

The Impact of Subsidized Access to Nurseries and Employment Services on Mothers’ Labor Market Outcomes and Child Development in Egypt: Pre-Analysis Plan

Stefano Caria, Bruno Crepon, Noha Fadl, Caroline Krafft, AbdelRahman Nagy

September 7, 2021

1 Introduction

The low participation of women in Egypt’s labor market is often attributed to a high opportunity cost of time, limited access to jobs and restrictive gender norms ([2]; [12]; [30], [31]). In this project, we will investigate these constraints and their interaction, specifically: (i) does lowering the cost of childcare increase the labor supply of economically vulnerable mothers? (ii) are childcare subsidies more effective when combined with employment services and when gender norms are more favorable to female employment? (iii) what is the impact of shifting to nursery care on children’s development?

We will use a randomized control trial in low-income communities in Egypt, a context where women’s employment is particularly low, just 17 percent [17]. While childcare costs are considered a key constraint on female employment in developing countries [15]; [16], the experimental evidence on the impacts of alleviating this constraint is scarce. Also, there is little research on how reducing the barriers to childcare may interact with labor demand and gender attitudes. We will cross-randomize two interventions targeting mothers with young children: childcare subsidies and employment services. We will study the direct impacts of these interventions on mothers’ labor market outcomes and children’s development.

2 Motivation and Conceptual Framework

2.1 Motivation

Globally, female labor force participation (FLFP) has remained stubbornly low. From 1980 to 2009, global FLFP inched from 50.2 percent to 51.8 percent [29]. Progress in developing countries has been uneven. Countries in the Middle East and North Africa region have the lowest rates of FLFP in the world [29]. Despite rising educational attainment, FLFP has declined in Egypt. In 2018, only 17 percent of women were employed [17]. There cannot be gender equity in employment in developing countries without progress in worst-performing countries such as Egypt.

2.2 Conceptual Framework

There are three main explanations as to why FLFP globally and in Egypt remains low: (1) women’s high opportunity cost of time (determined by childcare responsibilities) (2) weak labor demand and (3) restrictive gender norms ([2]; [15]; [16]; [30], [31]). We will test aspects of all three constraints on FLFP. Inability to pay for childcare is a major constraint. In Egypt, only 8 percent of children ages 0 to 4 years are enrolled in registered nursery programs ([26]). Private nursery costs average 324 Egyptian pounds (EGP) per child per month. With a fertility rate of 3.1 births per woman [17], after paying for childcare women earning the median private sector wage of 1000 EGP have net earnings of -4 EGP per month; the opportunity cost of time precludes employment. Indeed, when asked why they are not working, 69 percent of women say this is because of household responsibilities.

Access to jobs is a second important constraint. Employers may be reluctant to hire female workers because they expect that domestic responsibilities will make them less productive and harder to retain ([21], [8]). In Egypt, 60 percent of employers in a recent survey by Osman et al (forthcoming). openly reported that they prefer to hire men. Labor demand is also generally weak ([3]), which exacerbates women-specific constraints.

Gender norms are a further constraint on women’s work. In Egypt these norms are often not in favor of equity in employment ([18]). However, there is significant heterogeneity in attitudes to female employment and recent evidence suggests gender norms are malleable ([5], [10]).

The three constraints interact and reinforce each other. For example, inability to pay for childcare can reinforce employers’ worries about female employees. A unique aspect of our research is that our design will enable us to capture the dynamic interaction of the multiple barriers women face in the labor market.

