacts by experimentally varying the program's overall intensity and the degree of inter-group contact at the neighborhood level. The induced variation is substantial: the total labor hours for public works will exogenously vary between 1 – 10% of total adult labor supply, and the total compensation for public

works will range from 10 – 60% of the annual GDP at the neighborhood level. Furthermore, between 5 – 40% of households in a neighborhood will be randomly chosen to participate in mixed work groups.

To trace out the flow of funds through the economies targeted by RHISN, we will conduct unusually extensive and high-frequency data collection. We will survey 12,000 households at Baseline and Phase 1 Endline, 12 months after enrollment. To understand the general equilibrium impacts on local production and prices, we will conduct censuses of all fixed-location enterprises at Baseline and Phase 1 Endline. And we will visit all 25-30 weekly markets in our study area, and a 1-to-1 matched sample of control markets, every two months to collect price data for over 100 goods and services, the total number of sellers across categories, and conduct a short survey with a random subset of 25 sellers in each market during each of these bi-monthly visits.

To evaluate workgroup dynamics and productivity, we will conduct visits to each of the 133 public work groups in Dabat – one camp in the study area where randomized mixing was implemented with high fidelity. During these visits, we will engage direct observation to assess project characteristics, worker productivity, and group interactions. Additionally, we will conduct brief surveys with a selection of beneficiaries. Our short-run analysis will focus on social cohesion, attitudes towards outgroups, policy preferences, social networks, female empowerment, aspirations, and mental well-being. These will be measured through a mix of direct survey questions, observations of interactions and collaboration during public works, and participation in social cohesion activities. By integrating these findings with household surveys, we aim to construct a comprehensive view of how diversity in the workplace influences productivity and how inter-group relationships evolve in response to artificially introduced workplace contact and socializing activities outside of work.

***Contributions relative to existing work.*** This project makes several contributions to existing literature.

1. *Evidence on the impacts of humanitarian interventions.* This study examines a large-scale public works program in a humanitarian context. Although this topic has received increasing attention from economists (Verme and Schuettler 2021; Becker and Ferrara 2019), evidence is only starting to emerge on the effectiveness of interventions that help forced migrants become self-sufficient and maintain social cohesion with hosts. Most of the available evidence focuses on humanitarian-based interventions such as cash transfers (Özler et al. 2021; Altındağ and O’Connell 2023; Hızıroğlu et al., 2022) and few studies assess development-based programs and are mostly focused on examining the impacts of regularization programs (Ibáñez et. al, 2022, Urbina et al., 2023). A such RHISN is an innovative program that will allow us to understand the effects of integrating refugees into an existing national program and on varying exogenously their interactions with hosts.
2. *Evidence on impacts of public work programs in urban areas.* Our project is one of the few studying social protection programs operational in urban settings. Particularly, in Low- and Middle-Income Countries (LMIC), adapting such initiatives to urban environments is a key policy challenge, and there is limited evidence on how to do so effectively (Hanna & Olken, 2023). RHISN is also notable for its extensive scope, contributing to the nascent body of research on large-scale interventions and poverty traps (Balboni et al., 2022).

3. *Evidence on impacts of public work programs on female empowerment:* The program involves a designated adult from each household working on public projects for up to 20 hours weekly (or 40% of the household's labor capacity) over three years. Public works employment could displace other productive activities within the household. The compensation for this work is expected to exceed typical market wages, approximately doubling monthly household income through direct payments. Beneficiary selection is household-driven, and pilot data suggests a large majority of participants are likely to be women. Consequently, RHISN may radically alter women's roles, enhancing their mobility, labor market participation, and contribution to household income. This shift provides a unique opportunity to explore the effects of these changes on labor distribution and decision-making dynamics within households, contributing to the work examining the impacts of social protection on gender equality (Buller et al. 2018, Bastagli et al. 2019, Baranov et. al 2021, Olney et al. 2022, Perera et al. 2022, Puteh and Halim. 2023, Peterman et al. 2022, Timu & Kramer 2023). Importantly, one relevant contribution relative to previous work is that this project will assess the effects of the program not only on gender equality but also on the dynamics of gender norms.
  
4. *The general equilibrium effects of social protection programs.* Our study is designed to quantify the general equilibrium effects of a public works program using a two-stage clustered RCT, something only a few studies – and none in the refugee space or an urban setting – have been able to do (see Gassmann et al. 2023, or Egger et al. 2022 for the only RCT in rural Kenya). A rare exception is Franklin et al. (2023) who study the general equilibrium effects of a precursor to RHISN using its random roll-out over two years across neighborhoods in Ethiopia's capital Addis Ababa on the local labor market. Our study areas are much poorer than Addis, and RHISN consequently represents an exceptionally large shock to the local economy: it will target between 15 – 60% of all households living in refugee camps and surrounding cities. Public works remuneration and business grants together correspond to roughly 100-150% of local GDP across the three years of the program, and public works activities constitute up to 10% of adult labor supply. Rather than relying on random rollout alone, our two-stage clustered saturation design creates substantial variation in the intensity of the program across neighborhoods in host cities and blocks within refugee camps, allowing us to study both the direct and the local general equilibrium effects of the program. Local labor supply reductions allow us to trace out the labor demand curve. Local inflows of cash and expenditure will quantify the impact on local production, firm- and market-level supply elasticities, and effects on prices and wages in general equilibrium. Our extensive data collection will also allow us to measure market prices and enterprise outcomes and track a representative set of beneficiary and non-beneficiary households over a long-time frame.

5. *Intergroup contact and social cohesion.* This project builds on previous work exploring the effects of inter-group contact on social inclusion (Allport 1954, Rao 2019; Scacco and Warren 2018; Mousa 2020; Lowe 2021). We contribute by assessing the effects of intergroup contact through randomized multi-year co-working relationships on social cohesion. This unprecedented intensity and duration of contact, in a high-stakes real-world setting, will contribute to our understanding of how contact affects network formation, beliefs, attitudes and social cohesion more broadly. Moreover, we randomly vary outgroup exposure not only at the individual level but also at the group level (minority vs. majority refugee groups) and the neighborhood level, allowing us to study how the effects of contact vary by intensity and are shaped by within-group social dynamics (e.g. within-group segregation or selection of team leads). Lastly, using detailed bilateral baseline social and economic network data and variation in contact exposure across neighborhoods, we will be able to study how attitudes and social norms propagate through such networks.

6. *Impacts of diversity in productivity.* This study examines how team diversity affects workplace productivity. Most existing work has been focused on the impacts of diversity in lab experiments and classroom settings.<sup>3</sup> A rare exception is Hjort (2014) who studies how the mixing of different tribes affects productivity of work teams in flower production in Kenya. We will build on this literature by examining the effects of inter-group contact on public work productivity. The scale and duration of group interactions within the RHISN program, the variety of different real-world tasks performed, as well as the size and diversity of groups are novel. We will study how diversity affects within-group dynamics, including task and leadership assignment, cooperation, communication (including language skills), within-group segregation, individual and group productivity, bilateral subjective evaluation of other group members, and long-term social and economic network formation). We are also (to our knowledge) the first to separately vary after-work socializing and study its impact relative to on-the-job exposure.

***Policy relevance.*** Our research findings are expected to attract significant policy interest. Ethiopia, as one of the pioneering Sub-Saharan nations to assimilate refugees into its national social protection framework, represents a unique case study. By examining this program comprehensively, including its economic spillover effects on non-recipients, we aim to deepen our understanding of these initiatives, as well as the dynamics and interactions within refugee-host environments. Our large sample size enables an exploration of heterogeneous impacts across different social and economic environments and individual characteristics, thereby enhancing the generalizability of our findings.

---

<sup>3</sup> Most empirical evidence corresponds to case studies (Hamilton et al 2003, 2004; Leonard and Levine 2003; Kurtulus 2011), simple correlations between labor diversity and firms' outcomes (Grund and Westergaard-Nielsen 2008; Barrington et al 2001; Erhart et al 2003; Canella Jr et al 2008), or studies that only examine the effects of diversity of the firms' main manager, CEO, or board of directors (Kilduff et al. 2000, Giannetti and Zhao 2019, Hoogendoorn and Van Praag 2012).

This study will impact the lives of millions of households by refining future expansions of the Refugee and Host Integration through the Safety Net (RHISN) and of the safety net in general, shaping refugee policies in Ethiopia and more generally. Our insights will also inform policy development and funding strategies globally, addressing the challenges of forced displacement. A thorough assessment of the program's effectiveness will offer valuable insights into the role of social protection in meeting refugees' basic needs, supporting local community development, and promoting harmony between refugees and local populations. Given the innovative nature of RHISN, our study serves as a 'proof of concept', potentially guiding a global shift towards socio-economic integration in refugee policy.

## **2. RESEARCH DESIGN**

### **2.1 Program implementation and local context**

The RHISN program is led by the Ministry of Urban Infrastructure (MUI) and managed by the Federal Project Coordination Unit (FPCU), with the Refugees and Returnees Service (RRS) providing critical support. MUI is primarily responsible for RHISN's execution, while RRS focuses on targeting, safeguarding, and aligning the project with existing refugee policies and programs. A Steering Committee, consisting of ministers and senior officials from these organizations, oversees the project. Additionally, the RHISN Technical Working Group (TWG), which includes representatives from MUI, FPCU, RRS, and the World Bank, offers technical advice.

