Journal of Development Economics

What's in the Name? Volunteer and Employment Opportunities in Egypt

--Manuscript Draft--

Manuscript Number:	DEVEC-D-24-00836R1
Article Type:	Registered Report Stage 1: Proposal
Section/Category:	Insurance, Agriculture, Land, Credit, poverty, Africa - Micro, households, gender
Keywords:	Volunteering, employment, female labor force participation, female empowerment
Corresponding Author:	Marieke Kleemans UNIVERSITY OF ILLINOIS Urbana, UNITED STATES
First Author:	Brian Feld
Order of Authors:	Brian Feld
	Marieke Kleemans
	Abdel Rahman Nagy
Abstract:	Female labor force participation in Egypt is low, partly due to social norms regarding female employment. While in our study area only 17% believe that women should be allowed to be employed, 52% support women working as volunteers even though volunteers receive financial compensation. We will visit 7,500 households and encourage women to apply for work opportunities that are either referred to as "employment" or as "volunteering" but otherwise have identical characteristics. In addition to randomizing the name of the opportunity across 500 geographically distant agglomerations, we will randomize at the household level whether we discuss the opportunity with women only or whether we also involve other household members. After comparing application behavior across these four treatment arms, we will randomize work opportunities among applicants. This will allow us to estimate the impact of working on the applicants and their families, and compare work performance and satisfaction across treatment arms.
Response to Reviewers:	

Journal of Development Economics
Registered Report Stage 1: Proposal

What's in the Name?
Volunteer and Employment Opportunities in Egypt

January 2025*

Abstract
Female labor force participation in Egypt is among the lowest in the world, due, in part, to social norms that restrict what activities women can do outside of their homes. While in our study area only 17% believe that women should be allowed to work as an employee, 52% are supportive of women working as a volunteer even though volunteers receive financial compensation. In this project, we will visit 7,500 households and encourage women to apply for work opportunities that are either referred to as "employment" or as "volunteering" but otherwise have the same job characteristics, including compensation, hours and contract. In addition to randomizing the name of the work opportunity across 500 geographically distant agglomerations, we will randomize at the household level whether we discuss the opportunity with the woman only or whether we also involve other household members in order to better understand household decision making and how to alleviate constraints to women's work out of the home. After comparing application behavior across these four treatment arms, we will randomize work opportunities among applicants. This will allow us to estimate the impact of working on the female applicants and their families, and compare work performance and satisfaction across treatment arms.

Keywords: Volunteering, employment, female labor force participation, female empowerment

JEL codes: O12, J22, I32, J24, J16

Study pre-registration: We have registered this study in The American Economic Association's registry for randomized controlled trials: https://www.socialscienceregistry.org/

* We are grateful to Andrew Garin, Jules Gazeaud, Greg Howard, Suzanna Khalifa, Sikandra Kurdi, Greg Lane, Jeremy Magruder, Paul McNamara, Adam Osman, Lena Song, Rebecca Thornton, and Andrew Weaver for their helpful comments and suggestions. We are grateful to JPAL MENA, Life Makers, and the Sawiris Foundation for Social Development for the implementation of this project. We especially thank Youssr Abbas, Shaimaa Abdelrazik, Nada Alaa, Maha Balbaa, Hoda El-Abbadi, Sara El-Sarrag, Ahmed Elsayed, Amira El-Shal, Mohamed Gamal, Mahmoud Hussein, Hanan Khayal, Marina Maged, Mai Moataz, Fatma Mosaad, Sara Ragab, Rafal Ramadan, Mohamed Sayed, Nourhan Soliman, Nahed Yousry, and Mohamed Zanati for their commitment and support, and we are grateful to David Buller for excellent research assistance. We thank Jules Gazeaud, Reshmaan Hussam, Erin Kelley, Gregory Lane, and Fatima Zahra for generously sharing survey instruments. All errors remain our own.

Proposed timeline

The qualitative study including target group discussions was conducted in Oct 2021 – Dec 2021.

The pilot study was conducted in Oct 2022 – Apr 2023.

The study area was divided into agglomerations from Nov 2023 to Apr 2024.

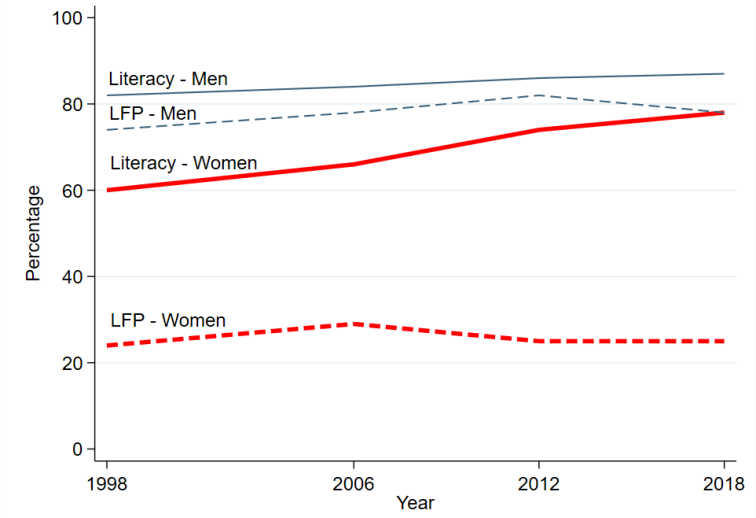
The main project is scheduled for 2.5 years from Mar 2025 to Aug 2027.

- Project implementation during household visits including the baseline survey will take place on a rolling basis for the first 1.5 years from Mar 2025 – Aug 2026.
- Midline surveys will be conducted 4 months after the baseline survey, so from Jul 2025 – Dec 2026.
- Endline surveys will be conducted 1 year after the baseline survey, so from Mar 2026 – Aug 2027.

1. Introduction

Countries in the Middle East and Northern Africa have made significant progress towards achieving gender parity in education over the past decades, but women's labor force participation continues to lag behind. Figure 1 illustrates this pattern using data from the Egyptian Labor Market Panel Survey (Krafft, Assaad, and Rahman, 2021) showing low and stagnant levels of female labor force participation of approximately 25%.[2] Female labor force participation is even lower in more disadvantaged areas, including Sohag, the second poorest Governorate in Egypt and the setting for this project, where, according to our pilot data, less than 12% of women are employed outside of their home. In contrast, more than 50% of the female population aged 15 and above participate in the labor force in low- and middle-income countries in Sub-Saharan Africa, East Asia and the Pacific, Europe and Central Asia, and Latin America and the Caribbean (World Bank, 2013).

Figure 1. Gender gap in literacy and labor force participation in Egypt



Notes: Labor force participation (LFP) is measured as the number of people aged 15 to 55 who are employed or unemployed over the total number of people aged 15-55.
Source: Egyptian Labor Market Panel Survey (Krafft et al, 2021).

There are various complementary explanations for the disproportionately low female labor force participation in the Middle East and Northern Africa. In addition to well-studied constraints such as labor market discrimination, limited availability of adequate work opportunities, and lack of support systems such as childcare, recent evidence has underscored the importance of social norms that consider female work outside of their home socially unacceptable (Bursztyn, González, and Yanagizawa-Drott, 2020); Osman, Speer, and Weaver, 2024; Assaad, Hendy, Lassassi, and Yassin, 2020).[3] Indeed, in our context, social norms appear to play an important role in limiting women's work out of the home. In target group discussions we carried out, social norms and child care responsibilities were mentioned as the main reasons why women do not work outside of their homes. Indeed, data from our pilot study shows that amongst women surveyed, 83% believe that

[2] The most recent data from the International Labour Organization indicates that female labor force participation had dropped to 15.7% in 2021 partially due to the COVID 19 pandemic (ILO, 2023). In contrast, male labor force participation in that same year remained high at 70.7%.

[3] In line with Bénabou and Tirole (2011), we consider social norms to be the set of social sanctions or rewards that promote specific behaviors.

women should not be allowed to work, and 70% agree that women who work cannot take care of their children well.

In stark contrast to views towards women being employed outside of their homes, views are much more favorable towards women working as volunteers: while only 17% of pilot study respondents thought that women should be allowed to work, 52% was supportive of women working as a volunteer. Indeed, Life Makers, a large nationwide NGO in Egypt and implementing partner in this study, relies heavily on female volunteers to help implement their projects. This contrast is surprising because volunteers do similar tasks as employees, and – unlike common practice in high-income countries – volunteers in Egypt adhere to a contract and prespecified work schedule, and, importantly, they receive financial compensation for their work. On paper, this compensation is meant to cover transportation and meal costs, but in practice, the amount far exceeds these costs and often exceeds minimum wage. Moreover, volunteers work outside of their home.

Our target group discussions suggested that this discrepancy between employment and volunteering can partly be understood by diverging social norms towards these two activities: if a woman is employed, the family can be perceived as poor and in need of additional income, whereas if she volunteers the family is well-regarded in the society, both from a religious and economic point of view because she can spend time doing charitable activities. Similarly, if a woman engages in employment, her husband could be perceived as not having sufficient control over her, whereas if she volunteers her husband could be seen favorably by allowing her to do so.

Motivated by these patterns, we will offer opportunities for women to work as either an employee or volunteer in this study. Other than the name of the work opportunity, we will keep everything else constant, including compensation, contract, work schedule, and activities and we will work in geographically distinct areas that are randomized into "employment" and "volunteering" agglomerations.[4] All work opportunities will be offered by our implementing partner, Life Makers, who aims to recruit 900 women to carry out a large development project in Sohag. Our research study centers around the recruitment of these workers. This NGO works with both employees and volunteers to implement their projects, making this a natural setting to randomly vary the form of contracting between "volunteers" and "employees."

In addition to the importance of the contract type, our pilot study suggests that household dynamics play an important role in the decision to apply for a work opportunity. When encouraged to apply during the pilot, 30% of the eligible women stated that they didn't want to apply to work for the NGO because they believed their father or husband would not approve of it. Therefore, in addition to varying the name of the work opportunity, we aim to further understand the household decision-making process by randomizing whom we target in our recruitment process. In our "Women Only" treatment arm, we will visit a woman by herself, providing her with informaion and encouraging her to apply. In the "Women Plus" treatment arm, we will instead encourage her to invite other household members to be present because they may have influence over her decision to apply for the work opportunity. This randomization will be carried out at the household level and will be implemented in both the Volunteering and Employment agglomerations. As such, there will be 4 treatment arms: 1. Volunteering with Women Only, 2. Volunteering with Women Plus, 3. Employment with Women Only, and 4. Employment with Women Plus.

[4] Henceforth, we use the term "work" to describe either "volunteering" or "employment".

For this project, we plan to visit 7,500 households living in 500 small and geographically distinct agglomerations of 15 households each. 250 agglomerations will be randomly assigned to the Volunteering treatment arms and 250 to the Employment treatment arms. In all agglomerations, households will have 50% chance of being assigned to the Women Only and Women Plus treatments. Based on our power calculatins and data from our pilot study, this will allow us to answer the following three research questions.

Our first research question focusses on application behavior. We will study whether there are differences in application behavior for Volunteering and Employment and for Women Only and Women Plus. As such, we aim to increase our understanding of how women can be encouraged to work out of their home. While Volunteering may be more socially acceptable, Employment may be perceived as more economically advantageous, and these patterns may differ by characteristics of the household and woman. Similarly, while involving additional household members in Women Plus households may increase understanding and support, the most direct test to date by Lowe and McKelway (2024) shows that, in India, involving husbands and encouraging discussion actually decreased application rates for a female employment opportunity. We will flexibly compare all four treatment arms and explore two-way comparisons given that there may be interactions between Volunteering versus Employment on the one hand and Women Only versus Women Plus on the other hand.

For our second research question, we aim to understand how working affects women and their families. Various studies have shown effects of working on household income, health, wellbeing, consumption, political participation, coping with shocks, subsequent labor market outcomes, and intimate partner violence (Bhanot, Han, and Jang, 2018; Blattman and Dercon, 2018; Breza, Kaur, and Shamdasani, 2021 ; Blattman, Dercon, and Franklin, 2022; Hussam, Kelley, Lane, and Zahra, 2022; Kotsadam and Villanger, 2022; Aalen, Kotsadam, Pieters, and Villanger, 2024). Similar to these studies, we will randomize work offers amongst the eligible women who decide to apply. We will estimate effects amongst those who are randomly selected to work in each of the four treatment arms on a variety of outcomes including consumption, household decision making, subsequent labor market outcomes, social norms, and physical and phychological wellbeing.

For our last research question, we will compare job performance and satisfaction amongst the four treatment arms. Even though the work opportunities are identical in terms of compensation, duration, hours, and tasks, women may exert effort differentially depending on the contracting type and depending on how much other household members were involved in the application process (Ashraf, Bandiera, Davenport, and Lee, 2020).