3 Sample and Interventions

3.1 Samples

3.1.1 Nurseries

Our experiment will take place in low-income neighborhoods in Greater Cairo. Within these low-income neighborhoods, we will identify nurseries registered with the Ministry of Social Solidarity (formal nurseries). We will offer these nurseries the opportunity to participate in our voucher experiment (to accept vouchers that subsidize part of the cost of care). If they agree to participate, we will survey each of these nurseries to gather capacity data for our child care voucher experiment, specifically:

- Nursery capacity in terms of number of slots, c_n , for nursery n .
 - We will pay specific attention to the current COVID-19 health crisis and policy response, which affect the capacity of nurseries. For example, nurseries were, as of early 2021, only allowed to operate at 50 percent capacity.
- The current number of slots occupied at the nursery o_n
- We will calculate the supply of available slots at the nursery $s_n = c_n - o_n$
- We will then generate a local target number t_n that we will define as three-quarters of the number of slots locally available: $t_n = 0.75s_n$.
 - This target number will be the number of households we register in the experiment in the catchment area around the nursery.
 - While only half of households will be offered subsidies, and not all will take up subsidies, households who do take up subsidies may have multiple children, so registering households at a level below capacity will ensure that nursery slots are available locally.
- The survey will also ask the monthly fee to attend the nursery, f_n .

3.1.2 Households

Our objective is to register in the study 5,000 households with women who have at least one child between the ages of 1 to 5 years old, living in the catchment area of the nurseries included in the experiment, and who are not yet a client of a nursery.

The procedure to identify and recruit these women is the following:

- The catchment area is defined by a 2km radius around each participating nursery.
- In cases where there are multiple nurseries with overlapping catchment areas we will combine the catchment areas and sum the household sample targets t_n .

- We will use the Facebook population projections to identify the GPS locations (pixels) where children aged 0 to 4 in 2020 lived and the number of such children [13]. These children will be 1 to 5 in 2021 when we are running the experiment.
- We will then identify the pixels that exist within our nursery catchment areas. Based on these identified pixels, we will evaluate the total number of children aged 0 to 4 in 2020. This will give us an estimate of the potential demand for slots.
- We will then draw (sample) points (pixels, which are locations with GPS coordinates) in each catchment area in a random order, probability proportional to child population.
- We will visit the nearest residential building to each selected point, check whether they meet the eligibility conditions and then register them if eligible.
- We will continue to register households in the catchment area until we reach the target t_n .
- We will continue to add nurseries and catchment areas until we identify 5,000 eligible women and their households.

3.2 Interventions

In our experiment, we will evaluate the impacts of two interventions: (i) childcare subsidies: mothers assigned to this intervention will be eligible for a subsidy on the price of a local NGO nursery, for one year. We will offer subsidies of 25 percent and 75 percent of median nursery cost; (ii) employment services: mothers assigned to the employment services intervention will be connected to local employers providing formal jobs. Mothers will be informed at the end of the baseline survey visit: (1) of whether and the level of subsidy they have been randomized into and (2) whether they have been randomized into employment services.

3.2.1 Subsidies

Subsidies will cover all children aged 0-5 in the household (fixed subsidy per child not per household). Subsidies can be used at any participating NGO nursery. Mothers randomized into the subsidy will be given the name(s) and contact information of nearby local participating nurseries.

The subsidy will be administered as follows:

1. At the end of the household survey, if she is randomized into a subsidy treatment, the mother is given a unique code corresponding to the level of support she is entitled to.
2. The mother goes to the nursery she wants to use and fills out a subsidy confirmation form provided to the nursery by JPAL. She must provide the code + ID card + signature / fingerprint.
3. Nurseries can call the partner NGO and confirm the names and ID numbers of mothers and the type of support they are receiving.
4. The nursery takes the children's attendance every day.
5. The partner NGO will make random visits in every nursery to ensure that the children who have subsidies are attending regularly.
6. The partner NGO transfers the support amount to each nursery at the end of each month.
7. There will be research team support in all these steps.

3.2.2 Employment services

To offer employment services we will partner with several recruitment platforms currently active in Egypt such as Shagahalni, National Employment Pact, Forasna and other similar platforms. These platforms work with firms that have vacancies to fill. The basic service they offers is simply to post the vacancies on their dedicated website, social media accounts and job fairs. Individuals searching for a job register to the platform, attend job fairs and they can directly apply through these platforms to the vacancy.