The project's geographical focus is on refugee camps and sites and their hosting communities. The decision to include these areas, along with beneficiary quota allocation, was made by the urban safety net's Steering Committee, guided by recommendations from the RHISN TWG. These recommendations were based on socio-economic indicators, program feasibility, proximity of host communities to camps, baseline social cohesion, local implementation capacity, and infrastructure.

The process of selecting beneficiary households, which involves eligibility determination, Proxy Means Test (PMT) targeting, and randomization, was developed by the RHISN TWG. The FPCU approved this approach, drawing on previous targeting methods, data availability, and experience in impact assessments within Ethiopia's social protection and refugee sectors.

Our study takes place in the following four study areas:

1. Bambasi in Benishangul-Gumuz, hosting primarily refugees from Sudan.
2. Pugnido in Gambella, hosting primarily refugees from South Sudan.
3. Asayita in Afar, hosting primarily refugees from Eritrea.
4. Dabat in Amhara, hosting primarily refugees from Eritrea.

Each camp and neighboring city consists of a refugee camp with between 15,000 and 60,000 refugees in each camp and a nearby host city, typically a small urban agglomeration with populations between 25,000 and 54,000. Host cities are within walking distance of refugee camps and although refugee camps are formally separate entities, and while refugees have the right to participate in the host's formal labor market and there are some basic though varying levels of social and economic integration around each camp, refugees struggle to obtain formal documentation attesting their right to work and so to access formal jobs and open formal businesses. Cities are divided into *kebeles*, the lowest administrative unit of the Ethiopian

government. Refugee camps are administratively divided into zones. For implementation purposes, *kebeles* are matched to zones, so that refugees from a certain zone only engage in public works in their matched *kebele*. Baseline integration varies substantially across sites.<sup>4</sup>

## **2.2 Intervention – Phase 1**

In Stage 1 (the first 12 months), program beneficiary households will be assigned to homogenous or mixed refugee-host groups and work, in those groups, in remunerated public work projects, producing small-scale public goods like street cleaning, greenery and small watershed management. Refugees also receive a residence permit allowing them to work legally. Public work groups consist of 30 individuals who work together on public work projects for up to 2 years, 4 days per week and up to 5 hours per day. Beneficiaries also attend livelihood trainings on financial management and life-skills in their groups and receive coaching. Some public work groups will be invited – by public lottery – to attend monthly teas, annual dinners, and a cultural festival to promote after-work socializing. Public works participation is remunerated with direct cash payments that equal approximately 100% of recipient households’ monthly incomes. In Stage 2 (12 – 24 months), beneficiaries receive additional livelihoods training, support for business plan development, and a USD 600 business grant.

## **2.3 Research Questions**

### Primary questions

- 1- What are the impacts of RHISN on the well-being of the program’s beneficiaries?
- 2- What are the impacts of team diversity on productivity and social cohesion?
- 3- What are the spillover effects of RHISN on local economic aggregates and social cohesion?

### Secondary questions:

- 4- How does varying the share of outgroup members impact team dynamics, task allocation, labor attrition, communication, and productivity?
- 5- Does additional after-work socializing (teas, dinners, cultural festivals) amplify the impacts of inter-group contact?
- 6- How do social norms, attitudes towards the outgroup, and social cohesion propagate through existing social and economic networks?

## **2.4 Empirical Design.**

RHISN will target over 22,500 beneficiary households, or approximately 10-60% of the local population in each camp.

---

<sup>4</sup> In Asayita, for instance, refugees and hosts share a tribal affiliation, religion, and language. Refugees typically visit the local market to buy and sell goods or work for wages. Host and local populations also use the same hospital. In Pugnido on the other hand, refugees and hosts often have different tribal affiliations, religion and there is sporadic violence. While three of the refugee camps were established a long time ago, and refugees have typically lived there for more than ten years (or were born there), Dabat is interesting because it was set up only 3 years ago as Eritrean refugees hosted in the Tigray region were displaced due to the recent conflict there.

### *Program Eligibility:*

- a) Refugees: all refugees are eligible. The full refugee registry as maintained by RRS represents the sampling frame. This refugee registry was posted for several weeks in refugee zone administration buildings, and open for appeals. Beneficiaries are selected at random from the post-appeals list up to the assigned saturation quota (details below).
- b) Host population: among host populations, eligibility is determined through a proxy-means test (PMT) of consumption poverty. Initially, all host households are censused, and a simple vulnerability score based on 11 questions is computed. Households scoring the highest (most vulnerable) in this exercise were then visited by enumerators for the PMT.<sup>5</sup> 37% of the beneficiary quota was then automatically assigned to the poorest households according to the PMT. The remaining 63% of the quota was then assigned randomly among the remaining households.

The urban safety net's Steering Committee allocated total quotas for each camp and host city, based on overall estimated demand. Because demand estimates were not fully accurate, the share of households overall that will participate in RHISN varies somewhat across sites. Within camps, caseloads were allocated proportionally to the number of longlisted households, i.e. those that were eligible and expressed interest in the program, across kebele and camp zones.

### ***Randomization:***

***T1 – Participation in RHISN:*** Eligible host community and refugee households are randomly selected for participation in remunerated public works, a legal work permit, livelihoods training, and USD 600 business grants.

***T2 – Contact through mixed refugee/host work teams:*** RHISN beneficiaries will be randomly assigned to public work groups. Work groups consist of ~30 individuals who work together on public projects. There are 4 types of groups: 'Refugee-only groups' and 'Host only groups' comprise only refugees or host, respectively. 'Equal mixed groups' have 15 refugees and 15 host members; and 'refugee minority groups' have approximately 20 hosts and 10 refugees. To ensure feasibility of mixing, camp zones and host kebeles are matched based on geographic proximity, and relative number of eligible households, forming a correspondence between host kebele and camp zones.

***T3 – Social cohesion activities:*** A random subset of public work groups will be invited – by public lottery – to attend monthly teas and dinners, and an annual cultural festival promoting after-work socializing.

***T4 – Local program saturation:*** We randomly vary the saturation of a) the program overall, and b) the extent of outgroup exposure within small geographic areas. Host communities are administratively organized into kebeles and ketenas of approx. 200 households each (with roughly 4-6 ketenas per kebele). On the refugee side, camps are organized into zones and block groups of roughly 250 households each. There are 83 ketenas and 97 block groups, or a total of 180 clusters in our study area. We cross-randomize at the cluster level in two dimensions.

A: We randomize the ***intensity of program saturation:*** Each cluster has an implied raw beneficiary quota if total city-level cases were assigned proportionally to each neighborhood's share of eligible

---

<sup>5</sup> The number of households visited for this activity was intended to be at least twice the RHISN caseload in each camp, though in some camps there were not enough households available in the census (as caseloads exceeded 50% of the censused population).

households. In high-saturation clusters, the total number of eligible households selected for RHISN will be 125% of the raw quota, in low-saturation clusters 75% of the raw quota. This induces random variation in program intensity across neighborhoods/clusters, with treatment intensity being 1.67 times larger as a share of the population in high-saturation clusters, allowing us to estimate economic spillover effects of the program on non-recipients, and on the aggregate economy.

B: We cross-randomize the *intensity of outgroup contact*: In high-exposure neighborhoods/clusters, 2/3 of RHISN beneficiaries are assigned to mixed groups, while in low-exposure clusters only 1/3 are mixed.

## 2.5 Sample and Statistical Power

We will collect data from households, public work groups, firms, and market sellers:

*Households*: For Baseline household surveys, we sampled 12,000 households overall across our 4 refugee camps from the longlist of eligible households. For refugees, all households are eligible, resulting in a representative sample of the entire refugee population. For hosts, the longlist of eligible households was intended to cover the entire population, though a small share of households may have been missed during the census. Sampling probabilities vary slightly across camps to consider slight differences in the share of longlisted households that were selected for treatment. The sample will be equally split between eligible refugees (6,000) and eligible hosts (6,000). For hosts, we only sample among those randomly selected (or not) for treatment, and do not sample from those automatically enrolled through PMT targeting. To be equally powered for the main research questions 1 (effect of RHISN), and 2 (effect of mixing/out-group exposure), we sample twice as many treated households as controls (8,000 vs.4,000). Target sample sizes are in Table 2. For Endline 1, we follow-up with all households reached at baseline (including replacement households).

**Table 2 – Saturation Design and Sampling**

	<b>High exposure (T2):</b> 1/3 refugee/host only group 1/3 equal groups 1/3 refugee minority group	<b>Low exposure (T2):</b> 2/3 refugee/host only group 1/3 equal groups 1/3 refugee minority group
<b>High Saturation:</b> 2/3 of eligible households selected for RHISN (T1)	45 clusters/neighborhoods 140 public work groups (70 with social cohesion activities – T3)  <b>Sample:</b> 3334 households 667 control, 2667 treated (889 refugee/host only, 889 equal group, 889 refugee minority)	45 clusters/neighborhoods 140 public work groups (70 with social cohesion activities – T3)  <b>Sample:</b> 3334 households 667 control, 2667 treated (1779 refugee/host only, 444 equal group, 444 refugee minority)

<p><b>Low Saturation:</b> 1/3 of eligible households selected for RHISN (T1)</p>	<p>45 clusters/neighborhoods 70 public work groups (35 with social cohesion activities – T3)</p> <p><b>Sample:</b> 2666 households 1333 control, 1333 treated (444 refugee/host only, 444 equal group, 445 refugee minority)</p>	<p>45 clusters/neighborhoods 70 public work groups (35 with social cohesion activities – T3)</p> <p><b>Sample:</b> 2666 households 1333 control, 1333 treated (889 refugee/host only, 222 equal group, 222 refugee minority)</p>
--------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

*Public work groups:* Beneficiaries will be working in a total of 549 public work groups of about 30 members each. We will directly observe productivity, group dynamics, interactions, leadership selection, etc. of all groups once each quarter over the first 12 months, yielding a total of 2,196 group level observations. We will randomly sample 1/4<sup>th</sup> of all group members in each quarter during group-level observations for short interviews about within-group dynamics and experiences, eventually sampling all 8000 households in the treatment sample at least once between Baseline and Endline 1.