Finally, if we find differences in application behavior as part of research question 1, we will estimate how much women are willing to forego in income to comply with social norms and conversely, how much we need to compensate women to deviate from social norms. We will back out this compensating differential using the Becker-DeGroot-Marschak method. In a small sample separate from our main sample, we will ask women to provide the minimum compensation for which they would take a volunteering (employment) opportunity, and allow them to apply only if their “bid” is equal to or lower than a compensation unknown to them. By comparing the bids of women offered the volunteer position to those of women offered the employment position, we can

estimate how much more or less a volunteer would need to be compensated compared to an employee to have a similar application rate. This will give us an estimate of the potential cost posed by social norms.

We believe this project will add to various strands of literature. First, we contribute to the literature on female labor force participation and its constraints. Studies carried out in both developed and developing countries have found that female labor force participation increases when women can reduce the time they spend on childcare (Berlinski and Galiani, 2007; Berlinski, Galiani, and Mc Ewan, 2011; Cortés and Tessada, 2011; Romiti, 2018; East and Velásquez, 2024), elder care (Peri, Romiti and Rossi, 2015), and household chores (Coen-Pirani, León and Lugauer, 2010; Dinkelman, 2011). However, these types of interventions have not shown a similar impact in the MENA region and in Egypt in particular. A closely related ongoing study by one of our authors (Nagy) and colleagues finds low take-up of childcare and employment opportunities in Cairo (Caria, Crepon, ElBehairy, Fadlalmawla, Krafft, Nagy, Mottaghi, Zeitoun, and El Assiouty, 2022). They cross-randomized free or highly-subsidized nursery access and employment opportunities for mothers. They found 11 percent take-up of nurseries services but less than 0.3% take-up of employment, suggesting that labor demand is not the sole binding constraint. When exploring reasons for low take-up, social norms appeared to play a crucial role: both men and women had unfavorable views towards women being employed and towards sending children to a nursery.

Recently, researchers have stressed the importance of social, cultural, and religious norms in the process of economic development and gender equality (Antecol, 2001; Jayachandran, 2015; La Ferrara, 2019; Anukriti, Herrera-Almanza, Pathak, and Karra, 2020; Ashraf, Bau, Nunn, and Voena, 2020; Jayachandran, 2021; Dhar, Jain, and Jayachandran, 2022, Khalifa, 2022; McKelway, 2022; Bau and Fernández, 2023). These studies also provide evidence of the barriers that such norms can present to female empowerment and, specifically, female labor force participation. Experimental evidence of the effects of women's work has focussed on South Asia (Dean, and Jayachandran, 2019; Field, Pande, Rigol, Schaner, and Troyer Moore, 2021; Hussam, Kelley and Lane, 2023; Ho, Jalota, and Karandikar, 2024; Jalota and Ho, 2024; McKelway, 2024). Within the MENA region, social norms are central in Bursztyn et al (2020) who find evidence of pluristic ignorance towards support for women working outside the home in urban Saudi Arabia: most Saudi men privately support the idea of women working, but incorrectly assume that the majority of other men are against it. This leads to a situation where men publicly oppose women working outside the home because they want to conform to what they mistakenly perceive as the prevalent social norm in order to avoid social backlash. Bursztyn et al. (2020) then correct these beliefs by revealing the actual widespread support among men for women working outside the home. This intervention leads to a significant increase in men's willingness to assist their wives in job searching. Subsequently, their wives are more likely to apply and interview for jobs outside the home, highlighting the impact of perceived social norms on behavior. Our paper follows on Bursztyn et al. (2020) by recognizing that social norms may limit women's opportunities to work out of their homes and we explore two additional avenues to make work opportunities for women comply with social norms: presenting work as volunteering instead of employment, and involving additional household members in the decision making process. We then extend on existing work by also randomizing work opportunities amongst applicants, allowing us to look beyond application behavior and also study the impact of working out of the home under various social norm regimes on women and their families.

An established literature has pointed to various benefits of labor force participation, including recent work showing improvements on overall well-being (Hussam et al., 2022; McKelway, 2024). In the case of female employment in particular, research has shown that providing job opportunities to women improves their position in the household (Majlesi, 2016), and the outcomes of their children (Qian, 2008; Majlesi, 2016). Our study will contribute to this literature in two ways. First, we can test if alternative forms of work, such as volunteering, have similar effects on women. Second, given that the work opportunity in our study covers a fixed amount of time (6 months) and we will collect data while women are working and after the work opportunity has ended, we will estimate the persistence of the effects of female work.

Finally, we believe this project is policy relevant and has the potential to help tackle the region's social and economic challenge of low rates of female labor force participation by making work opportunities comply with prevailing social norms. This presents similarities with the literature on "Islamic microfinance" that shows that altering financial products to make them compliant with religious laws can increase take-up (El-Gamal, El-Komi, Karlan, and Osman, 2014; Karlan, Osman, and Shammout, 2021). Indeed, Sulaiman (2011) argues that "no religion or law gives volunteerism as much importance as Islam does" and that volunteering is vital for Muslim communities' growth and development. As such, in this study we hope to learn if and how volunteering can be used as a strategy to increase women's work out of the home, and allow them to take advantage of relatively high levels of education to benefit their families and broader communities.

The rest of this paper is structured as follows: in Section 2 we describe the setting where our intervention will take place, our partners, and the findings of our qualitative target group study and pilot study. In Section 3 we describe in detail our intervention, and Section 4 lays out our hypotheses. In Section 5 we describe the multiple rounds of data collection, our sample, and measures to deal with attrition and non-complaince. Finally, in Section 6 we discuss the estimation methodology and power calculations and Section 7 concludes.

2. Background

In this section, we describe the context of our study area in Egypt and the partners we will work with. Then, we describe the main findings of our qualitative target group study and pilot study that informed the design of our project.

2.1 Context and partners

There is a high prevalence of poverty in Egypt, especially in remote rural areas in Upper Egypt. Sohag Governorate, where the project will be implemented, has a poverty rate of 60%, the second highest in Egypt (CAPMAS, 2018).[5] Poor villages in Egypt suffer from many challenges that make their residents marginalized and vulnerable, including lack of infrastructure, low levels of education, high illiteracy rates, and difficulty accessing basic services such as healthcare. Women

[5] Governorates are the first administrative divisions of Egypt.

face additional challenges, including a lack of work opportunities and little autonomy over their life due to the subordinate role of women in society (El Feki, Heilman, and Barker, 2017).

With the support of the Sawiris Foundation for Social Development, the NGO, Life Makers, is implementing a comprehensive development project to improve the livelihoods of the people in Sohag and to enhance the social role of women at the family and community level. As part of this project, Life Makers will need to recruite 900 women to implement various components of the project, including adult literacy classes and an awareness campaign about women's issues in the region. Our intervention will be implemented in the context of this NGO's recruitment of workers to conduct these activities.

Our implementing partner, Life Makers, is a nation-wide NGO that was established in 2011. Life Makers works directly through its offices and branches across Egypt to promote sustainable development, education, health, livelihoods, and basic needs. Our research partner is the Abdul Latif Jameel Poverty Action Lab, J-PAL, and we work closely with their Middle East and North Africa in Cairo. Following many years of collaboration, JPAL will work with the survey firm "Promising Egypt" for survey data collection.[6]

2.2 Target group study

Based on initial discussions with our partners dating back to May 2021, we conducted a qualitative study in Sohag from October to December of 2021. We conducted 12 target group discussions with approximately 120 men and women in separate groups. Our main focus was on female empowerment and, specifically, female employment and volunteering.

We learned that both men and women have rather unfavorable views towards women working out of their homes. As reasons for their hesitancy, they cited the abundance of women's home and childcare duties and fear of backlash from other members of the community. Women clarified that they would need permission from their husband or father and many expressed reluctancy to ask them out of fear of rejection. They explained that such decisions are often made at the household level. Both men and women seem to be more approving of women working as volunteers than as employees.

When asked about volunteering, many women expressed interest as long as they know the NGO, they work with trusted personnel, and work mostly with other women. Interest was espeically high among unmarried women still living with their parents and among married women with older children. We learned that women with high school certificates or university degrees tend to volunteer more and many women believed volunteering can improve their skills and help them with their subsequent search for employment. Similarly, various men said they would allow their wives to work as a volunteer outside the house in nearby areas and with trusted individuals on a part-time basis, such as two or three days per week.

[6] https://lifemakers.org/ and https://www.povertyactionlab.org/middle-east-and-north-africa

2.3 Pilot study

Motivated by our target group discussions, we conducted a pilot study from October 2022 to April 2023 in the village of Al Gabab in Tama district in Sohag.[7] Given that our primary goal was to learn how to best organize the recruitment process and given that we didn't want to use up too much of the sample area, we decided to only pilot the recruitment of volunteers, not employees. Our pilot consisted of five consecutive stages, allowing us to test if women who were recruited in one stage were able to recruit women for the subsequent stage. The NGO used the stages to learn what kind of activities they can ask women to do in the main project. These tasks are described in more detail in Section 3 and Appendix B and have been tested successfully in the pilot.[8]

We visited 516 households and conducted detailed surveys with 282 women of whom 169 were deemed eligible to work for the NGO.[9] Amongst those, 86 expressed interest in the volunteering position, which is about 50%.[10] Based on an open-ended question, the pilot showed that the reasons why women were not interested in the work opportunity were mainly that they did not want to work, that their families did not allow them, and that they had too many responsibilities at home.

Table 1 presents summary statistics for the women we interviewed. Women in our sample are relatively young (the average age is 35 years), and more than three quarters of them are married. Among those married, they have on average 2.26 children and a third of women lives with their mother-in-law. Women have on average 7 years of education and 70% of them are literate.

Only 28% of eligible women has ever worked and 11.7% of those women are currently working. Amongst those currenlty working, they work 26.7 hours per week and earned 1424 EGP in the past month, which is equivalent to 55 USD. Only 6.7 percent of women has ever volunteered and none of them are currently volunteering so no data on hours and income from volunteering was collected. Women spend on average 4.3 hours a day on household chores and 2 hours taking care of children.

Drawing extensively on lessons from our target group discussions and pilot study, we will now describe the experimental design of our project.

---

[7] This village will be excluded from the project sample.

[8] Based on the pilot, we decided to delete one task: data collection. We had initially intended to have the recruited workers visit households by themselves to first administer the baseline survey and, afterwards, recruit eligible women to also work for the NGO. We learned that not all workers felt comfortable visiting new households by themselves. Therefore, and also because we aim to achieve the highest-quality data, we decided to have professional enumerators from our survey firm collect all survey data. Some of our recruited workers will accompany enumerators during the baseline data collection and be responsible for the recruitment process, as described in more detail in Section 3.

[9] As explained in more detail in the next section, work opportunities are available only for women between the ages 18 and 55 who are literate and present at the time of the survey. During the pilot, the NGO experimented with additional educational requirements which reduced the number of eligible women, but the NGO decided that educational requirements are not needed for the main project.

[10] To be conservative, our power and sample size calculations in Sections 5 and 6 allows for lower applications rates, even though the financial compensation and recruitment intensity will be higher in the main project than it was in the pilot. As detailed in Sections 5 and 6, we aim to visit 7,000 eligible women in order to fill 900 positions.

3. Research design

In this section we describe our intervention, including the treatment arms, the criteria for women to be eligible to work for the NGO, and a description of how eligible women will be recruited.

3.1 Intervention

From November 2023 to April of 2024, our research partner, JPAL, mapped the entire study area in Sohag Governorate that has a population size of approximately 400,000. We divided the area into 500 agglomerations of 15 households each, with large "buffers" of homes in between them to avoid contamination. Agglomerations will be randomly assigned to one of the following two groups: 250 agglomerations will be assigned to the Volunteering group and 250 agglomerations will be assigned to the Employment group.

Table 1: Summary statistics from pilot study

Age	35.07
	(11.06)
Percent literate	69.13
	(46.24)
Years of education	6.85
	(5.85)
Percent married	75.99
	(42.75)
Percent separated/divorced	2.92
	(16.84)
Percent widow	4.97
	(21.76)
Number of children (if married)	2.26
	(1.67)
Mother-in-law lives in the household (if married)	32.96
	(47.06)
Time spent on household chores per day (hours)	4.31
	(3.00)
Time spent taking care of children per day (hours)	2.04
	(2.31)
Ever volunteered (percent)	6.74
	(25.11)
Ever employed (percent)	28.01
	(44.99)
Currently employed (percent) if ever employed	11.70
	(32.20)
Hours worked in the previous week, if currently employed	26.73
	(23.86)
Income in the previous month, if currently employed (EGP)	1424.10
	(1380.63)
Number of eligible women interviewed	169
Expressed interest in the volunteering position offered (percent)	86

Note: Survey data from pilot study carried out in Al Gabab, Tama, Sohag from October 2022 to April 2023. 516 housheolds were visited with 583 women in total. Data on employment, volunteering and time use was collected for 282 women only who were adminstered a more detailed survey. During the pilot, the exchange rate was on average 26 EGP/USD.