These platforms also offer placement services linking the firm and the job seeker. They have a large pool of operators on staff who are in charge of this matching process. To facilitate matching, at baseline (before randomization), all women will be asked about their labor market status and their criteria for potential jobs in terms of geographical location, occupation, wage and work hours. These criteria will help identify a set of firms who are likely to have vacancies fitting the women's search criteria. For each woman randomly assigned to be offered placement services, our partners will search in the pool of vacancies posted by these firms for vacancies that are suitable. Operators from the platform will then call back the mother within three weeks from conducting the baseline survey and once the potential matches are identified and propose a minimum of three vacancies. Operators will register the vacancies, if any, the women is interested to apply to and offer to organize a meeting. On a regular basis, the placement firms' operations teams will track all the mothers' responses, interviews and placement, then they will share this data with the research team. The research team will do back checks with at least 5 percent of mothers to check with mothers about this process. In case mothers did not accept the job opportunities, the placement services firms will offer up to three different job opportunities and will follow up three times for each set of job opportunities. For a sample of mothers who do not accept a job opportunity, the research team will conduct a 5 minute phone interview to understand the reasons behind their decision.

3.3 Randomization

We will randomize both the subsidy and employment services interventions at the level of the individual mother. We will assign one fourth of the sample (about 1250 mothers) to pure control (no subsidy and no employment services), one fourth of the sample to childcare subsidy but no employment services (evenly split between the two levels of the subsidy), one fourth to employment services but no subsidy, and one fourth to both subsidy and employment services (see figure 1).

Randomization of mothers will happen according to a simple stratification rule. Within each catchment area we will construct blocks of 8 individuals who (i) have been interviewed consecutively, (ii) are identical along the following two dimensions:

1. Age of youngest child (0-2 years old vs. 3-5 years old);
2. Ever having worked or not;

In each block, randomly:

- Two individuals will be assigned to the control group,
- Two individuals will be offered childcare subsidies (one 25 percent, one 75 percent subsidy) and will not be offered employment services,
- Two individuals will be offered employment services and will not be offered childcare subsidies,
- Two individuals will be offered both employment services and childcare subsidies (one 25 percent, one 75 percent subsidy).

This stratification will help ensure balance in terms of age of youngest child and ever worked, two key variables that will shape take up and outcomes. Using blocks of eight individuals rather than four will reduce the probability enumerators can guess the assignment. This stratification will be done on the tablet to allow randomization into interventions at the end of the baseline interview.

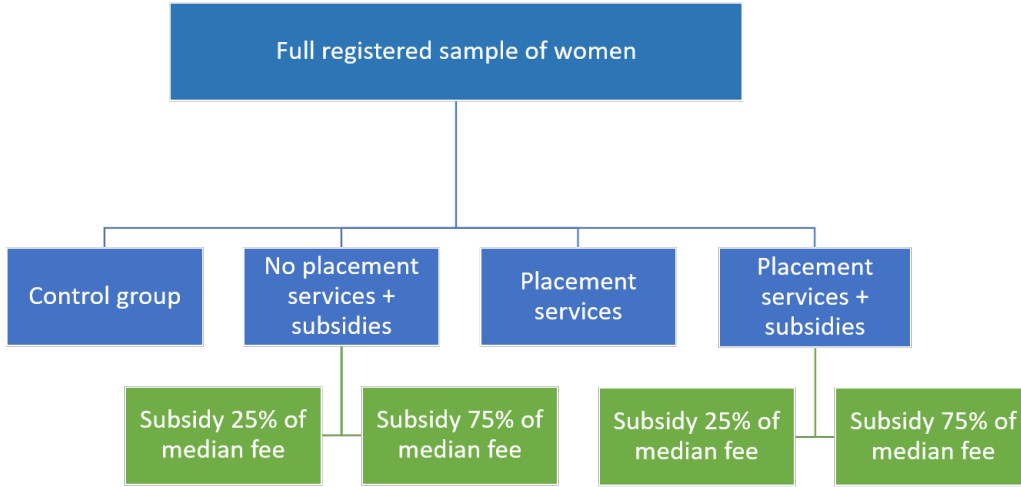


Figure 1: Experimental design

3.4 Risks

Our project poses minimal risk to participants. Our project will be reviewed primarily by the American University of Cairo Institutional Review Board (IRB), and subsequently by co-PI institutions as needed as well as Egypt’s Central Agency for Public Mobilization and Statistics (CAPMAS). The data collection involves no more risk than is typical for standard household survey questions on employment, gender role attitudes, or early childhood development. Care will be taken to minimize COVID-19 risks, including use of masks and sanitizers by enumerators and efforts to undertake interviews outdoors whenever possible. Some participants will be randomly selected to benefit financially from employment services or a child care subsidy.