*Market prices:* We will collect up to 3 price quotes for approximately 100 goods and services across all major expenditure categories every two months starting from Baseline up to 12 months after enrollment for approximately 23 markets (15 in treatment and 8 in control areas).

*Market sellers:* We will conduct short surveys with approximately 1,000 sellers in all markets.

*Fixed enterprises:* We have data from approximately 4,000 enterprises run by households in our household sample. We collect data on these enterprises as part of the household surveys. We additionally census, at Baseline and Endline 1, all enterprises in our study area with a fixed location (building, stall) or a fixed presence. We estimate approximately 5 – 6,000 enterprises across the 4 locations.

**2.6 Power Calculations**

We conduct power calculations for the main outcomes, and the research questions outlined above. While we may eventually conduct more involved analyses (see Analysis section below), the power calculations here are for relatively simple reduced-form treatment effects of our different treatment arms, where treatment is defined as a dummy variable. We focus here on household level outcomes, since the identification strategies for price effects and firm outcomes are different, and since we do not know the exact number of fixed enterprises that we will encounter. To answer each of the research questions we will use the following strategies:

- Question 1.* Compare randomly selected beneficiaries and control households.
- Question 2.* Compare beneficiaries randomly assigned to mixed vs. host/refugee-only groups.
- Question 3.* Compare treated and control households in high- vs. low-saturation clusters (for economic spillovers) and high- vs. low-refugee exposure clusters (for social spillovers).
- Question 4.* Compare those randomly allocated to equal groups vs. those assigned to refugee minority groups.
- Question 5.* Compare groups randomly invited to social cohesion activities to those not invited.

*Question 6.* We use the fact that variation at the cluster level together with detailed baseline social and economic network data creates variation in the share of a household’s connection that is exposed to the outgroup (conditional on one’s exposure).

We require estimates of ICC for the treatments that are going to be randomized at the cluster level. We obtain these from two different sources: i) The Socio-economic Survey of Refugees in Ethiopia (SESRE), which completed its first round between January and February 2023 and targeted all our impact evaluation sites, with a total sample of approximately 2,000 refugees and hosts; and ii) the administrative data collected as part of the RHISN program targeting (i.e. the proxy-means test and screening questions for determining program eligibility). We estimate ICC for key economic outcomes and social cohesion, as those will be our main primary outcomes. ICC estimates vary between 0.00 and 0.13. We, thus, use the highest and most conservative value (i.e., 0.13) for all cluster- and group-level randomizations for our power calculations (see Table 3 below).

**Table 3 – ICC estimates**

<b>Outcome</b>	<b>Hosts</b>	<b>Refugees</b>	<b>Source</b>
Total expenditure	0.02	0.01	SESRE
Food expenditure	0.00	0.04	SESRE
Food insecurity index	0.11	0.06	SESRE
Entrepreneurship indicator	0.02	0.00	SESRE
Locus of self-control index	0.04	0.09	SESRE
Outgroup attitudes index	0.13		
PMT score (log consumption)	0.03		RHISN
Vulnerability score		0.04	RHISN

We conduct power-calculations for our total sample, as well as for select sub-groups, including refugees and hosts only, and (for spillovers) beneficiaries vs. non-beneficiaries. Power calculations pool samples across cross-randomized treatments, i.e. we will not separately estimate interaction effects in our main specifications, though we may do so as part of more exploratory analyses. Table 4 suggests that we are adequately powered for all main research questions. We can detect changes in outcomes of between 0.05 and 0.17 standard deviations for all questions and subgroups.

**Table 4 – Power Calculations**

Research Question	Subgroup	Unit of Randomization	Sample: Treatment	Sample: Control	Number of Clusters or Groups	ICC	MDE (in SD)	MDE on consumption (in %)
<i>Primary</i>								
(1)	All	Household	8000	4000			0.05	3%
	Refugees	Household	4000	2000			0.08	4%
	Hosts	Household	4000	2000			0.08	4%
(2)	All	Household	4000	4000			0.06	3%
	Refugees	Household	2000	2000			0.09	5%
	Hosts	Household	2000	2000			0.09	5%

(3)	All	Cluster	6000	6000	90 vs. 90	0.13	0.16	8%
	Refugees	Cluster	3000	3000	90 vs. 90	0.13	0.17	9%
	Hosts	Cluster	3000	3000	90 vs. 90	0.13	0.17	9%
	Recipients	Cluster	4000	4000	90 vs. 90	0.13	0.16	9%
	Non-Recipients	Cluster	2000	2000	90 vs. 90	0.13	0.17	9%
<i>Secondary</i>								
(4)	All	Household	2000	2000			0.09	5%
	Refugees	Household	1000	1000			0.13	7%
	Hosts	Household	1000	1000			0.13	7%
(5)	All	Work group	4000	4000	275 vs. 274	0.13	0.10	5%
	Refugees	Work group	2000	2000	137 vs. 137	0.13	0.14	7%
	Hosts	Work group	2000	2000	137 vs. 137	0.13	0.14	7%

We are adequately powered for our MDE to be well below the meaningful economic range of possible effects. For a program of this scale, at this intensity, and with such large variation in saturation, we expect effects to be large, in particular:

- *Direct impacts:* RHISN stands out as a particularly generous social protection program, providing direct cash payments that equal approximately 100% of recipients' incomes. This significant financial support bolsters our confidence that the MDE we have established is sufficient to observe any economic impacts of the program. Uniquely, RHISN explores the effects of out-group contact in a professional setting with an intensity not seen in other programs. It involves randomly formed public work groups collaborating up to 4 days weekly for as long as 3 years. In contrast, existing research typically focuses on shorter-term interactions in sports or work teams, or prolonged mixing in school classrooms, but not in professional environments. Should RHISN fail to achieve at least our MDE in enhancing social cohesion, it's unlikely that other programs could surpass this benchmark.
- *Program saturation:* RHISN is a significant economic stimulus at the local level, with an estimated 15-60% of households eligible. Due to our saturation design, treatment intensity varies significantly, with 12-75% of households in treated regions impacted. This variation is a result of the selection ratio being 1.67 times higher in high-saturation areas compared to low-saturation ones. Consequently, even without increased local production, we expect a substantial rise in economy-wide spending, ranging from 12-75% of the local GDP.

Our research addresses novel questions, particularly among refugee populations and their host communities in urban and peri-urban areas. A related study by Egger et al. (2022) on the general equilibrium effects of cash transfer programs suggests potential significant economic spillovers. Assuming a local economy multiplier of 2.5, as indicated in Egger et al. 2022, we anticipate that the effects on the control group could range from 22% to 90% of those in the treatment group. Conversely, we are also prepared to detect a zero-multiplier scenario, where the treatment effects on the control group are negative, ranging from 18% to 150% of the treatment group. We are equipped for both scenarios, ensuring our MDE is robust enough to meaningfully bound spillover effects for policy implications.

An important consideration is the timing of direct and spillover effects. Differences in flow variables (like income and consumption) between treated and control groups can vary over time, even with a constant multiplier. To accurately capture these spillovers, we will collect high-frequency, short-term data on expenditures and prices, especially during periods when spillovers are less likely to have impacted the control group. This strategy helps us address concerns about control-group catch-up and enables an analysis of both the initial effects and the dynamics of spillover manifestation.

- *Outgroup exposure intensity*: In areas with high levels of mixing, two-thirds of all beneficiaries will participate in mixed groups, compared to only one-third in low-mixing areas. Given our program's saturation design and the high eligibility rate, this structure means that 5% to 40% of local households will be involved in mixed public works groups. Consequently, even non-recipients will likely be impacted through their social and economic networks. Despite the lack of prior studies specifically examining social spillovers from outgroup exposure, we hypothesize that the intensity and variation of contact in these mixed groups could lead to significant effects, at least as large as our MDE. On the other hand, if we fail to observe effects of at least the MDE, it suggests that few, if any, real-world programs could achieve economically meaningful effects of this nature.

## **2.7 Additional sample size considerations, risks, and deviations from planned implementation**

The somewhat elevated sample size and small MDEs are additionally justified by the following considerations:

1. Our objective is to ensure that the sample size is large enough to provide sufficient statistical power for each research question. The intervention design includes three levels of cross-randomization and focuses on three main research areas. Two of the three primary research questions relate to the extent to which out-group exposure (at work, and after work) affects social cohesion. These can only be studied within the treated group, thus rationalizing the larger sample size within the treated group relative to the control group. Consequently, a larger sample size assures that we will have adequate power to address all the outlined questions.
2. The RHISN program will be implemented in four refugee-hosting regions in Ethiopia, each exhibiting distinct socioeconomic statuses and levels of refugee-host integration. A larger sample size is advantageous for analyzing the program's effects within each region and for providing exploratory insights into outcome variations. While our main analysis will not separately estimate the interaction effects of different treatments, the broad scope of this program and the large sample size will allow us to delve into these aspects to a certain degree.
3. The program's implementation in potentially fragile regions of Ethiopia, coupled with the high mobility of refugee populations and the possibility of rapid and unpredictable local changes due to the reactivation of pre-existing conflicts adds a layer of complexity. These factors underscore the importance and relevance of our research but also introduce risks like attrition and non-compliance. Our initial power calculations assumed no attrition and complete compliance. To accommodate potential issues and unforeseen contingencies common in such projects, we selected a sample size that would remain robust and flexible enough to address all research questions, even with these

challenges. The larger sample size also provides a buffer against any political shifts in these regions.<sup>6</sup> In the case of contingencies, we will always try to keep as close as possible to the original pre-analysis plan design while considering political and security concerns in each region.