At the start of the project, our survey firm will visit all households in the sample to conduct the baseline survey. This survey includes a household roster and for each member collects data on their gender, age, and literacy that jointly determine eligibility to work for the NGO. Work opportunities are available only for women between the ages 18 and 55 who are literate and present at the time of the baseline survey visit. The survey also includes questions about household assets, consumption, household decision making, employment, volunteering, social norms, and wellbeing as detailed in Section 4.[11]

At the end of the survey, eligible women will be asked if they are interested in learning more about a work opportunity with the NGO. If so, they will be called by a recruiter from the NGO a few days after survey data collection to schedule a recruitment visit. During the recruitment visit, women will be asked if they are interested in applying for the work opportunity.

The name of the work opportunity will depend on the treatment arm assigned to the agglomeration:

- Volunteering agglomerations: women will be offered to apply to become volunteers for the NGO.
- Employment agglomerations: women will be offered to apply to become employees for the NGO.

All the characteristics of the work opportunity, including compensation, hours of work, work schedule, and tasks will be held constant across Volunteering and Employment treatment arms. All work opportunities will last six months, and women will work 12 days per month for, on average, 5 hours each day. The compensation for the work opportunities is around 2500 Egyptian pound (EGP) per month, which is approximately 50 USD and quite sizable for most households. The corresponding hourly rate of 42 EGP or 0.80 USD is higher than the minimum wage.[12] The compensation is also higher than wages received for those currently employed, see Table 1. For each training day attended, women will be compensated at a rate of 120 EGP.

In addition, households will be randomly assigned to one of two groups:

- "Women Only": the work opportunity will be discussed only with the woman during the recruitment visit.
- "Women Plus": the woman will be explicitly encouraged to invite others to attend the recruitment visit, in order to involve them in the decision to apply for the work opportunity.

[11] The baseline survey can be found here.

[12] For comparison, the minimum wage for a fulltime worker in the private sector is 3000 EGP per month (Ahramonline, 2023a) and for government workers 3500 EGP (Ahramonline, 2023b).

The randomization into Women Only and Women Plus will take place at the household level in both Volunteering and Employment agglomerations.[13] As such, and as shown in Figure 2, there are 4 treatment arms:

1. Encourage Women Only to work as Volunteers
2. Encourage Women Plus to work as Volunteers
3. Encourage Women Only to work as Employees
4. Encourage Women Plus to work as Employees

Survey data will be collected by professional enumerators. Section 5 below will discuss the data collection in detail. The baseline survey visit will be conducted by an enumerator and the recruitment visit will be carried out by a recruiter from the NGO. The recruiter will explain the work opportunity and show a brief video. At the end of the recruitment visit, women decide whether to apply. The recruitment process is detailed in Section 3.2 below. During both the baseline and recruitment visit, women are told that the opportunity is for 6 months only with no possibilities of extension, and we emphasize that opportunities will be offered to those randomly selected to receive it, among the women who apply.

Within the Volunteering and Employment treatment arms, we will randomly choose which women will be offered to work for the NGO amongst those who applied. This randomization will take place at the household level, such that all the women who applied from the same household will either all be offered the work opportunity or not. Figure 2 gives an overview of the experimental design including the different randomization levels and treatment arms.

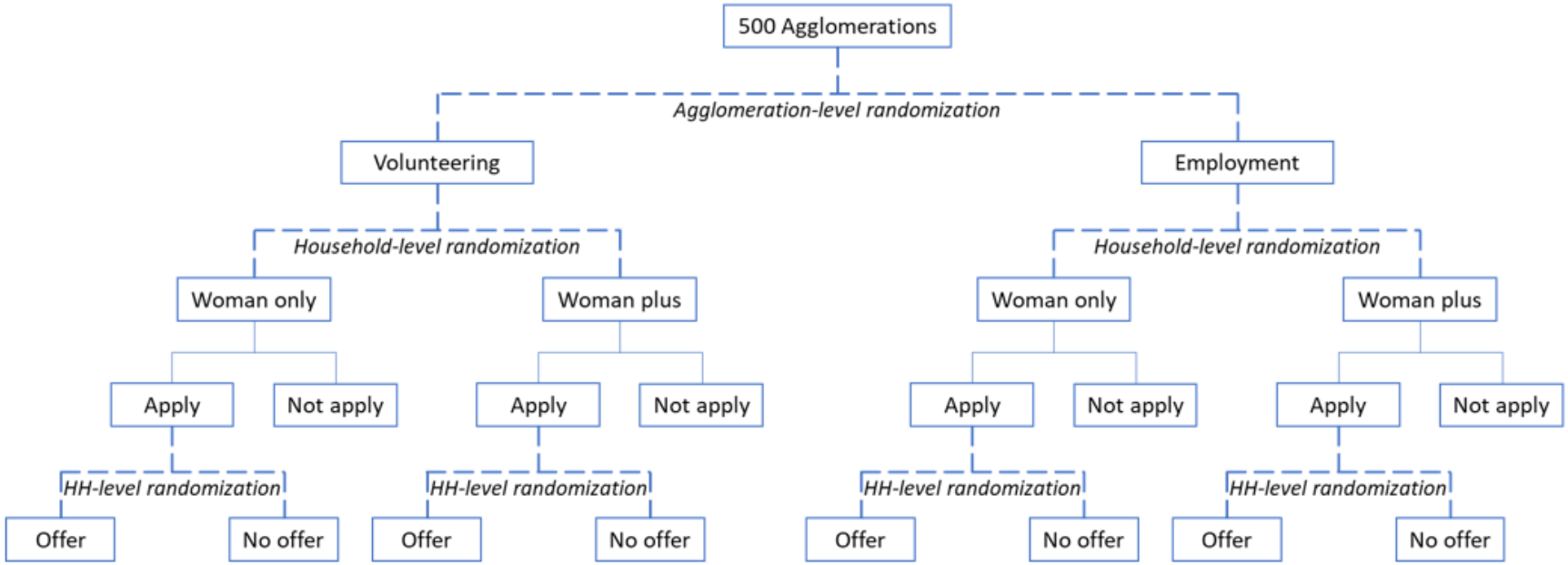
When calling women who have been selected to receive the work opportunity, the NGO will again emphasize that the work opportunities were assigned randomly. All work opportunities start with a multiple-day training. Volunteers and employees will receive training at different times and will work in geographically distinct areas, to minimize the probability that the two groups know about each other's existence. Appendix B lists the activities that women recruited to work for NGO will perform. These will be assigned by the NGO based on their needs and on the woman's education level, and this assignment will be done independently of the treatment arm.

3.2 Recruitment

As part of this project, we aim to recruit 900 women to work for the NGO. In order to do so, we plan to visit 7,500 households in Sohag Governorate to reach 7,000 eligible women. Please refer to Section 5.2 for more details on the sample size. In this section, we describe the recruitment process of our intervention.

[13] Note that randomization will take place at the household level and not at the individual level, so if there are multiple eligible women in a household, they will either all be assigned to Women Only or to Women Plus.

Figure 2: Experimental design



Note: Dashed lines indicate random assignment. Randomization into Volunteering and Employment will take place at the agglomeration level. Randomization into Women Only and Women Plus, and into Offer and No offer will take place at the household level.

First, a female enumerator from a survey firm (overseen by our research partner, JPAL) will visit households to conduct our baseline survey. The baseline survey includes a household roster with 3 eligibility questions for each household member. Members are eligible if they are (1) female, (2) between ages 18 and 55, and (3) present at the time of the survey. We will collect household-level data from the main respondent and individual data from all eligible women in the household.[14] For the individual part of the survey, the enumerator aims to survey eligible women without bystanders.

After administering the baseline survey, the enumerator will give a general introduction about the work opportunities available with the NGO.[15] Depending on the treatment arm, the enumerator will refer to the work opportunity as either "volunteering" or "employment". The enumerator will mention that work opportunities will be offered randomly among the pool of applicants. The enumerator will ask if the eligible woman is interested in learning more about the work opportunity. If so, she explains that a recruiter from the partner NGO will set up and conduct visit at a later date to provide additional information. We refer to this household visit as the "recruitment visit". If the household is randomized in the Woman Plus treatment, the enumerator encourages her to invite others during the recruitment visit explaining that inviting others may help gather support for her to apply for the work opportunity.[16] The recruiter then asks if she wants to invite others and if so, whom.

A few days after the baseline survey is conducted, a recruiter from our partner NGO calls the eligible women to follow-up on her interest in the work opportunity and asks when is a convenient time to visit to provide further details on the work opportunity.[17] Similar to the enumerator, the recruiter will refer to the work opportunity as either "volunteering" or "employment". If this is a Women plus treatment arm, the recruiter will ask if she still intends to invite the invitees she mentioned during the baseline survey and then asks if she wants to invite others. She will explain again that inviting others may help gather information and support for her decision.[18] The recruiter encourages the women to check if the invitees are available during the recruitment visit. Soon before the recruitment visit, the recruiter will briefly call all women again merely to remind them about the date and time of the recruitment visit. The goal of this reminder phone call is to help ensure that the eligible woman is home and help build trust in the NGO. Furthermore, for Women Plus households, the reminder phone call gives us an additional opportunity to encourage women

[14] The main respondent is the female spouse of the household head or the household head if she is female. If that person is not present or does not want to talk, the enumerator asks to speak with any woman aged 18 or older who lives in the household and is willing to answer questions.

[15] If there is more than one eligible woman in the household, the recruiter addresses all eligible women.

[16] However, we will make it clear that only she is allowed to apply for the position. Other women whom she may invite to the visit are not allowed to apply unless they are visited themselves.

[17] If more than one woman in a household is eligible, but only one of them expressed interest in the work opportunity during the baseline survey, the recruiter will call only the interested woman and during the recruitment visit the recruiter will direct the encouragement message only to her (and those she invited if the household belongs to the Women Plus treatment arm). If multiple eligible women in a household are interested, the recruiter will aim to visit the women simultaneously, and if that is not possible, the recruiter will schedule separate follow-up visits with all interested eligible women in the household.

[18] The exact wording that the recruiter will use is: "I highly encourage you to invite others to attend our follow-up meeting so that they can also learn about the volunteering position and help you decide if you'd like to apply. You can invite anyone you want, and as many people as you want."

in the Women Plus treatment arm to invite others and to make sure that those invited will attend the recruitment meeting.

To ensure high compliance with the Women Plus treatment, we ask women twice who she wants to invite for the recruitment visit: 1. At the end of the baseline data collection and 2. During the subsequent phone call setting up the recruitment visit. When asked who she wants to invite during this phone call, we prepopulate the survey with her answer from the baseline survey and ask if she still intends to invite each person. Similarly, during the reminder phone call, we mention who she had intended to invite so far. During both phone calls, we encourage her to make sure the invitees are available to attend the recruitment meeting and we emphasize that she can invite others too. Finally, at the start of the recruitment visit, we check with the list of invitees to see if everyone is present. If an invitee is at home (or working outside close to the home) but not yet attending, we ask if they will attend and we propose to wait until everyone has arrived .

The recruiter will conduct the recruitment visit by herself, providing details of the work opportunity and encouraging eligible women to apply. The recruiter will once again remind the woman that work opportunities will be offered randomly among those who apply. In case the household belongs to the Women Only treatment arm, the recruiter will explicitly and firmly request at the start of the visit to be alone with the women.[19] If the household belongs to the Women Plus treatment arm, the recruiter will direct the message to everyone whom the eligible woman invited. After delivering the message about the work opportunity, the recruiter will show a video where a woman who works for the NGO (either as a volunteer or an employee, depending of the treatment arm) talks about her experience. The objective of this video is to reinforce the idea that the work opportunity is serious and increase interest from the woman and her family.

After showing the video, the recruiter will ask if the woman wants to apply. If so, the recruiter asks the eligible women to fill out an application form that includes the woman's name, phone number and requires the woman the sign the form. The recruiter emphasizes that this is the only opportunity for women to apply and that women who decide not to apply will not be offered future work opportunities in the project.

4. Hypotheses

This section describes the hypotheses we will test as part of three main research questions. The data to test these hyptoheses will be obtained from 3 household surveys and from administrative data from the NGO, as shown in Table 2 and described in more detail in Section 5. Appendix A includes empty tables illustrating how we will show the results that test our main hypotheses.

4.1 Research questions and hypotheses

Research question 1: *How does application behavior differ with changes in the way the work opportunity is presented?*

[19] If this is not possible after multiple attempts, we will continue with the visit anyway and direct the message to the eligible woman. During all visits we will record information on bystanders.