One concern would be that the childcare provided may potentially be of low quality. We think that this is unlikely to be the case: all the nurseries are registered NGOs supervised by the Ministry of Social Solidarity. Our initial qualitative evidence suggests that the NGO nurseries that we are working with offer better services than available alternatives. Our expectation is thus that, in our context, attending a nursery will have a beneficial effect on children’s development. However, if this was not the case, there would still be substantial value to society from learning this and being able to use our evidence to mandate improvements in nursery care going forward.

A second concern is related to potential congestion in nurseries. The nurseries we are working with are currently under capacity and given our modest sample size in each nursery catchment area corresponding to availability of nursery capacity, we think that congestion is highly unlikely. Finally, employment effects may partly come from displacement. Given the size of the labor market, we do not expect this will be a first order concern. We expect that increases in demand for female workers are also likely as a result of the experiment and have constructed an experimental design that can credibly measure these effects.

NGO implementation biases may also be an issue. In some cases the NGOs’ staff may try to find other sources of subsidies for the control group. They could also try to have informal agreements with all families to use their subsidy fees to reduce the cost for all families within the same nursery. The research team will undertake close follow up for the implementation of the subsidies to ameliorate these risks. In addition, the research team will conduct back checks to better understand how the subsidies are implemented.

Ability to collect data and implement interventions during COVID-19 or any other future pandemic remains a challenge. We can not guarantee that data collection or interventions will be implementable as

planned after the baseline survey. However, we will collect as much contact information data as possible that could help us to contact participants for phone surveys if needed. We are monitoring the nursery availability and safety situation as well as employment services availability to ensure safety and feasibility.

The results of our research will not be used to discriminate or otherwise hurt specific individuals. All adults will consent to participate at each phase of the research, and we will obtain assent from children and consent from parents or guardians when assessing children, as well as following standard best practices for assessing children (e.g. offering neutral praise). Individuals will also be free to withdraw from the study at any point or not to answer any questions during the interviews.

3.5 Data sources

We plan to collect data from households at four points in time: at baseline (right before offering the interventions), four months after baseline (midline 1), ten months after baseline (two months before 12-month vouchers end, midline 2) and 18 months after baseline (six months after subsidies end, endline). The baseline interview will collect information about the mother (particularly her labor supply), her husband, the child's development, and the household's dynamics (gender role attitudes and time use). Midline 1, four months after baseline, and Midline 2, at 10 months, will be short interviews focused on measuring mothers' labor market outcomes. The endline, at 18 months, will measure the same outcomes as baseline. The timing of these two last interviews is chosen so that we will be able to measure outcomes when the nursery subsidy is ongoing as well as when the nursery subsidy is completed. The baseline and endline will be face-to-face, tablet-based interviews, safety permitting. The midlines will be administered over the phone.

The baseline and endline will use similar instruments, excluding time invariant information for households (e.g. date of birth for households). The midlines will be addressed to mothers, and use a subset of baseline/endline questionnaire questions, specifically those on current labor market activities and current child care use.

3.6 Timeline

- Baseline surveys of households (Months 1-2)
- Implementation of interventions (Months 1-13)
- Midline 1 survey of mothers (Months 4-5)
- Midline 2 survey of mothers (Months 10-11)
- Endline (post-subsidy) survey of households (Months 18-19)

4 Hypotheses and Outcomes

4.1 Primary outcomes

Our main hypothesis is that both childcare subsidies and employment services increase mothers' work activities, including wage employment and self-employment. In the short-run, mothers who receive subsidies may start work but earn lower average (conditional) wages. However, these effects may revert in the long-run, as experience accumulates and mothers are able to progress to better positions. Furthermore, it is possible that average (unconditional) earnings among all mothers increase also in the short-run, driven by increases in employment.