### 3. DATA

For this research, we will rely on exceptionally rich and extensive data collected from households, enterprises, markets, and public work groups:

#### 3.1 Household data (Key outcomes described in this PAP)

We will collect data from a sample of 12,000 households. The sample is split equally between hosts and refugees, and the sampling probability is twice as large for RHISN beneficiaries – an optimal strategy based on equally weighting research questions 2. Within each site, sampling is stratified by cluster and (within the treated group) by their assignment to mixed vs. non-mixed groups and is representative of each group. The sample size across sites was intended to be representative of the overall caseload across sites. However, in some camps, assigned caseloads relative to the local population were higher than in others, so that a proportional assignment of the sample across camps would leave too few control households in some sites. We consequently lowered the sample there in favor of other camps. We describe below how we intend to re-weight observations to obtain a representative treatment effect for the entire RHISN beneficiary population.

*Baseline and Phase 1 Endline:* The baseline and endline household surveys consist of two components:

Part A: Covers household-level (or primarily household-level) outcomes, including demographics, education, housing and assets, household finance, consumption, income from agriculture, livestock, self-employment, employment, assistance and transfers as well as questions on public goods and crime. The target respondent for this questionnaire is the household head.

Part B: Covers social networks, community groups and participation, language skills, social capital, social cohesion and attitudes, health and mental health, and female empowerment. The target respondent for this questionnaire is the (intended) RHISN beneficiary.

During the baseline, households that could not be located were substituted with another randomly selected household from the same sampling strata. This was capped at a maximum of 25% replacements per stratum. The exact replacement rate is currently unknown, but we plan to analyze if replacements vary by region and treatment status. Should significant disparities in replacement rates emerge across different locations

---

<sup>6</sup> Currently, there are security concerns at the Bambasi camp, affecting the implementation of mixed public work groups. A key footpath linking the camp to the nearest city is deemed unsafe. As a result, our partners have limited the mixing of public work groups to just one Kebele and one camp zone, applying it to all participants in these areas instead of a randomly selected subset. While program participation and saturation remain random, outgroup exposure is not. In response, we have made two adjustments: 1) Our analysis of out-group mixing's impacts will now focus exclusively on the three other camps, excluding Bambasi. 2) We have increased the sample size in Bambasi's likely implementation areas to maintain an average rate of mixing comparable to other camps. Concurrently, we have reduced Bambasi's overall sample size relative to its proportion in the RHISN program. To address changes in representativeness across camps, we plan to estimate treatment effects separately for each site and then re-weight them to calculate the average treatment effect for all program beneficiaries.

or demographic groups, re-weighting the sample to ensure representativeness may be considered. Endline 1 targets all households (including baseline replacements and replaced households). There will be no further replacements for households that aren't tracked. Instead, we will report the tracking rates and examine any selective attrition. We will conduct the analysis for two samples: a) including baseline households only, and b) including households that were missed at baseline.

Since the beginning of the baseline, the Amhara region has been experiencing a violent conflict between local militias and the central government. This conflict has led to internet disruptions, significantly hindering our capacity to gather baseline data from households in the Dabat refugee camp and its surrounding areas. Consequently, our baseline balance and attrition tests will exclude the Dabat camp, focusing only on the other three camps, with an exception for unchangeable household characteristics. For Dabat, we will rely on pre-intervention administrative data for balance checks, including vulnerability scores, Proxy Means Test (PMT) scores, and basic demographics, supplemented with data from UNHCR registries.

### **3.2 Enterprise data (outcomes will be specified in a separate PAP)**

We capture enterprise data in three distinct ways. First, enterprises operated by households in our household survey are captured during the household survey through an extensive self-employment module. Second, we will conduct Baseline and Phase 1 Endline 1 listing of *all* enterprises operating from a fixed location in all our study sites (cities and camps). During this listing, we will also collect basic information on enterprise ownership, category, activity, employment, revenue and profits, allowing us to measure overall economic activity in each neighborhood. Lastly, we will visit each of the 25 study area markets (the universe of all weekly markets in our sites) and matched control markets (more details in the price data section below) to conduct a short survey capturing the number of sellers, ownership, employment, sales, profits and imports every two months between Baseline until Endline 1 to measure market activity dynamically.

### **3.3 Public work group data (outcomes will be specified in a separate PAP)**

In the first 12 months after enrollment for the program, we will conduct one spot-check visit with each of the 133 public work groups at a work site/project where they are working in Dabat camp.<sup>7</sup> For each group, we will collect:

1. Direct public observations. This will capture details of the work carried out, and about work productivity that will depend on the particular project (e.g. acres cleared, kg of rubbish collected, number of trees planted, etc.). We will also capture attendance, and interactions between group members (e.g. who works together, communicates, etc.).

---

<sup>7</sup> We focus on Dabat because implementation monitoring suggests that randomized assignment and mixing of public work groups was implemented with high fidelity only in Dabat. In Bambasi, mixing was not randomized initially due to security concerns. In Pugnido, low implementation capacity and ethnic tensions between camp and local populations prevented mixing from being implemented. In Asayita, high temperatures and a lack of transportation options prevented refugees from working within the host city, and effectively restricted mixing only to refugees living within the local city.

2. Interviews with group members about their public work experience, task allocation, leadership assignment, satisfaction, and bilateral ratings of productivity and cooperativeness, etc. of other group members.
3. We will also collect behavioral measures of group cohesion, altruism, task assignment and coordination, including incentivized dictator and coordination games.

### 3.4 Price data (outcomes will be specified in a separate PAP)

We will collect up to 3 price quotes for approximately 100 goods, in 23 markets. Products were selected according to their weight in the Ethiopian CPI, and to be broadly representative of different product categories, as well as generally available in our study areas. We will visit markets every two to three months (including at baseline) over the first 12 months. We will also register the total number of sellers for each product and the availability of each product. This will yield a total of approximately 55,800 price quotes (of which 19,800 within our study area) over the course of the first 12 months.

## 4. EMPIRICAL SPECIFICATIONS

We will estimate a variety of different effects on different populations of interest:

### Question 1. Estimating the direct effects of RHISN

We will use the following ITT specification to assess the direct impacts of the program on its beneficiaries:

$$y_{ic} = \beta_0 + \beta_1 T_i + \beta_2 HiSat_c + X_i' \Lambda_2 + \gamma y_i^{BL} + \alpha M_i^{BL} + \epsilon_i \quad (1)$$

In this equation,  $i$  represents a household or individual, and  $c$  denotes a cluster or neighborhood.  $T_i$  is an indicator for whether individual  $i$  was chosen for the RHISN program, while  $HiSat_c$  indicates whether the cluster/neighborhood falls into the high- or low-saturation group. The control vector  $X_i'$  includes stratification variables used in randomization, such as camp zone indicators for refugees and kebele (local administrative units) for host individuals. The baseline outcomes  $y_i^{BL}$  are measured at baseline, with mean values substituted for any missing data, denoted by the missingness indicator  $M_i^{BL}$ . Standard errors are clustered at the cluster level (ketena/neighborhood or camp block group). In this model,  $\beta_1$  calculates the direct Intention-to-Treat (ITT) effect of selection for RHISN. In cases of significant non-compliance, Treatment-on-the-Treated (TOT) effects will also be assessed, using treatment assignment as an instrument for actual treatment uptake. Additionally, effects will be analyzed separately for refugees and hosts and by area.

### Question 3. Measuring the economic spillover effects of RHISN:

$\beta_2$  in Equation (1) estimates a simple version of neighborhood spillovers of RHISN. Economic spillovers, however, may not be contained within pre-defined neighborhoods. As alternative approaches to estimating spillover effects, we will use two complementary specifications.

First, we will measure the impact of the RHISN program based on geographical proximity, following established methods in existing research (e.g., Miguel & Kremer 2004, Muralidharan & Niehaus 2016, Egger et al. 2022). The core of our analysis is Equation (2):

$$y_{ic} = \beta_0 + \beta_1 T_i + \sum_r \beta_r \text{ShareTreat}_{cr} + X_i' \Lambda_2 + \gamma y_i^{BL} + \alpha M_i^{BL} + \epsilon_i \quad (2)$$

where  $\sum_r \beta_r \text{ShareTreat}_{cr}$  captures the spillover effects.  $\text{ShareTreat}_{cr}$  represents the proportion of eligible households within a certain radius  $r$  around cluster  $c$  that were selected for RHISN. The coefficients measure the impact at each radius. The  $\beta_r$  coefficients measure the impact at each radius. We will use geographic buffers around each household, firm, or market (depending on the outcome level) to define treatment exposure. These buffers will be set at 200m intervals. The number of these buffer zones included in the analysis will be determined by minimizing the Schwartz Bayesian Information Criterion (BIC), ensuring the most efficient model fit. The spillover effects ( $\beta_r$ ) at each radius will be presented individually, but we will also combine them to evaluate the average impact across different treatment shares. This approach helps in understanding both the localized and broader impacts of the RHISN program.