Hypothesis 1: *Application behavior differs between the four treatment arms: 1. Volunteering with Women Only, 2. Volunteering with Women Plus, 3. Employment with Women Only and 4. Employment with Women Plus*

Agglomerations are randomly assigned to either Volunteering or Employment, and household are randomly assigned to either Women Only or Women Plus. Therefore, these two dimensions of randomization are orthogonal to each other and we can add hypotheses 1A and 1B: hypothesis 1A compares Volunteering and Employment and hypothesis 1B compares Women Only and Women Plus.

Hypothesis 1A: *Application behavior differs between women who are offered a Volunteering work opportunity and an Employment work opportunity.*

Our pilot and target group discussions suggest that female employment outside the house is not always well-regarded in Egypt, especially in poor and conservative regions such as Sohag. This may be due to social norms regarding gender roles and family dynamics that dictate that men are breadwinners and women take care of children and household activities. In contrast, volunteering may be more well-regarded because it is seen as work to help others and the compensation is perceived to only cover the costs of volunteering (such as transportation and meals) and therefore does not threaten the men's position in the household as the breadwinner. In reality, the compensation is approximately 2500 Egyptian pound per month, which is approximately 50 USD and quite sizable for most households. All workers will be asked to work 12 days a month for 5 hours day on average, so the hourly rate is 42 EGP or 0.80 USD. Still, anecdotal evidence from our pilot indicates that households and especially husbands justify their choice to support their wives working by emphasizing that it is "merely volunteering", even though they themselves realize the compensation is substantial. Hence, take-up rates could be higher in the Volunteering treatment arm if social norms regarding women's employment outside of the home (but not volunteering) prevent them from looking for work.

Conversely, employment may be viewed as more attractive because it may be seen as more economically appealing (even though financial compensation is identical across treatment arms) or it could be associated with greater stability or perceived as a stepping-stone to subsequent employment (even though the duration of the contract is always fixed at 6 months). Adverse economic conditions that have affected Egypt in recent years may provide additional incentive to take on opportunities that women would otherwise have rejected. For these reasons, take-up rates could be higher in the Employment treatment arms.

As a robustness check, in addition to studying if there is differential interest in Volunteering and Employment for women, we will take a small sample, separately from our main sample, to also study if there is differential interest in Volunteering and Employment amongst men. If we found a gap in interest for women but not for men, that would provide additional support that social norms associated with the type of work affect only women's work out of the home.

While all characteristics of the work opportunity are identical between Volunteering and Employment treatment arms, we include questions to study perceptions about work responsibilities and norms to understand how perceptions may affect our results. We present women with 5

scenarios of a hypothetical woman in the community, and we randomize if we refer to them working as a volunteer or employee. On a five-point scale we ask how appropriate it is for this woman to skip work, never show up, show up late, decide not to put in much effort, and quit after one month. We ask about hypothetical women other than the respondent themselves to reduce potential social desirability bias and experimenter demand effects.

Hypothesis 1B: *Application behavior is different when women are encouraged to apply during a recrtuiment visit when she is by herself, compared to when other members of the household are explicitly invited to attend the recruitment visit.*

Women in Upper Egypt may lack autonomy to make the decision to work (El Feki et al., 2017; Caria et al., 2022). This was confirmed by our pilot and target group discussions, which suggest that the decision for women to work outside of the home usually requires the approval of other household members. Moreover, our pilot study suggested that 30% of women did not apply for a work opportunity because they thought their husband or parents would not approve of it.

On the one hand, recruiters may have more success getting women to apply if additional household members attend, for example because those household members may have more decision-making power and involving them directly in the recruitment process may provide them with information, increase their trust, or they may feel less threatened because of their involvement. On the other hand, involving other household members could instead lower application rates, as evidenced by Lowe and McKelway (2024). In India, they experimentally vary which spouse receives a ticket to a women's job opportunity and cross randomize whether spouses are encouraged to discuss the work opportunity. While experts they consulted predicted that discussion would increase take-up, they found the opposite with lower take-up rates amongst spouses who discussed the job opportunity. As such, we believe our paper will add new evidence on both the direction and magnitude of the effect of involving additional household members.

When asked during our pilot study, women differ in whom they prefer to invite to the recrtuiment visit. Some like to invite those with the most decision-making power (for example her husband or father), and some prefer to invite someone who she believes will stand up for her, such as her mother or sister. Therefore, in the Women Plus treatment arm, we will ask a woman to choose herself and we will record her choice as this may provide additional (non-experimental) insights in whom to target.[20] We will compare application behavior when we take to a woman by herself during the recruitment visit and when we encourage them to also invite others, in order to gain insight in whom to target when implementing measures to increase female labor force participation.

Ceteris paribus, changing the name of the work opportunity we offer and whom to target will help us understand if and how social norms constrain female labor force participation and what interventions can alleviate these constraints. For hypotheses 1, 1A and 1B, we will measure application behavior with the following outcomes:

[20] In this paper, we refer to inviting "other household members" in the Women Plus treatment arm because based on our pilot, we expect that most women will choose members of their own household. That said, she can also decide to invite others, such as a family member who does not live in the same house or a friend who could serve as a role model, for example because the friend has experience working out of her home.

Application behavior
We will use multiple measures of women's willingness to work outside of the home. First, we will use an indicator of whether women expressed interest in working for the NGO during the first household visit. Secondly, our main outcome variable for research question 1 is an indicator of whether women decide to apply for the work opportunity offered during the second or third household visit.

We will use these indicators to estimate the share of women interested in working outside of the home and to compare the application rates between volunteering and employment offers, and between women who are encouraged to involve other individuals in the decision and those who are not. In addition, we will use these indicators, together with the demographic and socioeconomic information from each woman (age, education, marital status and number of children) to study which characteristics correlate with the decision to express interest and apply for a work opportunity.

During the recruitment visit, we will also record the number of men and women who attend, and their relationship to the respondent (father, mother, siblings, etc.). We will compare who the women intended to invite and who actually showed up. Finally, we will compare who, amongst the attendents, asked questions and raised comments.

Research question 2: *What is the impact of working outside of the home on the female workers?*

Hypothesis 2: *The impact of working outside of the home differs between the four treatment arms: 1. Volunteering with Women Only, 2. Volunteering with Women Plus, 3. Employment with Women Only and 4. Employment with Women Plus.*

We will estimate and compare effects in all four treatment arms according to the statistical model described in Section 6.1. This model includes interactions of each treatment with the endogenous choice to apply as well as with the randomized work offers. This allows us to disentagle treatment effects from selection into applying, all of which may differ between the four treatment arms.

In addition to comparing the four treatment arms in hypothesis 2, we are also interested in estimating the effect of any work opportunity, irrespective of the treatment arm. Therefore, in hypothesis 2A, we will compare the outcomes of all applicants who were offered a work opportunity and all applicants who were not:

Hypothesis 2A: *Giving women the opportunity to work outside of the home impacts her and her household.*

For hypotheses 2 and 2A, we will test for a variety of effects from working for the NGO (be it in the form of employment or volunteering). For example, the work opporunity may serve as a stepping-stone towards increasing subsequent female labor force participation and incentivize women to look for additional work once the contract with the NGO has finished. In addition, the income received by women may increase household consumption and assets, and could provide women with additional autonomy which may translate into more influence in the decisions made in the household. Additionally, we may see effects on social norms and on her physical and mental

wellbeing. Finally, we may see broader effects on other household members, such as spouses, mothers-in-law, and children. Section 6 provides more details on these spillover effects.

Based on successful adherence to the household-level randomization of offers during the pilot study, we do not anticipate applicants who are not randomly selected to receive an offer to end up working for the NGO anyway. It could, however, happen that those who are randomized to receive the offer do not end up working for the NGO, for example because they changed their mind. As a result, for both hypotheses contained in research question 2, we will be estimating an intention to treat (ITT) effect.

We will test hypotheses 2 and 2A by conducting surveys with women who applied for the work opportunity four months and one year after their application, and comparing those who were randomized to receive a work offer and those who were not. The surveys will include questions to measure the following outcomes:

Labor market outcomes
We will collect information about women's status in the labor force. We will do this by asking several questions drawn from the 2018 Egypt Labor Market Panel Survey. Among those who are in the labor force, we will ask whether they are employed and, among those who are employed, we will ask for their earnings and hours of work and whether they work in the formal or informal sector. For those unemployed, we will collect information about search intensity, such as the amount of time spent looking for work and the methods used to look for work. If they are married, we will ask them for their husbands' opinion about their employment, and for those who have children, we will ask who cares for the children while they work.

Volunteering
We want to understand women's perceptions about volunteering and how those compare to their status in the labor force (obtained from the labor market outcomes questions) and job search (among the unemployed). To do that, we will ask women similar questions to the ones used to determine their status in the labor force, but this time refering to volunteering. Concretely, we will ask women if they participate or ever participated in volunteering, if they would be available to volunteer, and if not, for which reason(s). For those who are volunteering at the time of the survey, we will ask for their earnings and the periodicity with which they volunteer. If they are married, we will ask them for their husbands' view about their volunteering work, and for those who have children, we will ask who cares for the children while they volunteer.

Household decision making
We will measure women's involvement in household decisions by asking for their involvement in decisions related to using money earned by them and their spouse, working outside of the home, making small and large purchases, family planning, and children's education.

Household assets
We will collect information about the type of assets owned by the household. This will consist of questions about whether the household owns arable land and its size, the number of animals (cows, poultry, goats, etc.) that the household owns, the characteristics of the dwelling, the number of cellphones in the household, and whether any household member owns certain appliances and/or

vehicles. We will also collect basic financial information, such as whether they have a bank account and whether they participate in a ROSCA and if so, why.

Household consumption
In a household-level consumption module, we ask for the quantity and price of food and non-food items consumed. We ask respondents to include self-produced items and ask for the estimate value of such iterms. As our main measure we use per capita consumption and we use adult-equivalent per capita consumption as a robustness check.

Social norms and female empowerment
We will assess the impact of working on social norms and female empowerment by presenting women with several questions from the World Values Survey (the seventh edition) and from a study by Gazeaud et al. (2024) who measures social norms in Egypt specifically. These questions relate to acceptable behavior by women and men in the house, women's ability to work and its consequences for the household, and gender equality regarding children's education. Women will be asked to respond using a five-point Likert scale. We will combine the answers into an index of female empowerment to study the impact of working on this index, and we will also analyze each question separately.

Time use
We will measure the time allocated to various activities, including household chores, caring for children and elder household members, eating, sleeping, talking to friends and family, education, etc. To do this, we will ask women to recall the things they did in the previous day and how much time they spent on each activity. Following Hussam et al. (2021) we will also measure the self-reported average number of hours that respondents spend idle (this is framed as "sitting around with nothing to do").

Physical health
We ask women about their physical health, whether they became ill in the 30 days up to the survey and how many days they were ill. We also ask women whether they felt pain anywhere in their body in the 14 days to the survey and the severity of the pain on a four-point scale.

Psychological wellbeing
We will measure the effect of work on women's psychological wellbeing using several outcomes:

Depression: We will measure depression using two of the nine questions in the depression scale of the Patient Health Questionnaire (PHQ-2), a standardized screening tool that assesses mental and emotional health disorders.
Anxiety: We will measure anxiety using the seven-question depression scale of the General Anxiety Disorder (GAD-7), a standardized screening tool for the presence of a clinically significant anxiety disorder.
Stress: Following Hussam et al. (2021), we will measure stress using an index comprised of three questions from Cohen's Perceived Stress Scale, the most widely used tool for measuring the perception of stress.

Life Satisfaction: Drawing from Hussam et al. (2021), we will measure life satisfaction with with Diener's Satisfaction With Life Scalethat includes five statements measured on a seven-point Likert scale.
Sociability: We will obtain a measure of women's sociability by asking about the interactions that they have had during the day prior to the survey day. We will list the number of different people the respondent had a conversation with and how many of these interactions left them feeling positive.
Purpose: We follow Hussam et al. (2021) and ask four questions related to respondent's beliefs about their contribution to their family and community to construct and index of self-worth. All four questions are measured on a scale from 1-10.
Stability: We will measure women's feeling of stability in their current lives and in the future drawing from Hussam et al. (2021), who in turn adapt the Cantril Self-Anchoring Striving Scale (Cantril, 1965). The questions ask respondents to consider an eleven-step ladder, with the most secure life being a 10, and the least secure life being a 0. Respondents are asked to position themselves in step on the latter they feel they are on at present and where they anticipate standing in five years.

As shown in Table 2 all outcome variables for research question 2 will be measured during the baseline and endline survey, and the midline survey will include only a subset of questions.[21]

In order to estimate potential backlash effects of female labor force participation and get a clearer picture of the effects of the work opportunity on household dynamics, we will survey a subsample of husbands at midline to ask them the same questions regarding household decision making, social norms, and time use.