We are also interested in whether the interaction of these interventions generates additional employment gains. There are two ways in which this can happen. First, some mothers may experience both barriers to employment and thus may be induced to take up wage employment only by the combination of childcare subsidies and employment services. For example, a mother disconnected from the labor market may decrease her reservation wage when she receives a subsidy, but may nevertheless struggle to find open vacancies at this lower wage. Second, employment services may help mothers to fully capture the labor market gains associated to childcare subsidies. For example, placement services might allow women to be more selective and to end up with a better job.

H_a The subsidies, employment services and their combination will increase mothers' employment rates, hours, and (unconditional) earnings.

To study this hypothesis, we will investigate impacts on the following outcomes:

1. Work activity: Dummy variable for the mother spending at least one hour on remunerated work in the past 7 days. Work can include formal wage employment (with social insurance), informal wage employment (without social insurance), or self-employment.
2. Hours of work: Total number of hours of remunerated work over the past 7 days that the mother spent on work (in total).
3. Labor income: Total income over the past 7 days earned by the mother.

⇒ **Family 1** Our first family of primary hypotheses comprises outcomes (1)-(3). We will measure each outcome at each midline interview, and we will pool both midline observations in the same regression.

4.2 Mechanisms

4.2.1 Childcare

Our next family of hypotheses concerns the mechanisms behind the interventions. In particular, we hypothesize that, when offered childcare subsidies, households adjust the organization of childcare, using more nursery care and less family care.

H_b Treated households will use more nursery care and less family care

We will test this hypothesis by studying impacts on the following outcomes:

1. Total hours children aged 1-5 spend in NGO nursery
2. Total hours children aged 1-5 spend in other formal care
3. Total hours children aged 1-5 spend in informal care (home nursery, neighbor, close family member etc.)
4. Total hours children aged 1-5 spend under father's care
5. Total hours children aged 1-5 spend under mother's care

⇒ **Family 2** Our second family of hypothesis comprises outcomes (1)-(5). These outcomes will be measured at both midlines. We will pool observations from both midlines in the regressions.

4.2.2 Reservation wages and reservation job quality

Our next family of hypotheses focuses on a second set of mechanisms related to the types of jobs that treated mothers will accept.

H_c Childcare subsidies, employment services, and their combination will lead mothers to change reservation job quality and reservation wages.

We will study this hypothesis by estimating impacts on the following outcomes

1. Reservation wage for private sector job
2. Reservation job quality 1: maximum commuting time
3. Reservation job quality 2: requires flexible working hours
4. Reservation job quality 3: requires ability to take time off work at short notice (paid leaves)

5. Reservation job quality 4: requires childcare facility at place of work
6. Reservation job quality 5: requires part-time work
7. Targeted occupation is a white-collar occupation
8. Number of targeted occupations

⇒ **Family 3** Our third family of hypotheses will comprise outcomes (1)-(8). These outcomes will be measured only at the first midline interview.

4.2.3 Job search effort

Our next family of hypotheses focuses on a third and final set of mechanisms related to job search effort.

H_d Childcare subsidies, employment services, and their combination will lead mothers to increase their job search effort.

We will study this hypothesis by estimating impacts on the following outcomes

1. A dummy indicating whether they have done any job search activity since the previous interview
2. Number of applications since the previous interview
3. Number of interviews invited to
4. Number of interviews attended

⇒ **Family 4** Our fourth family of hypotheses will comprise outcomes (1)-(4). These outcomes will be measured only at the first midline interview.

4.3 Downstream outcomes

We then have a set of hypotheses about the downstream effects of the interventions.

4.3.1 Endline employment

First, we hypothesize that the employment effects will persist after the end of the interventions:

H_e The subsidies, employment services and their combination will increase mothers' employment rates, hours, and (unconditional) wage earnings for mothers at endline.