Second, we will use an exposure approach based on baseline exposure metrics (mostly relevant for outcomes in future PAPs, and represented here for completeness). This method is important because it recognizes that economic activities and entities do not always align neatly with geographic distances or administrative boundaries. For instance, when we measure outcomes like the number of firms, their revenue, market prices, or even a firm's revenue, we need to consider more nuanced factors than just location. For example, when  $y_{ic}$  is the number of firms, revenue, or prices in a local market, or the revenue of a firm, we will use baseline data on shopping patterns by households to calculate treatment exposure:

$$y_i = \beta_0 + \beta_1 T_i + \beta_2 \text{MktShareTreat}_c + X_i' \Lambda_2 + \gamma y_i^{BL} + \alpha M_i^{BL} + \epsilon_i \quad (3)$$

where  $\text{MktShareTreat}_c$  represents the share of RHISN-eligible households that reported shopping at a specific market  $r$  at baseline, weighted by their total expenditure in that market, and who were later selected for RHISN. This helps to understand the program's influence based on actual shopping behaviors and patterns.<sup>8</sup>

Equations (2) and (3) are designed to measure the spillover effects of the RHISN program within and across neighborhoods in each of our study cities. However, it's possible that RHISN could have a significant impact on broader economic indicators like inflation and overall market activity in an entire city or camp. To assess this wider impact, we compare the prices and market activities in each treatment region with 3-5 similar markets outside our study area.

These comparison markets are selected based on their proximity and relevance. They are the nearest markets where the Ethiopian national office of statistics regularly collects monthly price data for the same products we survey. To ensure the validity of our comparisons, these markets are located at least 50 to 100km away from our study sites. This distance helps us minimize the potential of spillover effects influencing our

---

<sup>8</sup> Depending on the outcome, we may also consider Equations (1), (2) and (3) as 'reduced-forms' for a larger IV analysis, and use the endogenous regressors as instruments for another variables (such as the total amount of money transferred from RHISN into buffer  $r$ , or the average RHISN public works hours by households in buffer  $r$ , etc.) when we are interested in quantifying the impact not of treated shares, but of more economically meaningful local economic aggregates. The reduced-form (Equation 2) will be identical in each case. We may also use such reduced-form estimates to estimate parameters in structural models. The particular model that we will choose may determine exactly which specifications we run, and how to compute model-consistent exposure measures for Equation (3) (e.g. in a standard quantitative spatial GE commuting model, these would be market access terms).

results. Our analysis will utilize a matched difference-in-difference approach, comparing prices and market activities in our study sites to those in the control markets. This method will allow us to discern the overall impact of RHISN on each site.

Additionally, to gain a deeper understanding of trends over time, we may also examine microdata on prices from the control markets and some of our study area markets. These data, collected by the Ethiopian statistical office, will provide further insights into the pre-existing economic conditions and trends prior to the implementation of RHISN. This retrospective analysis will help us contextualize our findings within the broader economic landscape of the regions under study.

#### Questions 2 and 4. Estimating inter-group contact effects of RHISN

Our approach will be analogous to the direct and spillover effects of RHISN participation. A crucial difference, however, is that mixing only occurs in the treated groups.

$$y_{ic} = \beta_0 + \beta_1 T_i + \gamma_1 \text{MixedGroup}_i + \beta_2 \text{HiSat}_c + \gamma_2 \text{HiMix}_c + X_i' \Lambda_2 + \gamma y_i^{BL} + \alpha M_i^{BL} + \epsilon_i \quad (1a)$$

$y_{ic}$  is now a household/individual or workgroup level outcome (e.g. social cohesion, outgroup activity, etc.).  $\text{MixedGroup}_i$  is an indicator for whether a recipient was assigned to a mixed group (i.e. refugee and hosts mixed), and  $\text{HiMix}_c$  is a cluster with high outgroup exposure (i.e. 2/3 of groups being mixed).  $\gamma_1$  now identifies the additional impact of RHISN on individuals assigned to mixed groups (with a total effect of  $\beta_1 + \gamma_1$  for this group), and  $\gamma_2$  estimates the spillover effects on neighborhoods with a high degree of outgroup exposure/mixing. As above, we will estimate these separately for refugees and hosts, by area. We will also investigate spillovers separately for the treated and control groups.

For some outcomes that we only measure for treated individuals (e.g. workgroup productivity, workgroup cohesion, etc.), we will restrict this analysis to the treated group only (thus dropping  $T_i$  from the above specification).

In secondary analyses, we will replace  $\text{MixedGroup}_i$  with two indicators for whether a workgroup has an equal number of refugees and hosts, or a minority number of refugees (which was again randomly determined) to test whether the effect of mixing depends on the relative shares of both groups:

$$y_{ic} = \beta_0 + \beta_1 T_i + \gamma_1 \text{EqualMixedGroup}_i + \gamma_2 \text{RefugeeMinorityGroup}_i + \beta_2 \text{HiSat}_c + \gamma_2 \text{HiMix}_c + X_i' \Lambda_2 + \gamma y_i^{BL} + \alpha M_i^{BL} + \epsilon_i \quad (1b)$$

#### Questions 3 and 5. Social spillovers of RHISN

We will estimate the social spillovers of the program on beneficiaries and non-beneficiaries from a higher level of out-group exposure among neighbors further, using analogous specifications as Equations (2) and (3), replacing  $\text{ShareTreat}_{cr}$  with the share of beneficiaries assigned to mixed groups in buffer  $r$ , and  $\text{MktShareTreat}_c$  with the share of individuals in a meaningful social unit (e.g. a community group, or a social network) that is assigned to mixed groups. For example, for the case of studying how out-group contact spillovers through existing social networks, we will estimate:

$$y_{ic} = \beta_0 + \beta_1 T_i + \beta_2 \text{NetShareMixed}_i + X_i' \Lambda_2 + \gamma y_i^{BL} + \alpha M_i^{BL} + \epsilon_i \quad (3a)$$

where  $NetShareMixed_i$  is the share of RHISN eligible baseline social connections (relatives, friends, economic connections) who were assigned to mixed groups.  $\beta_2$  now estimates the impact on  $i$  of their connections experiencing randomly higher out-group exposure.

**Question 6: Measuring impacts of after-work socializing**

We cross-randomized the provision of after-work socializing activities, which includes monthly dinners and teas) across different workgroups. To test the additional impact of these activities, we will estimate the following model:

$$y_{ig} = \beta_0 + \beta_1 CohesionActivity_g + X_i' \Lambda_2 + \gamma y_i^{BL} + \alpha M_i^{BL} + \epsilon_i \quad (1c)$$

for outcomes for individual  $i$  in group  $g$  (or group-level outcomes), where  $CohesionActivity_g$  is an indicator for group  $g$  being assigned to such after-work social activities. As above, we will control for all other treatments and randomization clusters, and will investigate impacts separately for refugees and hosts, mixed vs. non-mixed groups, across areas, etc.

Lastly, we may investigate complementarities between different treatments (e.g. mixing and after-work socializing, or high-saturation and high-exposure) in specifications that include interactions between each of the treatments in secondary analyses.

**5. OUTCOMES (This PAP only covers outcomes collected in the baseline and midline surveys)**

We examine the effect of the RHISN program on a broad range of outcomes, organized by group and family of outcome within each of our main research questions.<sup>9</sup> We will correct for multiple hypotheses within each outcome family. In this pre-analysis plan we are only detailing the outcomes that are measured at the individual or household level. All market outcomes and spot checks data, as well as enterprise and price outcomes will have a separate pre-analysis plan.

**PART 1: What are the direct effects of the RHISN program on beneficiaries’ well-being?**

Group 1: Economic Impacts – Survey: Phase 1 Midline Survey			
ID	Outcome Family	Unit	Variables within each family
1.1	Assets	HH (household)	<b>Primary outcome:</b> Total Assets excl. land and housing (sum 1.1.1 to 1.1.5) 1.1.1 Value of livestock 1.1.2 Value of ag and livestock productive assets (excl. livestock) 1.1.3 Value of enterprise-related assets 1.1.4 Value of household durables 1.1.5 Net financial savings

<sup>9</sup> For monetary values, we will report outcomes in nominal terms as well as real USD PPP, where we use the price index collected at the closest market to each observation as our deflator back to baseline. We will winsorize monetary values at 2% and 98% for variables that take both positive and negative values. When constructing indices, we will follow Anderson (2008). Indexes of indicators variables will generally be constructed as averages of those variables making sure that all the variables move in the same direction. If there is a mix of dichotomous and continuous variables, they will first be redefined so that they move in the same direction, then they will be standardized, then averaged, and then standardized again to create an index following Kling et al. (2007). When specific outcomes are listed as index components, we also intend to analyze these individual outcomes.