Research question 3: *Does worker productivity, performance and satisfaction differ between those who received offers in the four treatment arms?*

Hypothesis 3: *Worker productivity, performance and satisfaction differs between the four treatment arms: 1. Volunteering with Women Only, 2. Volunteering with Women Plus, 3. Employment with Women Only and 4. Employment with Women Plus.*

Participants in the study may consider the work relationship differently depending on whether it is called "volunteering" or "employment". On the one hand, volunteering may be treated by women as a more flexible and lower-commitment activity than employment. On the other hand, women who choose to volunteer may be intrinsically motivated to help others through the activities they are recruited for. Similarly, women's work performance and experience may depend on support from other household members which may differ by whether other household members were involved in the recruitment process. For example, if household responsibilities such as child care and household chores get shifted away from the study participant, this may allow her to concentrate more on work and perform better at it.

We will compare the following outcomes for women in the four treatment arms for hypothesis 3:

[21] The baseline survey that includes the exact framing of outcome variables can be found here.

Absences
We will record the number of days a woman was absent from work or training, and the number of days she showed up more than 1 hour later than the agreed time.

Supervisor assessment
Two NGO supervisors will independently review each worker on a five-point Likert scale in terms of their overall performance and performance on the following four components: 1. Productivity, 2. Quality of work, 3. Commitment, 4. Punctuality. In addition to reviewing the separate components, we will create an assessment index taking the average of the overall scores and the four components across the two supervisors. The NGO will share basic characteristics of the supervisors allowing us to test for potential biases by supervisor characteristic. In addition to that, the NGO will have performance reviews done by an independent reviewer who is not informed of whether the worker is a volunteer or employee in order to test for bias by job type (volunteering or employment). This will be done for a random subset of 10% of the workers and will allow us to test if reviews differ by reviewers who are and aren't aware of the job type. Finally, to test for bias in the performance review, we will compare the more subjective supervisor assessment with the more objective measure of number of days missed or late for work.

Work satisfaction
We will ask respondents who work to express how satisfied they are with their work on a five-point scale. We will ask this question for their work overall and for the following four components of their work: 1. Type of activities, 2. Work hours and flexibility, 3. Compensation (monetary and/or non-monetary), 4. People you work with.

While we will randomly assign work offers, those who are offered to work may subsequently decide not to work for the NGO after all. Therefore, as with research question 2, we will estimate an intention to treat effect for research question 3. However, unlike research question 2, the sample to test this hypothesis is only composed of women who decided to apply to the work opportunity and ended up working for the NGO, because all outcome variables for research question 3 can only be measured for those who work. Therefore, we can neither use the control group applicants not selected nor the control group of non-apllicants for causal inference. Selection into applying may differ between the four treatment arms, as we will study as part of research question 1, so the effects estimated as part of research question 3 should be thoughts of as the combined effect of selection into wanting to work and the direct effects of working. The insights from research 3 will, however, inform the NGO on the most appropriate way to recruit workers in order to maximize productivity.

4.2 Heterogeneity
Our household surveys start with a household roster that collects data about each household member's relation to the household head, age, education, marital status and number of children.

We will use this information to estimate heterogeneous effects by the women's age, marital status, number of children under age 18 and under age 5, education level (below and above the median for women), total values of consumption (below and above median of per capita consumption), and baseline level of answers to social norms questions toward female empowerment (below and above the median amongst respondents).

5. Data

This section describes the rounds of data collection, the construction and size of the sample, and measures to deal with non-compliance and attrition.

5.1 Survey rounds

The data for our project will be collected through three household surveys and through administrative data collection by our partner NGO. All three household surveys will be conducted in person and administered by our survey firm:

1) Baseline data will be collected during the first household visit as explained in Section 3.2 and occur between March 2025 until August 2026 as the project is rolled out in the study area.

2) The midline survey will be administered four months after the baseline survey so between July 2025 and December 2026. All work opportunities last for 6 months, so women who were selected to work for the NGO will be working during the midline survey.

Table 2. Data collection

	Baseline survey	Midline survey	Endline survey	Admin data
Research question 1				
Application behavior	X			X
Research question 2				
Employment	X	X	Reduced	
Volunteering	X	X	Reduced	
Household decision making	X	X	X	
Household assets	X		X	
Household consumption	X		X	
Social norms and female empowerment	X	X	X	
Time use	X	X	X	
Physical health	X	X	Reduced	
Depression (PHQ-2)	X	X	X	
Anxiety (GAD-7)	X	X	Reduced	
Stress	X	X	X	
Life satisfaction	X	X	X	
Sociability	X	X	X	
Purpose	X	X	Reduced	
Stability	X		X	
Research question 3				
Absences				X
Supervisor assessment				X
Work satisfaction		X		

Notes: Baseline, midline and endline data is collected by the survey firm, and admin data is provided by the NGO. Questions on employment and volunteering are reduced during the midline survey because these are only asked to

those who do not work as an employee or volunteer respectively as part of the project and we only ask basic questions on labor force participation, hours and total earnings. Questions on physical health, depression, anxiety and self-worth are reduced in the midline because we only ask 1 overall question rather than all components of the index.

3) The endline survey will be administered one year after the baseline survey, so between March 2026 and August 2027. All work opportunities provided by the project will have ended at the time of the endline survey.

Following Egypt's cultural norms, we will only work with female enumerators.[22]

In addition to household data, the NGO will collect administrative data on application behavior and work performance, as detailed in Section 4 and Table 2.

5.2 Sample

The NGO seeks to fill 900 work opportunities, 450 as Volunteers and 450 as Employees. Based on our pilot, we aim to reach 7,000 eligible women in order to get enough applications for Employment and Volunteering such that we can randomize who actually receives an offer. We expect to receive approximately 2,200 applications for the 900 work opportunities that we have. Also based on the pilot, there is on average one eligible women per household. In order to be conservative, we plan to visit 7,500 households who live in 500 agglomerations of 15 households each.

We expect approximately 20% of noncompliance among women who are offered the work opportunity, so we will make offers to 1,100 women among those who apply for the work opportunity. We will administer the baseline survey to all 7,500 households in the sample. The midline and endline surveys will be conducted with all the women who apply for the work opportunity. We will also carry out midline and endline surveys with a random subset of 1,000 women in treatment agglomerations who decided not to apply, equally divided amongst the four treatment arms. Therefore, we expect the sample for the midline and endline surveys to consist of 3,200 women.

5.3 Measures to deal with non-compliance

A potential concern in our study is that some women who applied and were randomly selected to be offered a work opportunity may decide not to work for the NGO after all. Therefore, we will estimate intention to treat effects for the hypotheses contained in research questions 2 and 3. To reduce the likelihood of non-compliance, women will be required to fill out a form when they apply that asks for the person's name, phone number and address. This application form states that the applicant will accept the work opportunity if selected and women are required to sign the form if they want to apply. Signing a document is a strong commitment in the Egyptian culture, so we expect that this will further reduce non-compliance.

[22] It is not socially accepted for women to talk to a man whom they do not know and visits their house.

5.4 Measures to deal with attrition

Another concern of this study is that women may not be found, or they may refuse to respond to the midline or endline survey. If the reasons for this are different between treated and untreated women, our estimates of the effect of working on female empowerment would be biased. To reduce the likelihood of attrition, we will collect phone numbers of all women who are interested to learn more about the work opportunity, so that in case a woman moved to another location we may still be able to reach her. In addition, and in order to incentivize participation, we will give survey respondents 100 EGP in the form of mobile credit for answering each of the two follow-up surveys. During the baseline survey, we will also collect GPS coordinates and precise descriptions of the household's location. Finally, given that each agglomeration consists of 15 neighboring households, we will check with neighbors if we cannot reach anyone in the household. In line with previous studies in the region, we expect to reach a recontact rate of approximately 90% of the women in our sample with these measures.

6. Estimation methodology

In this section, we outline the statistical model we will estimate to test the hypotheses described in Section 4. All equations will be estimated using the ANCOVA method. We then discuss the minimum effects we will be able to detect based on our design and sample size, as well as the measures we will put in place to deal with spillovers and the fact that we are testing several hypotheses simultaneously.

6.1 Statistical model

Table 3 gives an overview of the hypotheses introduced in Section 4 including the outcome category and which groups are compared. This subsection will introduce the equations we use to test these hypotheses.

Table 3. Hypotheses

Hypothesis	Outcome category	Comparison	Equation	Number of observations	Clusters	MDES
1	Application	All 4 treatment arms	(1)	7500 (baseline sample)	Agglomerations	0.147
1A	Application	Volunteer vs. Employee	(2)	7500 (baseline sample)	Agglomerations	0.147
1B	Application	Women only vs Women plus	(3)	7500 (baseline sample)	Households	0.076
2	Empowerment	All 4 treatment arms	(4)	3200 (mid/endline sample)	Agglomerations	0.168
2A	Empowerment	All offered vs All not offered	(5)	2200 (applicants only)	Agglomerations	0.184
3	Performance	All 4 treatment arms	(1)	900 (workers)	Agglomerations	0.225

Notes: The reported MDES (minimum detectable effect sizes) are the smallest effects that can be found with 90% power and α of 0.05. A conservative intra-class correlation coefficient of 0.2 is used to calculate the MDES for all the hypotheses that are clustered at the agglomeration level.

In hypothesis 1, we study application behavior in all four treatment arms by estimating the following equation:

$$Y_i = \beta_1 Volunteering_i \times WomenOnly_i + \beta_2 Volunteering_i \times WomenPlus_i + \beta_3 Employment_i \times WomenOnly_i + \beta_4 Volunteering_i \times WomenPlus_i + \Gamma X_i + \varepsilon_i \quad (1)$$

Y_i is one of the outcome of interest listed in Section 4. We do not include a constant so coefficients β_1 to β_4 show the mean of the outcome variable in each of the four treatment arms. X_i is a set of individual characteristics and ε_i is the error term. We cluster our standard errors at the agglomeration level because this is the largest geographical level of randomization, and the other level of randomization – households – are contained within agglomerations.

We test for differences between the four treatment arms by comparing the β coefficients. Specifically, testing if $\beta_1 = \beta_2$ reveals if there is a difference in Y_i between Women Only and Women Plus households in Volunteering agglomerations. Testing if $\beta_1 = \beta_3$ tells us if there are differences between Volunteering and Employment agglomerations when only women are targeted for recruitement, etc.

Appendix Table A1 illustrates what coefficients we will estimate using our model without control variables. For each outcome variable, we will add two additional columns. In the second column we will add the following individual-level control variables: age, marital status, number of children, and years of education. In the third and last column we will also add month of the year and village fixed effects. Unlike in many other countries, villages in Egypt are quite large with an average population size of over 10,000. Given that the four treatment arms are assigned randomly, the control variables and fixed effects are primarily used to increase precision of our estimation.

Because agglomerations are randomly assigned to either Volunteering or Employment, and households are randomly assigned to either Women Only or Women Plus, these two dimensions of randomization are orthogonal to each other. Therefore, we can compare Volunteering and Employment in hypothesis 1A and Women Only and Women Plus in hypothesis 1B. We use the same outcome variables for hypotheses 1, 1A and 1B, see Section 4.

We estimate hypothesis 1A using the following equation:

$$Y_i = \beta_1 Volunteering_i + \beta_2 Employment_i + \Gamma X_i + \varepsilon_i \quad (2)$$

Given that we do not include a constant, we can test for differences between Volunteering and Employment by testing if $\beta_1 = \beta_2$. Given that Volunteering and Employment is assigned at the agglomeration level, we cluster our standard errors at the agglomeration level.

We compare Women Only and Women Plus for hypothesis 1B in an equation similar to equation (2):

$$Y_i = \beta_1 WomenOnly_i + \beta_2 WomenPlus_i + \Gamma X_i + \varepsilon_i \quad (3)$$

Comparing if $\beta_1 = \beta_2$ will reveal if there are differences in the outcome variables between Women Only and Women Plus. Because randomization is assigned at the household level, we can cluster our standard errors at the household level, which increases power for this hypothesis. Appendix Table A1 shows our estimation of hypotheses 1, 1A and 1B.