1. Work activity Dummy variable for at least one hour of remunerated work over the past 7 days.
2. Hours of work Total number of hours of remunerated work over the past 7 days.
3. Labor income Total income over the past 7 days earned by the mother.

⇒ **Family 5** This family will include outcomes (1)-(3), measured at endline.

Note that we will also provide separate regressions to break down these impacts on remunerated work in their three underlying component: formal work, informal work, self-employment. In this analysis, we will also include non-remunerated family work.

4.3.2 Endline job quality and net labor income

H_f Childcare subsidies, employment services and their combination will change job quality and net labor income.

We will investigate this hypothesis by computing the following outcomes:

1. Commuting time higher than control median¹
2. Has flexible working hours
3. Ability to take time off work at short notice (paid leaves)
4. Childcare available with employer
5. Labor income net of transport costs, meal expenses for work lunch, and childcare costs

⇒ **Family 6** This family will consider outcomes (1)-(5), measured at endline.

Note that we will also report separate regressions reporting changes in each specific cost category: (i) transportation undertaken for work purpose during the past 7 days, (ii) meal expenses for lunch taken at or to work, (iii) childcare costs, for each childcare type category reported in family 2.

4.3.3 Welfare

Next, we will investigate impacts on mother welfare.

H_g Childcare subsidies, employment services and their combination will improve mother's welfare.

We will measure **welfare** with these two outcomes:

- (1) Mothers' perceived food insecurity The questionnaire includes a series of questions intended to detect food security as well as the impact of COVID-19 on food security. The questions ask: In the past 7 days, have you or any household member experienced any of the following? (select all that apply)
(1) Difficulties in going to food markets due to mobility restrictions imposed by government/closures
(2) Unable to buy the amount of food we usually buy because of shortages of food in markets
(3) Unable to buy the amount of food we usually buy because the price of food increased
(4) Unable to buy the amount of food we usually buy because our household income has dropped
(5) Had to reduce the number of meals and/or the portion of each meal we would usually eat. While (1) and (2) are about the availability of food, (3), (4), and (5) could be affected by changes in employment and income brought about by our intervention. We will therefore test for changes in each of these secondary outcomes.
- (2) Mother psychological well-being While employment services and child care subsidies may decrease maternal depression, it is also possible that additional employment creates stress and difficulties combining employment and domestic responsibilities could actually increase maternal depression. We will use the WHO-5 well-being index to screen for maternal depression [24]. This is one of the most widely used measures of subjective well-being. The WHO-5 scale asks "Over the past two weeks..." for five questions such as "I have felt cheerful and in good spirits" with responses of (5) All of the time (4) Most of the time (3) More than half the time (2) Less than half the time (1) Some of the time (0) at no time. The WHO-5 will be summed for an overall raw score and multiplied by four to scale from 0 (worst possible well-being) to 100 (best possible well-being). A cut-off score of 50 is commonly used to screen for depression [24]. We will use both the overall raw score, as a potentially more sensitive measure of maternal well-being, as well as a binary measure of maternal depression, as a particularly important secondary outcome.

⇒ **Family 7** will include outcomes (1) and (2), measured at endline.

¹For outcomes 1-4, if a respondent does not have a job, the variable takes a value of 0. If there are significant results, we will undertake bounding exercises to calculate the intensive versus extensive margins.

4.3.4 Empowerment

Next, we will investigate impacts on mother’s empowerment. This is a complex and contested concept [4, 14]. We will use the common measures of mobility and self-efficacy as our two metrics of empowerment.

- Mobility The mobility measures are a series of questions on ”If you need to go to any of the following places, can you go on your own without permission or do you need to inform someone or get permission or you just can’t go?” We will undertake exploratory factor analysis to create a standardized factor from four variables for different destinations (the local market; going to the doctor for treatment; bringing children to the doctor; visiting home of relatives, friends, or neighbors) with the responses (4) Go alone without permission (3) Go alone after informing them (2) Go alone, but must be granted permission first and (1) Cannot go alone. Higher scores will indicate higher levels of mobility.
- Self-efficacy The self-efficacy questions are the ten-question generalized self-efficacy scale, asking ”How often do you feel...” for a variety of statements, such as ”I can usually handle whatever comes my way.” Responses range from from (1) No (2) Scarcely (3) Often (4) Always. Higher scores will indicate higher levels of self-efficacy. We will undertake exploratory factor analysis to create a standardized factor from these variables.