			1.1.6 Estimated housing value (estimated through rental/lease value) 1.1.7 Estimate of value of land owned or hold rights for (estimated through rental/lease value)
1.2	Household Income (annualized)	HH (household)	<b>Primary outcome:</b> Total income (sum of 1.2.1 through 1.2.6) 1.2.1 Profits from agriculture and livestock 1.2.2 Profits from self-employment 1.2.3 Wages, salaries, and in-kind transfers earned 1.2.4 Net value of remittances and goods received 1.2.5 Income from transfers, aid, and assistance 1.2.6 Rental income
1.3	Household Consumption (annualized)	HH (household)	<b>Primary Outcome:</b> Total consumption (sum of 1.3.1 + 1.3.2) 1.3.1 Food consumption 1.3.2 Non-food consumption 1.3.3 Cash expenditure (excl. consumption of household production) 1.3.4 Durables expenditure (non-food consumption excl. non-durables) 1.3.5 Housing / land related expenditure 1.3.6 Education expenditure 1.3.7 Health / medical expenditure
1.4	Financial Access	HH (household)	<b>Primary Outcome:</b> Financial Access Index (1.4.1 through 1.4.4) 1.4.1 Has Bank account (=1) 1.4.2 Has Mobile money account (=1) 1.4.3 Community savings/loan group participation (=1) 1.4.4 Has any loan outstanding (=1) 1.4.5 Total financial savings 1.4.6 Total loans outstanding
1.5	Self-employment revenue (annualized)	HH (household)	<b>Primary Outcome:</b> Self-employment revenue (1.5.3+1.5.4) 1.5.1 Dummy for any agricultural or pastoral activity run by any household member 1.5.2 Dummy for any non-agricultural activity run by any household members 1.5.3 Revenue from agricultural or pastoral activities run by members of the household 1.5.4 Revenue from non-agricultural enterprises run by members of the household 1.5.5 Total expenditure on agriculture and livestock activities (investment + operating expenses) 1.5.6 Total expenditure on non-agricultural self-employment (investment + operating expenses)
1.6	Labor supply	All adults in the household, summed up (1.6.1-1.6.5 & 1.6.8) & intended beneficiary (1.6.6-1.6.7)	<b>Primary Outcome:</b> Hours worked (1.6.1). 1.6.1 Household hours worked: Total productive hours worked 1.6.2 Household hours worked: agriculture/pastoral activities 1.6.3 Household hours worked: non-ag self employment 1.6.4 Household hours worked: paid employment 1.6.5 Unemployment indicator (share of adult household members actively looking for a job) 1.6.6 Hours spent on household chores: hours spent on household chores 1.6.7 Hours spent on leisure: hours spent on leisure 1.6.8 Hourly wage: Average hourly wages across all paid employment spells (conditional on having any paid employment). We will run this at the individual level including all household members.
1.7	Food security index	HH (household)	<b>Primary Outcome:</b> Food Insecurity Index (1.7.1-1.7.6) 1.7.1 Not enough food or money to buy food (=1).

			1.7.2 Consume less expensive food (Likert scale, standardized) 1.7.3 Borrow food (Likert scale, standardized) 1.7.4 Limit portion size at mealtimes (Likert scale, standardized) 1.7.5 Restrict consumption by adults for small children to eat (Likert scale, standardized) 1.7.6 Reduce the number of meals eaten in a day
--	--	--	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Group 2: Health, Mental, and Social Impacts - Survey: Phase 1 Midline**

ID	Outcome Family	Unit	Variables within each family
2.1	Health Status	Intended Beneficiary	<b>Primary Outcome:</b> Health index (2.2.1-2.2.6) 2.2.1 Overall self-described health status 2.2.2 Number of visits to healthcare facilities 2.2.3 Number of days of school or work missed due to health 2.2.4 Healthcare access: wanted to go but did not 2.2.5. Had any major health issues (=1) 2.2.6. Health expenditure
2.2	Mental Health and Subjective Well-being	Intended Beneficiary	<b>Primary Outcome:</b> Mental health index (index of 2.1.1-5) 2.1.1 Depression: CES-D score 2.1.2 Life satisfaction 2.1.3 Happiness Index 2.1.4 Perceived stress scale (PSS-4) 2.1.5 Locus of control index
2.3.	Aspirations	Intended beneficiary	<b>Primary Outcome:</b> Aspirations (Average of 2.3.1, 2.3.2) 2.3.1 Economic Aspirations 2.3.2 Educational Aspirations
2.4	Social capital	Intended Beneficiary	<b>Primary Outcome:</b> Number of social connections (sum 2.4.1- 2.4.3) 2.4.1 Number of close friends 2.4.2 Number of good friends 2.4.3 Number of economic relationships 2.4.4 Number of community group members (incl. church/mosque) 2.4.5 Hours spent with community groups/church/mosque

**Group 3: Outgroup Attitudes - Survey: Phase 1 Midline**

ID	Outcome Family	Unit	Variables within each family
3.1	Outgroup attitudes	Intended Beneficiary	<b>Primary Outcome:</b> Outgroup Attitude Index (3.1.1 through 3.1.12) 3.1.1 Job perception game: average difference in ranks between in-group and out-group 3.1.2 Comfortable having outgroup as neighbors 3.1.3 Comfortable with offspring marrying outgroup 3.1.4 Often eating dinner with outgroup 3.1.5 Outgroup returning something valuable 3.1.6 Share of support received from outgroup 3.1.7 Attitude toward working with outgroup 3.1.8 Attitude toward working under outgroup 3.1.9 Attitude accepting outgroup officer 3.1.10 Attitude toward same schools with outgroup 3.1.11 Net effect of refugees in Ethiopia (only hosts)

			3.1.12 Sum of activities that are easy between refugees and host community. (=1 for each option)
3.2	Refugee policies	Intended Beneficiary	<b>Primary Outcome:</b> Policy Attitudes Index (3.2.1 through 3.2.6) 3.2.1 Refugees allowed to be citizens 3.2.2 Refugees with unrestricted permits 3.2.3 Refugees granted free mobility 3.2.4 More money from humanitarian community towards refugees 3.2.5 More money from gov. community towards refugees 3.2.6 Negative effects of refugees (=1 for each option)

<b>Group 4: Social Cohesion - Survey: Phase 1 Midline</b>			
<b>ID</b>	<b>Outcome Family</b>	<b>Unit</b>	<b>Variables within each family</b>
4.1	Economic interaction	Household, Seller	<b>Primary Outcome:</b> Economic interaction index (4.2.1 through 4.2.5) 4.2.1 Number of times visited out-group market or business 4.2.2 Share of sales to outgroup 4.2.3 Co-ownership with outgroup (=1) 4.2.4 Share of outgroup as co-workers 4.2.5 Share of number of remittances from/to the out-group
4.2	Social interaction	Intended Beneficiary	<b>Primary Outcome:</b> Social interaction index (Index of 4.3.1-7) 4.3.1 Share of close friends from outgroup 4.3.2 Share of good friends from outgroup 4.3.3 Share of economic relationships from outgroup 4.3.4 Share of community group members from outgroup 4.3.5 Having tea with outgroup 4.3.6 Children friends with outgroup 4.6.7 Children play with outgroup
4.3	Social Cohesion	Intended Beneficiary	<b>Primary Outcome:</b> Social cohesion index (5.3.1 through 5.3.5) 5.3.1 Hypothetical dictator game, share of donation to outgroup 5.3.2 Incentivized dictator game, share of donation to outgroup 5.3.3 Signature to petitions to support refugees (=1, two questions) 5.3.4 Self-reported trust in outgroup 5.3.5 Self-reported integration of refugees 5.3.6 National identity
4.4	Language Proficiency	Intended beneficiary	<b>Primary Outcome:</b> Language proficiency index (5.1.1 through 5.1.6) 5.1.1 Reading proficiency in local language 5.1.2 Speaking proficiency in local language (For household head) 5.1.3 Speaks local language (=1) 5.1.4 Scoring on local language test 5.1.5 Language test comprehension 5.1.6 Self-assessment on language test
4.5	Political Integration	Intended beneficiary	<b>Primary Outcome:</b> Political integration index 4.5.1 Has any formal ID 4.5.2 Holds a position of responsibility in camp or city 4.5.3 Citizen participation index (section 15 political participation four questions, =1 for each choice and then average)

			4.5.4 Trust and satisfaction with local government
--	--	--	----------------------------------------------------

<b>Group 5: Other Impacts - Survey: Phase 1 Endline</b>			
<b>ID</b>	<b>Outcome Family</b>	<b>Unit</b>	<b>Variables within each family</b>
5.1	Migration	HH (household)	<p><b>Primary Outcome:</b> Migrated elsewhere (6.1.1)</p> <p>5.1.1. Migrated elsewhere 5.1.2. Change in the number of households members 5.1.3. Intentions to migrate elsewhere</p>
5.2	Local Amenities	HH (household)	<p><b>Primary Outcome:</b> Amenities index (6.3.1 through 6.3.7)</p> <p>5.3.1 Quality of sewerage and drainage systems 5.3.2 Cleanliness of streets 5.3.3 Rate availability of green space in the neighborhood 5.3.4 Availability and quality of public toilets 5.3.5 Respondent notices smells of drains or sewerage in local area 5.3.6 Respondent notices smell of solid waste or trash in local area. 5.3.7 Crime and Safety Index</p>
5.3	Female Empowerment	Intended Beneficiary	<p><b>Primary Outcome:</b> Female empowerment index (6.4.1 through 6.4.7)</p> <p>5.4.1 Control of resources (only female respondents) 5.4.2 Decision's control index (only female respondents) 5.4.3 Domestic discord 5.4.4 Attitudes toward gender freedoms 5.4.5 Controlling behaviors 5.4.6 Emotional violence from partner 5.4.7 Number of women older than 18 years that work for a wage outside of the household. 5.4.8 Girls' school attendance</p>