For hypothesis 2, we study the effects of the four treatment arms on a variety of outcome variables listed in Section 4, while taking into account potential differential selection into the four treatment arms. After all, the number and characteristics of those who decide to apply may differ between the four treatment arms. We will estimate this simultaneously in the fully interacted model:

$$\begin{aligned} Y_i = {} & \beta_1 Volunteering_i \times WomenOnly_i + \beta_2 Volunteering_i \times WomenPlus_i \\ & + \beta_3 Employment_i \times WomenOnly_i + \beta_4 Employment_i \times WomenPlus_i \\ & + \beta_5 Volunteering_i \times WomenOnly_i \times Apply_i \\ & + \beta_6 Volunteering_i \times WomenPlus_i \times Apply_i \\ & + \beta_7 Employment_i \times WomenOnly_i \times Apply_i \\ & + \beta_8 Employment_i \times WomenPlus_i \times Apply_i \\ & + \beta_9 Volunteering_i \times WomenOnly_i \times Apply_i \times Offer_i \\ & + \beta_{10} Volunteering_i \times WomenPlus_i \times Apply_i \times Offer_i \\ & + \beta_{11} Employment_i \times WomenOnly_i \times Apply_i \times Offer_i \\ & + \beta_{12} Employment_i \times WomenPlus_i \times Apply_i \times Offer_i \\ & + \Gamma X_i + \varepsilon_i \end{aligned} \quad (4)$$

We do not include a constant, so β_1 to β_4 reflect the mean value of Y_i in each of the four treatment arms. The interactions between the treatment arms and whether or not person i decided to apply for the work opportunity, $Apply_i$, quantifies the selection into applying in terms of Y_i. For example, if we measure consumption for Y_i, then β_5 reflects the higher or lower baseline value of consumption for women who decided to apply to work as a Volunteer after being encouraged through the Women Only encouragement. Similarly, β_6, β_7 and β_8 measure selection into the other three treatment arms.

In order to estimate the causal effect of working, we rely on the randomized work offers reflected by $Offer_i$, which equals 1 if person i received a work offer. By interacting this term with each of the four treatment arms, we quantify the change in Y_i as a result of receiving the job offer.[23] For example, β_9 captures the causal effect of receiving a work offer in terms of Y_i in the Volunteering and Women Only treatment arm. For the other three treatment arms the causal effect is captured by β_{10}, β_{11} and β_{12}.

For example, comparing if $\beta_5 = \beta_6$ will test if in Volunteering agglomerations there is a baseline difference in Y_i between those who decided to apply in Women Only and Women Plus households. Testing if $\beta_6 = \beta_8$ reveals if there is a difference in the importance of Women Plus in the decision to apply between Volunteering and Employment agglomeration, etc. Similarly, we can test for differences in the effects of receiving a work offer: comparing $\beta_9 = \beta_{10}$ reveals if Y_i is affected by involving other household members in encouraging women to work in Volunteering agglomerations. And comparing $\beta_9 = \beta_{11}$ will tell us if there is a diffential effect in Y_i between those in Volunteering and Employment agglomerations, both of which were encourage to apply through Women Only.

[23] We multiply the last four interaction terms also by the decision to apply, $Apply_i$, but given that only those who apply should receive a work offer, this should not affect the estimation. Based on our pilot, we do not expect any non-applicants to received a work offer, so we multiple by $Apply_i$ merely for clarity.

Appendix Tables A2, A3 and A4 show our estimation of research question 2 on three groups of outcomes: Table A2 shows how we will estimate effects on subsequent labor market outcomes; Table A3 will show effects on household-level outcomes, household decision making and social norms, and Table A4 will show effects on physical and pshychological wellbeing, as detailed in Section 4.

We will be estimating intention to treat (ITT) effects of receiving the work offer, rather than the effect of actually working. After all, some of the women who receive an offer may not end up working for the NGO for various reasons. Based on the pilot and previous studies that offered work opportunities to women, we expect this to be around 20 percent (Donald and Grosset-Touba, 2024; Jalota and Ho, 2024). As an exploratory exercise, we will use the randomized work offers as an instrument for actually working for the NGO to estimate the local average treatment effect (LATE) for those who complied to the randomized work offers by working for the NGO.

In order to estimate the effect of receiving any work offer, hypothesis 2A focusses on applicants only and compares all those who received a work offer to all those who did not. As such, this captures a weighted effect of receiving a work offer across all four treatment arms and is estimated using the following equation:

$$Y_i = \beta_0 + \beta_1 Offer_i + \Gamma X_i + \varepsilon_i \quad (5)$$

The sample consists of applicants only and $Offer_i$ is an indicator that takes the value of one if the woman was offered a work opportunity (be it as Volunteering or Employment and through Women Only or Women Plus), and zero otherwise. Given that we include an intercept here, β_1 captures the effect of randomly receiving a work offer on Y_i. Appendix Table A5 shows our estimation of hypothesis 2A.

For research question 3, we will estimate the combined effects of selection and treatment for all four treatment arms, as detailed in Section 4.1. We will do so by estimating equation (1) for the sample of women who end up working for the NGO. After all, those are the women for whom we can measure worker productivity, performance and satisfaction. As such, the sample consists of all 900 workers. Y_i is the outcome of interest and coefficients β_1 to β_4 show the combined effect in each of the four treatment arms. As before, we cluster our standard errors at the agglomeration level. Our estimation of research question 3 is shown in Appendix Table A6.

Finally, there may be important interactions between Volunteering or Employment on the one hand and Women Only or Women Plus on the other hand. For example, the effect of Women Plus may be different in Volunteering Agglomerations than in Employment Agglomerations. Acknowledging these interactions, we include four subhypotheses each to research questions 1, 2 and 3. In Table A7 we list these hypotheses. For example, in subhypotheses 1.1, we take the sample of Volunteering agglomerations, and amongst those, we compare application behavior between Women Only and Women Plus. For subhypotheses 1.2, we do the same in Employment agglomerations. Then in subhypotheses 1.3 and 1.4, we compare application behavior between Volunteering and Employment in the sample of households that were randomized into Women Only and Women Plus, respectively. Similarly, subhypotheses 2.1, 2.2, 2.3 and 2.4 study effects

as part of research question 2 in the same four samples, and subhypotheses 3.1, 3.2, 3.3 and 3.4 study effects on performance. These 12 subhypotheses allows us to flexibly test for all possible interactions.

6.2 Power calculations

Our study area will be divided into 500 agglomerations, 250 in which we will recruit employees and 250 in which we will recruit volunteers. Based on our pilot, we assume that there will be on average one eligible woman per household, and each agglomeration will consist of 15 households. Table 3 shows the minimum detectable effect sizes (MDES) for each hypothesis, assuming a conservative intraclass correlation coefficient (ICC) of 0.2 when applicable. We can detect an effect size of 0.147 of a standard deviation in the application rates between Volunteers and Employees in the Women Only and Women Plus groups with 90% power, following hypothesis 1. The power curves for various ICC values are shown in Appendix Figure C1.

Following hypothesis 1A, we can also detect an effect size of 0.147 of a standard deviation in the application rates between the entire Volunteering and Employment groups with 90% power given an ICC of 0.2. For hypothesis 1B, we will randomize who is encouraged (Women Only versus Women Plus) at the household level. Assuming 7,500 households (15 eligible women per agglomeration in each of the 500 agglomerations), we can detect an effect size of 0.076 of a standard deviation in the application rate between offering an opportunity in the Women Only versus Wwomen Plus treatment with 90% power. Power curves for hypotheses 1A and 1B are shown in Appendix Figure C2.

When studying the effects of Volunteering or Employment in the Women Only and Women Plus groups in hypothesis 2, we expect to have 3,200 women in our midline and endline surveys who were encouraged to apply for work. Again assuming an ICC of 0.2, we should be able to detect an effect of 0.168 standard deviations with 90% confidence for hypotheses 2, which use the same sample but estimate effects on different outcomes. Finally, when we estimate the effects of being offered a work position – whether that be Volunteering or Employment – in hypothesis 2A, we expect to have 2,200 applicants, 1,100 of whom will receive work offers. This gives us the capacity to detect an effect of 0.184 standard deviations with 90% power given an ICC of 0.2. Power curves for hypotheses 2 and 2A are also shown in Appendix Figure C2.

When estimating the LATE for hypothesis 2 and 2A, since we expect a noncompliance rate of 20%, the minimum effects we are able to detect are 0.21 and 0.23, respectively. If the noncompliance rate were to increase to 30%, the minimum effects we would be able to identify would be 0.24 and 0.263, respectively.

For hypotheses 3, the sample size is smaller, namely the 900 workers for whom the outcome variables can be measured. As a result, we expect to be able to detect an effect of 0.225 standard deviations with 90% confidence for hypotheses, assuming an ICC of 0.2.

Finally, Table A7 and Appendix Figure C3 show the minimum detectable effect sizes and power curves for all 12 subhypotheses. Especially for subhypotheses 3.1 to 3.4 the samples become small because we draw from the sample of workers only, so the minimum detectable effect sizes

are large. As such, it will be less likely that we can detect significant effects so these effects may be seen as suggestive only.

6.3 Measures to deal with spillovers

Spillovers in our survey can happen for various reasons: first, if women offered a volunteering position learn that other women are being offered the same position in the form of employment or vice-versa), this may affect their decision to apply and/or their work performance and thus bias our results. We aim to limit this form of spillover by dividing the study area of 400,000 individuals into 500 agglomerations of 15 households each. Between any two agglomerations we will include a "buffer" of homes that will not be included in our study in order to spread out agglomerations within our study area, maximizing the distance between them. Given the large and spread-out study area, we believe the probability is low that households in any two agglomerations will know each other. Nonetheless, we will collect GPS coordinates of all sample households, allowing us to test if effects are different for Employment agglomerations that randomly happen to be closer to Volunteering agglomerations and vice versa.

Secondly, within an agglomeration, the behavior of some women could affect the behavior of others. For example, women who decide to apply may increase the likelihood of others applying too, or women in Women Plus households who choose to invite more household members to the recruitment meetings may increase the number for invitees for other women in the Women Plus treatment arm as well. We believe these are interesting spillovers and we aim to quantify them by using the fact that the randomization into Women Only and Women Plus occurs at the household-level and is independent of the agglomeration. As a result, there will be substantial variation in the share of Women Plus between agglomerations. So as long as application behavior differs sufficiently between Women Only and Women Plus, as we will test in hypothesis 1B, we can estimate if randomly having a higher shares of Women Plus households in an agglomeration leads to different application rates and to a different number of attendees for Women Plus households.

Thirdly, the outcomes of other household members may be affected by women working in their household. These effects may be negative, for example because household and childcare responsibilities shift to those not working, or positive, for example because people are inspired by seeing their female family members working as she serves as their role model. In order to quantify such spillovers, our three household surveys will ask her about the activities performed by all other household members. To be precise, we will ask about each household member's marital status, number of children, education (whether they are or have ever been in school), employment (whether in the labor force, and if so, whether employed for a wage, self-employed, or not employed), hours worked in the week prior to the survey (for those who are working), and whether they are volunteering and if so, how many hours they volunteered in the week prior to the survey.

6.4 Correcting for multiple hypothesis testing

Because we will test several hypotheses, we risk finding significant results on some outcomes by chance. Because our outcomes can be grouped into "families" based on the hypotheses listed in Section 4, we will control for the family-wise error rate using the permutation procedure proposed in Westfall and Young (1993). The advantage of this procedure over other corrections for multiple

hypothesis tests is that it allows for dependence across outcomes, which is a likely feature in our setting.

In addition to correcting for multiple hypothesis testing, we will follow Kling, Liebman, and Katz (2007) to create indexes of groups of related outcomes and estimate effects on these indexes. To construct these indexes, we combine related outcome variables into groups and define each outcome so that a higher value corresponds to more favorable outcomes. We then standardize each outcome into a Z-score by subtracting the mean and dividing by the standard deviation. For each group of outcomes we average all the Z-scores and standardize the average relative to the control group for each hypothesis. Then, rather than estimating effects on individual outcome variables, we estimate effects on these indexes according to the same statistical models discussed in Section 6.1.[24]

7. Conclusion

In this project we aim to learn how to increase female work out of the home in rural Egypt. In order to fill 900 work opportunities with a local NGO, we will cross randomize two distinct treatments. First, in geographically distinct areas, we encourage women to either work as volunteers or as employees for otherwise identical work opportunities. Secondly, at the household level, we will randomize whether we encourage only the woman to work, or whether we involve other household members in the decision-making process.

We aim to contribute to a rapidly growing literature on the importance of social norms and how these may impede women's opportunities to work. In the case of Saudi Arabia, Bursztyn et al. (2020) provide encouraging evidence that, in an environment characterized by pluristic ignorance, correcting second-order beliefs may increase women's work out of the home. Given that we do not find evidence of pluristic ignorance in our sample area in rural Egypt, we propose two different strategies to make women's work opportunities conform with prevailing social norms: presenting the opportunity as volunteering instead of employment and involving not just the potential applicant but also other household members in the recruitment process.

We furthermore extend on existing work by randomizing work opportunities amongst applicants, which allows us to look beyond application rates and study the effects of working on women and their families and whether and how those effects differ by social norm regimes. Therefore, in light of a steady decrease in the gender education gap is recent decades, we believe this project will provide policy-relevant evidence to help overcome the region's major challenge of low female labor force participation.

[24] This approach was also used by Banerjee, Duflo, Goldberg, Karlan, Osei, Parienté, Shapiro, Thuysbaert and Udry (2015) to assess how poverty graduation programs in six different countries affect a range of outcomes.