⇒ **Family 8** will include outcomes (1) and (2), measured at endline.

4.3.5 Time use at endline

Nurturing care is critically important to early childhood development [9]. Changes in maternal employment or use of nurseries may change the amount of time children receive nurturing care at home. We will examine maternal time use as an outcome for mothers and key mechanism for effects on children’s development. We focus on mothers, since men provide effectively nil caregiving even when their wives work in the context of Egypt[27].

H_h Childcare subsidies and employment services will change the amount of time women spend on childcare, chores, and leisure.

We have a 24 hour time use history for mothers in the baseline and endline questionnaires. The time use module includes both primary and secondary activities, which are critical for capturing caregiving. We will use hours engaged in specifically ”childcare and instruction” as a primary or secondary activity (the sum of the two) as our outcome. We will also contextualize any changes in caregiving time with corresponding changes in time spent in nursery or other care, since the two may be substitutes, albeit of potentially varying quality.

⇒ **Family 9** will include 1) hours per day of child care 2) hours per day of leisure 3) hours per day of chores

4.4 Primary outcomes and hypotheses: Children

An important question about using child care subsidies to increase women’s employment is whether increases in nursery care are beneficial, harmful, or neutral in their effects on children. We hypothesize:

H_i Increased usage of child care services will result in development gains for children.

The tools we use to measure children’s development depend on their age at the time of each wave. For children aged 0-35 months we use the Caregiver-Reported Early Development Instruments (CREDI). For children aged 36-83 months we use the Measure of Development and Early Learning (MODEL) direct assessment and parent/caregiver report tools.

4.4.1 Children’s development ages 0-35 months

The CREDI short form is designed to measure ECD across five dimensions for ages 0-35 months. Dimensions include motor, cognitive, language, socio-emotional, and mental health skills. The CREDI short form has been validated with results from 8,022 children across 17 low-, middle- and high-income countries, including in Arabic in Jordan and Lebanon [20]. Caregivers are asked 20 yes/no questions about children’s development within age-specific six-month bands (e.g. 12-17 months). An example item would be “Can the child follow simple directions (e.g., ‘Stand up’ or ‘Come here’)?” The results of the CREDI provide a single score across all domains [20]. As per the scoring manual [19], a raw scaled score (percentage of age-specific 20 items that are a yes) will be used as the child development outcome for children aged 0-35 months.

4.4.2 Children’s development ages 36-83 months

The Measuring Early Learning Outcomes and Quality (MELQO) tools include the MODEL measures of ECD for children aged 3-8 years (pre-primary aged children, globally) [25]. We will use these tools for children aged 36 months to 83 months (children would enter primary school in Egypt once aged six). There are specifically instruments for child direct assessment and parent/caregiver reports of development. Both instruments cover dimensions of: executive function, socio-emotional skills, and pre-academic skills (literacy and math).

The MODEL instruments have been validated across low- and middle- income contexts and used in more than 13 countries [22]. The measures are designed to be summarized through factor analysis. Previous analyses indicated all developmental dimensions loaded on to a single factor of ”school readiness” [22]. We will therefore undertake exploratory factor analysis to confirm this structure and factor scores will be used as the primary developmental outcome for ages 36-83 months. The MODEL tools had previously been adapted to the Egyptian context for a study of learning in pre-primary, and the adapted tools will be used for this study.

4.4.3 Children’s development ages 0-83 months

In order to estimate the impact of the interventions on children aged 0-83 months overall, we will also calculate the percentage of items correct (for the direct assessment) or yes (for the parent/child report) and create a raw scaled score combining the CREDI and MODEL measures. If significant effects are found for children overall, we will explore which specific items may be driving results.

⇒ **Family 9** will include 1) CREDI raw scaled score 2) MODEL raw scaled score 3) combined raw scaled score. All of these will be measured at endline.