<b>Group 6: Intergenerational outcomes - Survey: Phase 1 Endline</b>			
<b>ID</b>	<b>Outcome Family</b>	<b>Unit</b>	<b>Variables within each family</b>
6.2	Child well-being	HH (6.2.1-6.2.3)/ Caregiver (6.2.4-6.2.9)	<p><b>Primary Outcome:</b> SDQ Scale (6.2.9)</p>
6.1	Education	Household	<p><b>Primary Outcome:</b> Education index (Index of 6.1.1-4)</p> <p>6.2.1 Per capita education expenditure for children: total divided by number of children in the household. 6.2.2 School enrollment for all kids in the household 6.2.3 School attendance for all kids in the household 6.2.4 Educational aspirations</p>
6.2	Child health	Household	<p><b>Primary Outcome:</b> Child health index (Index of 6.2.1-3)</p> <p>6.2.1 Child health index (number of symptoms) 6.2.2 Child healthcare access (share of symptoms treated) 6.2.2 Child nutrition index</p>
6.3	Child labor	Household	<p><b>Primary Outcome:</b> Child labor hours (6.3.1)</p> <p>6.3.1 Child labor hours 6.3.2 Time use on domestic work and chores 6.3.3 Income earned from child labor</p>
6.4	Child development	Household	<p><b>Primary Outcome:</b> SDQ scale (6.4.1)</p> <p>6.4.1 SDQ scale</p>

**PART 2: What are the impacts of team diversity and inter-group contact on productivity and social cohesion? To be added in a separate PAP.**

**PART 3. What are the spillover effects of the program on individual level outcomes?** We will examine spillover effects on all the outcomes listed in the table above which use data from the midline surveys. We will also further specify how we study the impacts on local enterprises, markets, prices and local economic activity in a separate PAP.

### **5.1 Multiple outcomes and multiple hypothesis testing**

In our analysis of the RHISN program's impact, we report naive p-values for all outcomes, catering to readers with a specific interest in whether RHISN influences these outcomes. To address the issue of multiple inferences, we calculate sharpened q-values within each family of outcomes. We follow Anderson (2008) and report the lowest q-value at which each hypothesis is rejected. Our results will include both standard p-values and these minimum q-values, reflecting our commitment to current best practices in economic research, amidst evolving norms around multiple testing.

### **5.2 Covariate adjustment**

To enhance the precision of our analysis, we will apply Double/Debiased Machine Learning (DDML), as outlined by Chernozhukov et al. (2018), to estimate versions of Equations 1, 2, and 3. This advanced statistical technique integrates machine learning algorithms with econometric models, offering a more robust and accurate estimation by correcting for biases that can arise in traditional regression analyses. By leveraging DDML, we aim to ensure that our estimates are not only precise but also reliable, providing a deeper understanding of the data and the impacts of the RHISN program.

### **5.3 Exact tests of treatment effects**

In addition to the large-sample approach outlined in Section 3.1, we will perform Monte Carlo approximations of exact tests of the treatment effect (Fisher 1935).<sup>10</sup>

### **5.4 Heterogeneous impacts**

Our large sample will also allow us to do a range of sub-group analyses. We will compare the effects of the program on all outcomes between i) hosts and refugees, ii) cities and camps, and iii) across the four study sites. In addition, we will study the following dimensions of heterogeneity:

- For households: i) by gender of the beneficiary, ii) by whether recipients were married/cohabiting, iii) by recipient age, iv) by whether children lived in the household v) by whether households have below median household consumption (relative to their camp/city) and PMT vulnerability assessment, vi) by whether the household had a self-employed business at baseline, and vii) by whether the household was employed at baseline, viii) education/baseline language skills (using years of education) iv) baseline integration (using number of close friends from outgroup).

<sup>10</sup> Randomization inference allows us to test the Fisherian sharp null hypothesis that  $y(t)_{ic}=y(c)_{ic} \forall t$  for every unit  $iv$ . We will calculate exact p-values for all our main effects of interest (and all main research questions) under the null hypothesis using a Fisher permutation test, taking 10,000 permutations of treatment assignment and roll-out exactly as it was implemented in the actual intervention.

- For enterprises: i) by sector: agriculture and livestock, manufacturing, services, non-food retail, food retail, ii) by whether the enterprise existed at baseline, iii) by baseline enterprise revenue (relative to other enterprises in their city/camp and sector)
- For prices: i) by sector: food, livestock, non-food durables, non-food nondurables, services, ii) tradables and non-tradables.<sup>11</sup>

### 5.5 Balance and Attrition

To assess whether the attrition of households between baseline and Endline 1 surveys confounds our results, we conduct the following analyses.<sup>12</sup> Let  $r_{ic}$  be an indicator for whether individual  $i$  in cluster  $c$  is observed at baseline but not at Endline 1. First, Equation (4/4a) estimates whether the magnitude of attrition varies with treatment status, and we follow equation (1) for our standard errors:

$$r_{ic} = \beta_0 + \beta_1 T_i + \beta_2 HiSat_c + \epsilon_{ic} \quad (4)$$

$$r_{ic} = \beta_0 + \beta_1 T_i + \gamma_1 MixedGroup_i + \beta_2 HiSat_c + \gamma_2 HiMix_c + \epsilon_{ic} \quad (4a)$$

Second, Equation 5 assesses whether observation status varies with a vector of baseline characteristics  $X_{ic}^{BL}$  (including for all primary outcomes observed at baseline, and all pre-defined dimensions of heterogeneity):

$$r_{ic} = \omega X_{ic}^{BL} + \epsilon_{ic} \quad (5)$$

If we find worrying levels of differential attrition, we will adjust for potential bias by bounding our parameter of interest following Lee (2005) or newer developments in this area.

---

<sup>11</sup> Besides pre-specified dimensions of heterogeneity, we may also use machine learning-based approaches to estimate heterogeneous impacts. For instance, we may generalize random forests to data from eligible households in order to predict treatment effects. We note that this is an active area of research and new tools continue to be developed; we expect to employ approaches such as Chernozhukov et al. (2019) and Athey et al. (forthcoming) but will adopt new approaches if they become standard or offer other desirable properties for our setting.

<sup>12</sup> For Dabat, we do not have a baseline. We will instead conduct the same analysis for attrition, i.e. whether sampled households were differentially likely to be reached and surveyed across treatment groups. For balance and Equation 5, we have to restrict ourselves to immutable demographic characteristics, and pre-treatment administrative data in Dabat.

## REFERENCES

- Allport, Gordon Willard, Kenneth Clark, and Thomas Pettigrew, "The nature of prejudice," 1954.
- Altındağ, Onur and Stephen D O'Connell, "The short-lived effects of unconditional cash transfers to refugees," *Journal of Development Economics*, 2023, 160, 102942.
- 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.
- Aygün, Aysun Hızıroğlu, Murat G Kirdar, Murat Koyuncu, and Quentin Stoeffler, "Keeping Refugee Children in School and Out of Work: Evidence from the World's Largest Humanitarian Cash Transfer Program," Available at SSRN 4024053, 2022.
- Balboni, Clare, Oriana Bandiera, Robin Burgess, Maitreesh Ghatak, and Anton Heil, "Why do people stay poor?," *The Quarterly Journal of Economics*, 2022, 137 (2), 785–844.
- Baranov, Victoria, Lisa Cameron, Diana Contreras Suarez, and Claire Thibout, "Theoretical underpinnings and meta-analysis of the effects of cash transfers on intimate partner violence in low-and middle-income countries," *The Journal of Development Studies*, 2021, 57 (1), 1–25.
- Barrington, Linda, Kenneth Troske et al., "Workforce diversity and productivity: An analysis of employer-employee matched data," *Economics Program Working Papers*, 2001, (01-02).
- Bastagli, Francesca, Jessica Hagen-Zanker, Luke Harman, Valentina Barca, Georgina Sturge, and Tanja Schmidt, "The impact of cash transfers: a review of the evidence from low-and middle-income countries," *Journal of Social Policy*, 2019, 48 (3), 569–594.
- Becker, Sascha O and Andreas Ferrara, "Consequences of forced migration: A survey of recent findings," *Labour Economics*, 2019, 59, 1–16.
- Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli, "Adaptive linear step-up procedures that control the false discovery rate," *Biometrika*, 2006, 93 (3), 491–507.
- Buller, Ana Maria, Amber Peterman, Meghna Ranganathan, Alexandra Bleile, Melissa Hidrobo, and Lori Heise, "A mixed-method review of cash transfers and intimate partner violence in low-and middle-income countries," *The World Bank Research Observer*, 2018, 33 (2), 218–258.
- Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins, "Double/debiased machine learning for treatment and structural parameters," 2018.
- Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker, "General equilibrium effects of cash transfers: experimental evidence from Kenya," *Econometrica*, 2022, 90 (6), 2603–2643.
- Erhardt, Niclas L, James D Werbel, and Charles B Shrader, "Board of director diversity and firm financial performance," *Corporate governance: An international review*, 2003, 11 (2), 102–111.
- Fisher, Ronald A, "The logic of inductive inference," *Journal of the Royal Statistical Society*, 1935, 98 (1), 39–82.
- Franklin, Simon, Clement Imbert, Girum Abebe, and Carolina Mejia-Mantilla, "Urban public works in spatial equilibrium: Experimental evidence from Ethiopia," *Technical Report* 2023.
- Gassmann, Franziska, Ugo Gentilini, Julieta Morais, Conrad Nunnenmacher, Yuko Okamura, Guilio Bordon, and Giorgia Valleriani, "Is the Magic Happening?: A Systematic Literature Review of the Economic Multiplier of Cash Transfers," 2023.