References

Aalen, L., Kotsadam, A., Pieters, J., and Villanger, E. (2024). Jobs and political participation-evidence from a field experiment in Ethiopia. *The Journal of Politics*, 86(2), 656-671.

Ahramonline. (2023a, June 20). *Egypt raises minimum wage for private sector workers by 11%.* Ahram Online. https://english.ahram.org.eg/News/503365.aspx

Ahramonline. (2023b, March 2). *Egypt raises minimum wage for government employees to EGP 3,500 & pensions by 15% as of 1 april - politics - egypt*. Ahram Online. https://english.ahram.org.eg/News/491010.aspx

Antecol, H. (2001). Why is there interethnic variation in the gender wage gap?: The role of cultural factors. *Journal of Human Resources*, 36(1), 119-143.

Anukriti, S., Herrera-Almanza, C., Pathak, P. K., and Karra, M. (2020). Curse of the Mummy-ji: the influence of mothers-in-law on women in India. *American Journal of Agricultural Economics*, 102(5), 1328-1351.

Ashraf, N., Bandiera, O., Davenport, E., and Lee, S. S. (2020). Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. *American Economic Review*, 110(5), 1355-1394.

Ashraf, N., Bau, N., Nunn, N., and Voena, A. (2020). Bride price and female education. *Journal of Political Economy*, 128(2), 591-641.

Assaad, R., Hendy, R., Lassassi, M., and Yassin, S. (2020). Explaining the MENA paradox: Rising educational attainment yet stagnant female labor force participation. *Demographic Research*, 43(28), 817-850.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B., and Udry, C. (2015). A multifaceted program causes casting progress for the very poor: evidence from six countries. *Science*, 348(6236), 1260799.

Bau, N., and Fernández, R. (2023). Culture and the Family. *Handbook of the Economics of the Family*, 1(1), 1-48.

Bénabou, R., and Tirole, J. (2011). Laws and norms. NBER Working Paper 17579.

Berlinski, S., and Galiani, S. (2007). The effect of a large expansion of pre-primary school facilities on preschool attendance and maternal employment. *Labour Economics*, 14(3), 665-680.

Berlinski, S., Galiani, S., and Mc Ewan, P. J. (2011). Preschool and maternal labor market outcomes: Evidence from a regression discontinuity design. *Economic Development and Cultural Change*, 59(2), 313-344.

Bhanot, S. P., Han, J., and Jang, C. (2018). Workfare, wellbeing and consumption: Evidence from a field experiment with Kenya's urban poor. *Journal of Economic Behavior & Organization*, 149, 372-388.

Blattman, C., and Dercon, S. (2018). The Impacts of industrial and entrepreneurial work on income and health: Experimental evidence from Ethiopia. *American Economic Journal: Applied Economics*, 10(3), 1-38.

Blattman, C., Dercon, S., and Franklin, S. (2022). Impacts of industrial and entrepreneurial jobs on youth: 5-year experimental evidence on factory job offers and cash grants in Ethiopia. *Journal of Development Economics*, 156, 102807.

Breza, E., Kaur, S., and Shamdasani, Y. (2021). Labor rationing. *American Economic Review,* 111(10), 3184-3224.

Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in Saudi Arabia. *The American Economic Review*, 110(10), 2997-3029.

Caria, S., Crepon, B., ElBehairy, H., Fadlalmawla, N., Krafft, C., Nagy, A., Mottaghi, L., Zeitoun, N., and El Assiouty, S. (2022). Child care subsidies, employment services and women's labor market outcomes in Egypt: first midline results. *World Bank*.

Central Agency for Public Mobilization and Statistics. (2018). Household income, expenditure, and consumption survey 2017-2018 (HICECS 2017-2018). Cairo, Egypt: Economic Research Forum, Central Agency for Public Mobilization and Statistics. https://www.erfdataportal.com/index.php/catalog/168

Coen-Pirani, D., León, A., and Lugauer, S. (2010). The effect of household appliances on female labor force participation: Evidence from microdata. *Labour Economics*, 17(3), 503-513.

Cortés, P., and Tessada, J. (2011). Low-skilled immigration and the labor supply of highly skilled women. *American Economic Journal: Applied Economics*, 3(3), 88-123.

Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., and Zamora, P. (2013). Do labor market policies have displacement effects? Evidence from a clustered randomized experiment. *The Quarterly Journal of Economics*, 128(2), 531-580.

Dean, J. T., and Jayachandran, S. (2019). Changing family attitudes to promote female employment. *AEA Papers and Proceedings*, 109, 138-142.

Dhar, D., Jain, T., and Jayachandran, S. (2022). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India. *American Economic Review*, 112(3), 899-927.

Dinkelman, T. (2011). The effects of rural electrification on employment: New evidence from South Africa. *American Economic Review*, 101(7), 3078-3108.

Donald, A., and Grosset-Touba, F. (2024). Complementarities in Labor Supply: How Joint Commuting Shapes Work Decisions. Working paper.

East, C. N., and Velásquez, A. (2024). Unintended consequences of immigration enforcement: Household services and high-educated mothers' work. *Journal of Human Resources*, 59(5), 1458-1502.

El Feki, S., Heilman, B., and Barker, G. (2017). Understanding masculinities: results from the international men and gender equality survey (IMAGES) - Middle East and North Africa. UN Women and Promundo-US.

El-Gamal, M., El-Komi, M., Karlan, D., and Osman, A. (2014). Bank-insured RoSCA for microfinance: Experimental evidence in poor Egyptian villages. *Journal of Economic Behavior & Organization*, 103, S56-S73.

Field, E., Pande, R., Rigol, N., Schaner, S., and Troyer Moore, C. (2021). On her own account: How strengthening women's financial control impacts labor supply and gender norms. *American Economic Review*, 111(7), 2342-2375.

Ho, L., Jalota, S., and Karandikar, A. (2024). Bringing work home: flexible work arrangements as gateway jobs for women in West Bengal. Working Paper.

Hussam, R., Kelley, E. M., and Lane, G. (2023). The gendered value of employment. Working Paper.

Hussam, R., Kelley, E. M., Lane, G., and Zahra, F. (2022). The psychosocial value of employment: evidence from a refugee camp. *The American Economic Review*, 112(11), 3694-3724.

International Labour Organization. ILO modelled estimates database ILOSTAT. Accessed April 25, 2023. https://ilostat.ilo.org/data/.

Jalota, S., and Ho, L. (2024). What works for her? How work-from-home jobs affect female labor force participation in urban India. Working Paper.

Jayachandran, S. (2015). The roots of gender inequality in developing countries. *Annual Review of Economics* 7(1): 63–88.

Jayachandran, S. (2021). Social norms as a barrier to women's employment in developing countries. *IMF Economic Review*, 69(3), 576-595.

Karlan, D., Osman, A., and Shammout, N. (2021). Increasing financial inclusion in the Muslim world: Evidence from an islamic finance marketing experiment. *The World Bank Economic Review*, 35(2), 376-397.

Khalifa, S. (2022). Female genital cutting and bride price. Working paper.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1), 83-119.

Kotsadam, A., and Villanger, E. (2022). Jobs and intimate partner violence-evidence from a field experiment in Ethiopia. *Journal of Human Resources.* 58(6).

Krafft, C., Assaad, R. and Rahman, K.(2021). Introducing the Egypt labor market panel survey 2018. *IZA Journal of Development and Migration*, 12(12), 1-40.

La Ferrara, E. (2019). Presidential address: Aspirations, social norms, and development. *Journal of the European Economic Association*, 17(6), 1687-1722.

Lowe, M., and McKelway, M. (2024). Coupling labor supply decisions: an experiment in India. Working Paper.

Majlesi, K (2016). Labor market opportunities and women's decision making power within households. *Journal of Development Economics*, 119, 34-47.

McKelway, M. (2022). Women's employment and empowerment: Descriptive evidence. *AEA Papers and Proceedings*, 112, 541–45.

McKelway, M. (2024). Women's self-efficacy and economic outcomes: Experimental evidence from India. Working Paper.

Osman, A., Speer, J., and Weaver, A. (2024). Discrimination against women in hiring. Forthcoming in *Economic Development and Culture Change*.

Peri, G., Romiti, A., and Rossi, M. (2015). Immigrants, domestic labor and women's retirement decisions. *Labour Economics*, 36, 18-34.

Qian, N (2008). Missing women and the price of tea in China: the effect of sex-specific earnings on sex imbalance. *The Quarterly Journal of Economics*, 123(3), 1251-1285.

Romiti, A. (2018). The effects of immigration on household services, labour supply and fertility. *Oxford Bulletin of Economics and Statistics*, 80(4), 843-869.

Schwarzer, R., and Jerusalem, M. (1995). Generalized self-efficacy scale. *J. Weinman, S. Wright, and M. Johnston, Measures in health psychology: A user's portfolio. Causal and control beliefs*, 35, 37.

Sulaiman, Kabuye Uthman (2011) Volunteerism from Islamic perspective. *International Conference on Humanities*.

Westfall, P. H., and Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*. John Wiley & Sons.

World Bank. (2013). Opening doors: gender equality and development in the middle east and north Africa. MENA Development Report. Washington, DC. http://hdl.handle.net/10986/12550

Appendix A – Tables

Table A1. Application behavior (Hypotheses 1, 1A, and 1B)

	Expressed interest	Application decision	Number of attendees
Hypothesis 1	(1)	(2)	(3)
Volunteering × WomenOnly	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenPlus	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)
Employment × WomenOnly	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)
Employment × WomenPlus	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)
N			
R-squared			
Hypothesis 1A	(1)	(2)	(3)
Volunteering	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)
Employment	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)
N			
R-squared			
Hypothesis 1B	(1)	(2)	(3)
WomenOnly	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)
WomenPlus	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)
N			
R-squared			

Notes: Standard errors are clustered at the agglomeration level for hypotheses 1 and 1A. Standard errors are clustered at the household level for hypothesis 1B. As explained in Section 6, every column will be repeated twice: first with individual-level control variables age, marital status, number of children, and years of education and, second, also including month of the year and village fixed effects. In addition to this table showing the main results for research question 1, appendix tables will show results for the following outcomes variables: application decision during second and third household visit separately, number of women and men present separately, number of attendees raising comments or questions, share of those invited who show up.

Table A2: Impact of working on labor market outcomes (Hypothesis 2)

	Labor force participation	Employment	Volunteering	Hours of work	Log earnings	Formal sector	Job search intensity	Work index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Volunteering × WomenOnly	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenPlus	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenOnly	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenPlus	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenOnly × Apply	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenPlus × Apply	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenOnly × Apply	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenPlus × Apply	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenOnly × Apply × Offer	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenPlus × Apply × Offer	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenOnly × Apply × Offer	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenPlus × Apply × Offer	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
N								
R-squared								

Notes: Standard errors are clustered at the agglomeration level. As detailed in Section 5, the sample consists of 2,700 women eligible to work, divided between 900 applicants who randomly received a work offer, 900 applicants who did not receive a work offer, and 900 non-applicants. As explained in Section 6, every column will be repeated twice, sequentially adding control variables and fixed effects. Please refer to Section 4 for details on the outcome variables. All labor market outcomes are measured at the time of the endline survey when the work opportunity will have ended. Columns 4, 5 and 6 are conditional on working and column 7 is conditional on not working. Formal sector is an indicator for working formally. Work index is defined according to Section 6.4, following Kling et al (2007).

Table A3: Impact of working on household outcomes, decision making and social norms (Hypothesis 2)

	Log HH consumption	Log HH assets	Hours spent HH & care	HH decision making index	Financial independence	Mobility independence	Empower-ment index	Social norms index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Volunteering × WomenOnly	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenPlus	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenOnly	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenPlus	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenOnly × Apply	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenPlus × Apply	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenOnly × Apply	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenPlus × Apply	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenOnly × Apply × Offer	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenPlus × Apply × Offer	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenOnly × Apply × Offer	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenPlus × Apply × Offer	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
N								
R-squared								

Notes: Standard errors are clustered at the agglomeration level. As detailed in Section 5, the sample consists of 2,700 eligible women. As explained in Section 6, every column will be repeated twice, sequentially adding control variables and fixed effects. Please refer to Section 4 for details on the outcome variables. Columns 1 and 2 are measured at the household level and divided by the number of household members. We will repeat these columns with adult equivalent per capita values in the Appendix. Columns 3 to 8 are measured at the individual level. Column 3 measures the hours spent on household chores and care for children and elderly. Indices in columns 4, 7 and 8 are defined according to Section 6.4, following Kling et al (2007). Columns 5 and 6 measures whether the respondent can make independent decisions about her finances and her mobility measured as leaving her home.

Table A4: Impact of working on physical and pshychological wellbeing (Hypothesis 2)

	Physical health index	Depression index	Anxiety index	Stress index	Life satisfaction index	Sociability index	Self-worth index	Stability index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Volunteering × WomenOnly	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenPlus	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenOnly	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenPlus	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenOnly × Apply	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenPlus × Apply	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenOnly × Apply	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenPlus × Apply	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenOnly × Apply × Offer	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenPlus × Apply × Offer	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenOnly × Apply × Offer	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Employment × WomenPlus × Apply × Offer	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
N								
R-squared								

Notes: Standard errors are clustered at the agglomeration level. As detailed in Section 5, the sample consists of 2,700 eligible women. As explained in Section 6, every column will be repeated twice, sequentially adding control variables and fixed effects. Please refer to Section 4 for details on the outcome variables and the variables that comprise each index. All indices are defined according to Section 6.4, following Kling et al (2007).

Table A5: Impact of any type of work amongst applicants (Hypothesis 2A)

	Labor force participation	Employment	Volunteering	Hours of work	Log earnings	Formal sector	Job search intensity	Work index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Offer	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Control mean								
N								
R-squared								
	Log HH consumption	Log HH assets	Hours spent HH & care	HH decision making index	Financial independence	Mobility independence	Empower-ment index	Social norms index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Offer	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Control mean								
N								
R-squared								
	Physical health index	Depression index	Anxiety index	Stress index	Life satisfaction index	Sociability index	Self-worth index	Stability index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Offer	coef	coef	coef	coef	coef	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Control mean								
N								
R-squared								

Notes: Standard errors are clustered at the agglomeration level. As detailed in Section 5, the sample consists of 1,800 applicants, divided between 900 applicants who randomly received a work offer and 900 applicants who did not. As explained in Section 6, every column will be repeated twice, sequentially adding control variables and fixed effects. Please refer to Section 4 and the notes to Tables A2, A3, and A4 for details on the outcome variables.

Table A6: Impact on work productivity, performance and satisfaction (Hypothesis 3)

	Number of days missed or late	Overall performance	Overall work satisfaction
	(1)	(2)	(3)
Volunteering × WomenOnly	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)
Volunteering × WomenPlus	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)
Employment × WomenOnly	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)
Employment × WomenPlus	coef	coef	coef
	(s.e.)	(s.e.)	(s.e.)
N			
R-squared			

Notes: Standard errors are clustered at the agglomeration level. As detailed in Section 5, the sample consists of 900 women who end up working for the NGO. As explained in Section 4, a worker is counted as late if she shows up at least one hour after the agreed time and the performance scores are average values given by two supervisors. As explained in Section 6, every column will be repeated twice, sequentially adding control variables and fixed effects. In addition to this table showing the main results for research question 3, appendix tables will show results for the following disaggregated outcomes variables: numbers of days missed, number of days at least 1 hour late, average score of each of the four components of performance and score of each of the four components of work satisfaction. Please refer to Section 4 for additional details on these outcome variables.

Table A7. Hypotheses and subhypotheses estimating interactions

Hypothesis		Outcome category	Comparison	Equation	Number of observations	Clusters	MDES
1		Application	All 4 treatment arms	(1)	7500 (baseline sample)	Agglomerations	0.147
	1.1	Application	WO vs WP amongst volunteers	(1)	3750 (baseline sample)	Households	0.107
	1.2	Application	WO vs WP amongst employees	(1)	3750 (baseline sample)	Households	0.107
	1.3	Application	V vs E amongst WO	(1)	3750 (baseline sample)	Agglomerations	0.160
	1.4	Application	V vs E amongst WP	(1)	3750 (baseline sample)	Agglomerations	0.160
1A		Application	Volunteer vs. Employee	(2)	7500 (baseline sample)	Agglomerations	0.147
1B		Application	Women only vs Women plus	(3)	7500 (baseline sample)	Households	0.076
2		Empowerment	All 4 treatment arms	(4)	3200 (mid/endline sample)	Agglomerations	0.168
	2.1	Empowerment	WO vs WP amongst volunteers	(4)	1600 (baseline sample)	Households	0.164
	2.2	Empowerment	WO vs WP amongst employees	(4)	1600 (baseline sample)	Households	0.164
	2.3	Empowerment	V vs E amongst WO	(4)	1600 (baseline sample)	Agglomerations	0.201
	2.4	Empowerment	V vs E amongst WP	(4)	1600 (baseline sample)	Agglomerations	0.201
2A		Empowerment	All offered vs All not offered	(5)	2200 (applicants only)	Agglomerations	0.184
3		Performance	All 4 treatment arms	(1)	900 (workers)	Agglomerations	0.225
	3.1	Performance	WO vs WP amongst volunteers	(1)	450 (workers)	Households	0.265
	3.2	Performance	WO vs WP amongst employees	(1)	450 (workers)	Households	0.265
	3.3	Performance	V vs E amongst WO	(1)	450 (workers)	Agglomerations	0.292
	3.4	Performance	V vs E amongst WP	(1)	450 (workers)	Agglomerations	0.292

Notes: The reported MDES (minimum detectable effect sizes) are the smallest effects that can be found with 90% power and α of 0.05. A conservative intra-class correlation coefficient of 0.2 is used to calculate the MDES for all the hypotheses that are clustered at the agglomeration level. We use abbreviations for our four treatment arms. "V" stands for Volunteering and "E" stands for Employment; WO stands for Women Only and WP stands for Women Plus.

Appendix B – Work description

This appendix describes the work opportunities that will be available for women during the study. To be eligible, a woman must be aged 18 to 55, and be present in the household at the time of the baseline survey. There will be a total of 900 job opportunities for women to work for the NGO. All job opportunities will last six months, and women will work 12 days per month and 5 hours each day, at a rate of approximately 2500 EGP per month. There is no difference between treatment arms in the characteristics of the work opportunities.

All workers will work for the NGO, Life Makers, as part of the "Comprehensive Development and Women's Empowerment Project" in Sohag. Their project's theory of change centers on empowering women in marginalized communities: if women in marginalized communities are equipped with essential awareness and knowledge, while simultaneously addressing the critical needs of their families and implementing interventions that enhance per capita income, this will elevate the role of women in society. As a result, these communities are more likely to embrace change, acknowledge and support the leadership roles of women, and assist in enabling women to realize their full potential.

The project includes various components such as an awareness campaign and adult literacy classes. The NGO will assign tasks to each worker based on their needs within the broader project and the worker's education level. For example, some workers will work as awareness campaign facilitators, conducting group awareness sessions with 10-15 women at a time and one-on-one sessions during home visits. The awareness campaign with follow Egypt's AWSO guide and Bab Amal's awareness messages. Other workers will contribute to the adult literacy classes, either as literacy classes facilitators, delivering informal small-scale classes to 10-12 adult women in each class, or as administrative assistants for the literacy classes arranging spaces to conducting classes, making sure that the literacy classes run smoothly, and taking attendance. Additional work opportunities include nursery nanny, nursery teacher, seamstress at a quilt workshop, secretary, and community and outreach facilitator.

When assigning tasks, the NGO will not factor in the work type – Volunteering or Employment. The NGO will provide us with administrative data on the type of tasks performed by each woman (as well as measures of their performance that we use for research question 3). This will allow us to test directly if there are indeed no systematic differences in the tasks performed by volunteers and employees. Similarly, during the midline survey we will ask workers what tasks they perform, to ensure that there are no differences between treatment arms.

Appendix C – Power curves

Figure C1. Power as a function of effect size for hypothesis 1 for various ICC values

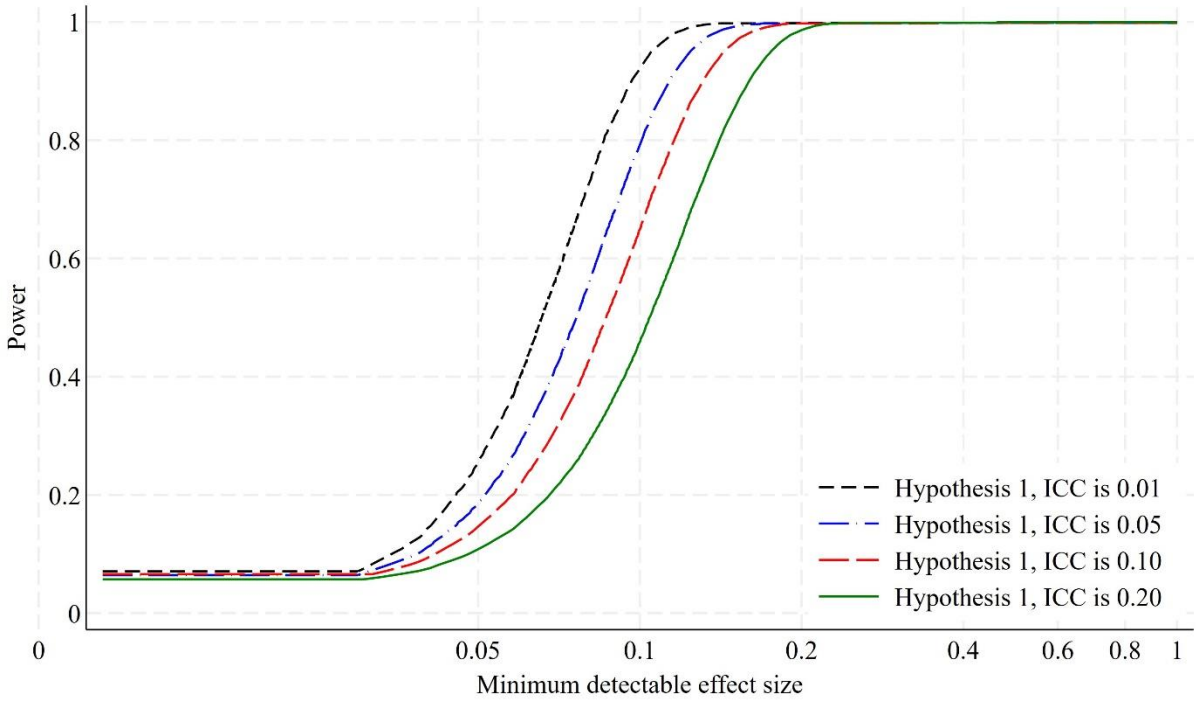


Figure C2. Power as a function of effect size for hypotheses 1A, 1B, 2, 2A, and 3

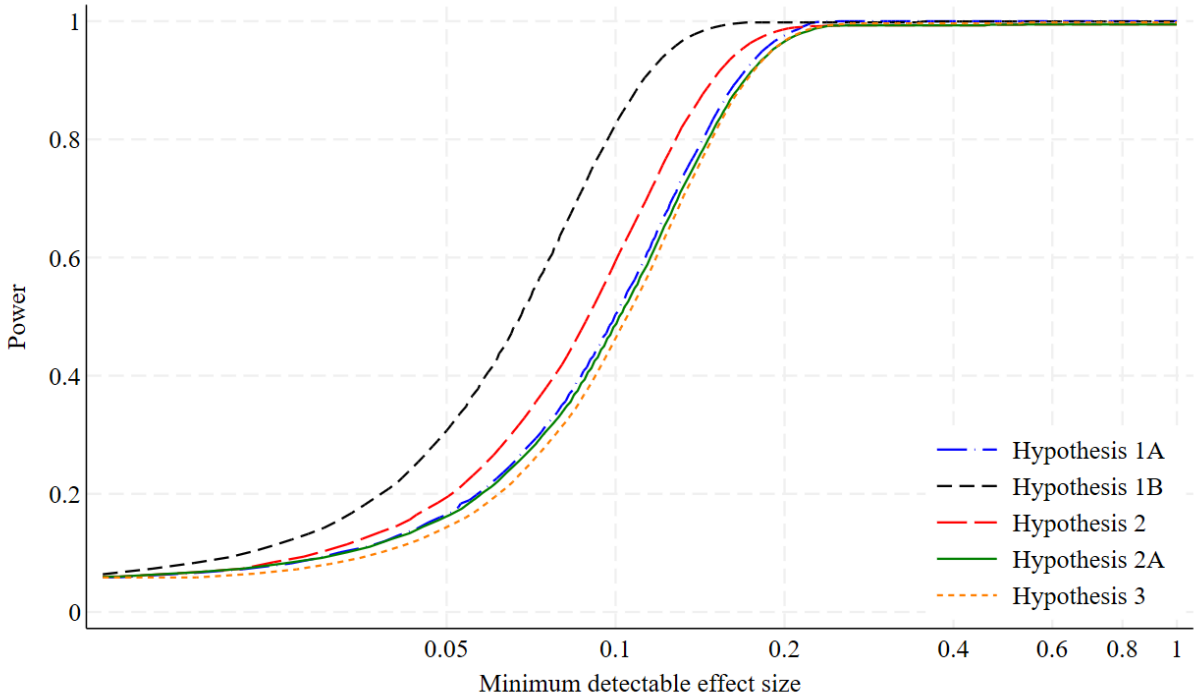


Figure C3. Power as a function of effect size for subhypotheses