Giannetti, Mariassunta and Mengxin Zhao, “Board ancestral diversity and firm-performance volatility,” *Journal of Financial and Quantitative Analysis*, 2019, 54 (3), 1117–1155.

Grund, Christian and Niels Westergaard-Nielsen, “Age structure of the workforce and firm performance,” *International Journal of Manpower*, 2008, 29 (5), 410–422.

Hamilton, Barton H, Jack A Nickerson, and Hideo Owan, “Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation,” *Journal of Political Economy*, 2003, 111 (3), 465–497.

Hanna, Rema and Benjamin A Olken, “Social Protection in Low and Middle-Income Countries,” J-PAL-EPoD Social Protection Initiative Evidence Review, 2023.

Hjort, Jonas, “Ethnic divisions and production in firms,” *The Quarterly Journal of Economics*, 2014, 129 (4), 1899–1946.

Hoogendoorn, Sander and Mirjam Van Praag, “Ethnic diversity and team performance: A field experiment,” 2012.

Ibáñez, Ana María, Andres Moya, María Adelaida Ortega, Sandra V Rozo, and Maria Jose Urbina, “Life out of the shadows. impacts of amnesties in the lives of refugees,” *Policy Research Working Paper*, 2022, 9928.

Jr, Albert A Cannella, Jong-Hun Park, and Ho-Uk Lee, “Top management team functional background diversity and firm performance: Examining the roles of team member colocation and environmental uncertainty,” *Academy of management Journal*, 2008, 51 (4), 768–784.

Kilduff, Martin, Reinhard Angelmar, and Ajay Mehra, “Top management-team diversity and firm performance: Examining the role of cognitions,” *Organization science*, 2000, 11 (1), 21–34.

Kurtulus, Fidan Ana, “What types of diversity benefit workers? Empirical evidence on the effects of co-worker dissimilarity on the performance of employees,” *Industrial Relations: A Journal of Economy and Society*, 2011, 50 (4), 678–712.

Lee, David S, “Training, wages, and sample selection: Estimating sharp bounds on treatment effects,” 2005.

Leonard, Jonathan S and David I Levine, “Diversity, discrimination, and performance,” *Institute of Industrial Relations Working Paper No. iirwps-091-03*, 2003.

Lowe, Matt, “Types of contact: A field experiment on collaborative and adversarial caste integration,” *American Economic Review*, 2021, 111 (6), 1807–1844.

Miguel, Edward and Michael Kremer, “Worms: identifying impacts on education and health in the presence of treatment externalities,” *Econometrica*, 2004, 72 (1), 159–217.

Mousa, Salma, “Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq,” *Science*, 2020, 369 (6505), 866–870.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar, “Building state capacity: Evidence from biometric smartcards in India,” *American Economic Review*, 2016, 106 (10), 2895–2929.

Olney, Deanna K, Aulo Gelli, Neha Kumar, Harold Alderman, Ara Go, and Ahmed Raza, “Social assistance programme impacts on women’s and children’s diets and nutritional status,” *Maternal & Child Nutrition*, 2022, 18 (4), e13378.

Özler, Berk, Çiğdem Çelik, Scott Cunningham, P Facundo Cuevas, and Luca Parisotto, “Children on the move: Progressive redistribution of humanitarian cash transfers among refugees,” *Journal of Development Economics*, 2021, 153, 102733.

Perera, Camila, Shivit Bakrania, Alessandra Ipince, Zahrah Nesbitt-Ahmed, Oluwaseun Obasola, Dominic Richardson, Jorinde Van de Scheur, and Ruichuan Yu, “Impact of social protection on gender equality in low-and middle-income countries: A systematic review of reviews,” *Campbell systematic reviews*, 2022, 18 (2), e1240.

Peterman, Amber Roy et al., Cash transfers and intimate partner violence: A research view on design and implementation for risk mitigation and prevention, *Intl Food Policy Res Inst*, 2022.

Puteh, Fadilah and Wan Puspa Melati Wan Abdul Halim, “Localizing SDG in Eastern Region Malaysia: Key Challenges of Project Impact Evaluation,” *Journal of Technical Education and Training*, 2023, 15 (3), 212–223.

Rao, Gautam, “Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools,” *American Economic Review*, 2019, 109 (3), 774–809.

Scacco, Alexandra and Shana S Warren, “Can social contact reduce prejudice and discrimination? Evidence from a field experiment in Nigeria,” *American Political Science Review*, 2018, 112 (3), 654–677.

Timu, Anne G and Berber Kramer, “Gender-inclusive,-responsive, and-transformative agricultural insurance: A literature review,” *Global Food Security*, 2023, 36, 100672.

UNHCR, “Global Trends: Forced Displacement in 2022,” Technical Report, The United Nations High Commissioner for Refugees 2023.

Verme, Paolo and Kirsten Schuettler, “The impact of forced displacement on host communities: A review of the empirical literature in economics,” *Journal of Development Economics*, 2021, 150, 102606.

### **Administrative Information**

**Funding:** We are grateful to the Forced Displacement Multi-Donor Trust Fund, Prospects, Knomad, IPA displaced livelihood initiative, and the Research Support Budget of the World Bank to provide financial support for this impact evaluation.

**Institutional Review Board (ethics approval):** This research project has been approved by Oxford's University Ethics Committee (approval number ECONCIA22-23-17).

**Declaration of interest:** The authors have no disclosures relevant to this study.

## Appendix A: RHISN Program Details

### *RHISN public works component*

The RHISN Phase II public work program spans 24 months, from January 2024 to December 2024, during which beneficiaries receive a monthly cash payment for their labor contributions. RHISN public work groups are of four kinds: host-only, refugee-only, refugee-and-host mixed groups with equal proportions, and refugee-minority mixed groups. A typical RHISN public work group consists of about thirty beneficiary households. In host-only groups, all members are host households, while refugee-only groups consist solely of refugee households. Equal mixed groups comprise about 15 hosts and 15 refugees, whereas in refugee minority groups, there are 10 refugees and 20 hosts.

*Table: Number of public work groups and beneficiary households in RHISN Phase-II locations*

	woreda	By group type								Total	
		Refugee only		Host only		Equal mixed		Refugee minority mixed			
		Groups	House holds	Groups	House holds	Groups	House holds	Groups	House holds	Groups	House holds
Impact evaluation (IE) sites	Asayita	30	634	59	1519	40	933	25	576	154	3662
	Bambasi	45	1318	66	1944	6	170	8	230	125	3662
	Dabat	22	601	42	1237	24	627	41	1197	129	3662
	Pugnido	45	1213	37	1069	24	648	41	1146	147	4076
	<b>IE Total</b>	<b>142</b>	<b>3766</b>	<b>204</b>	<b>5769</b>	<b>94</b>	<b>2378</b>	<b>115</b>	<b>3149</b>	<b>555</b>	<b>15062</b>
Non-IE sites	AwBaree-Sheder	10	701	12	874	20	1534			42	3109
	Kebribeyah	8	434	7	567	13	897			28	1898
	<b>Non-IE total</b>	<b>18</b>	<b>1135</b>	<b>19</b>	<b>1441</b>	<b>33</b>	<b>2431</b>			<b>70</b>	<b>5007</b>
	<b>Overall</b>	<b>160</b>	<b>4901</b>	<b>223</b>	<b>7210</b>	<b>127</b>	<b>4809</b>	<b>115</b>	<b>3149</b>	<b>625</b>	<b>20069</b>

Each household can have a maximum of four individuals benefiting from the program (if a household consists of four or more individuals), depending on the household size. A household with four eligible beneficiaries is required to work five days per week, whereas a household with only one beneficiary works for a total of five days per month. Each beneficiary works for 4 to 5 hours, usually in the morning, depending on the hardship of the location as determined by the program. Beneficiary households receive payments based on the number of days worked, with each workday compensated with 150 birr.<sup>13</sup> Consequently, a household with four eligible beneficiaries could potentially earn up to 3,000 birr monthly, while households with just one eligible beneficiary could receive 750 birr monthly. Local administrations, in consultation with the refugees and host representatives, identified specific projects for public works. These projects involve tasks such as street cleaning, dry waste collection, urban greenery, and small-scale infrastructure (such as building latrines).

### *RHISN livelihood grant*

RHISN not only offers labour-intensive public work opportunities for beneficiaries but also supports them in business development and provides livelihood grants to move them into sustainable livelihoods. The livelihood program is phased in with the public works by providing training and coaching to support beneficiaries to enter self-employment and wage employment. Beneficiaries will undergo life skills and financial literacy training twice a month for the first five months and monthly for an additional 11 months, from April 2024 to July 2025. Then, participants will select their livelihood path (self-employment or wage-employment) between June and July 2024, and receive business plan and job search plan training in December 2024. Technical skill training follows from December 2024 to February 2025. At the end of the public work cycle and upon successful completion of training, each beneficiary household will receive a US\$600 grant in two installments to either start a business or to enter wage employment. The first installment grant (80%) will be disbursed in January 2025 and the second (20%) in May 2025. Throughout the program, from April 2024 to December 2025, participants will receive group coaching and mentoring sessions.

<sup>13</sup> Based on the February 2024 official exchange rate, 150 birr is equivalent to USD 2.7.

## Appendix C: Market price survey details