5 Estimation methodology

5.1 Main effects-Intent to treat on individuals

In our study we implement a two sided multi-arm experiment. Households are registered in the catchment area of nurseries (see section 3.1.2) and they are randomly assigned to be offered different levels of subsidies (see section 3.2.1): 0, 25 or 75% and offered placement services (see 3.2.2). This defines a set of household variables that we denote as follows:

- V^{25} is equal to one for households who are offered a childcare subsidy corresponding to 25% of the median fee
- V^{75} is equal to one for households who are offered a childcare subsidy corresponding to 75% of the median fee
- E is equal to one for households who are offered employment services

The main analysis of the data collected at the individual level from our household surveys will be based on the estimation of the following intention-to-treat regression equation, in which S denotes the set of randomization strata dummies and i denotes the household:

$$y_i = \alpha + \beta_1 E_i + \beta_2 V_i^{75} + \beta_3 V_i^{25} + \beta_4 E_i * V_i^{75} + \beta_5 E_i * V_i^{25} + \text{selected}(x_i) \beta_l + \sum_k \delta_k S_i^k + \varepsilon_i$$

This model includes a set of control variables chosen using the double post lasso method proposed in [7] which has the advantage to automatically select the relevant subset of variables, avoiding specification search.²

We will identify neighborhoods in greater Cairo that correspond to our target criteria (see section 3.1.1). Within these areas we aim to consider the largest possible sample of NGO nurseries and actually will consider including them all. Within the corresponding catchment areas, we will randomly assign households to the different treatment groups. As a result, following [1], we will not cluster standard errors at the catchment area level and use the simple Eicker-Huber-White robust standard error. In models where we pool more than one wave, we will cluster standard errors at the level of the individual respondent.

There are several hypotheses we want to test. We will implement these tests for our outcome variables y :

- H1: No impact of the employment service intervention in the absence of the childcare subsidies: $\beta_1 = 0$.
- H2: No impact of childcare subsidies in the absence of employment services: $\beta_2 = \beta_3 = 0$.
- H3: No differential impact of childcare subsidies depending on the amount of the voucher: $\beta_2 = \beta_3$.
- H4: No interaction between childcare subsidies and employment services: $\beta_4 = \beta_5 = 0$.

5.2 Treatment effect heterogeneity

There are two key dimensions of treatment effect heterogeneity we will explore for our main outcomes:

- Household income (in the baseline questionnaire the question: Last month, what was your household's total monthly income (from all sources)). We will examine treatment effect heterogeneity by baseline income tertiles.
- Gender role attitudes will be measured by 11 questions in the baseline questionnaire. The baseline questionnaire asks whether respondents strongly disagree (1), disagree (2), neutral (3), agree (4), or strongly agree (5) for each of the following: (1) The husband should help his working wife raise their children. (2) The husband should help his working wife with household chores. (3) Girls should go to school to prepare for jobs, not just to make them good mothers and house-wives. (4) A woman who works outside the home cannot be a good mother. (5) For a woman's financial autonomy, she must work and have earnings. (6) A woman's work interferes with her ability to keep a good relationship with her husband. (7) Women should continue to obtain leadership positions in society. (8) Boys and girls should get the same amount of schooling. (9) Boys and girls should be treated equally. (10) When jobs are scarce, men should have more right to a job than women. (11) Girls/women who are harassed deserve it if they are dressed provocatively. We will reverse code (4), (6) and (11) so that all measures indicate more gender equitable values consistently and use factor analysis on these measures to create a continuous index. We will examine treatment effect heterogeneity by baseline gender role attitude index tertiles.

5.3 Threats

5.3.1 Partial Compliance

We will conduct our analysis using mainly ITT parameters. Nevertheless, the interventions we consider consist of offering services. An important first step in our analysis is to describe the take-up in the different

²These are the variables that will be included in the first stage of the double-lasso procedure. First, the baseline value of all variables we consider as outcome in any of our tables. Second, a set of variables corresponding to the marital status, the presence of family members in the close neighborhood, the role of the mother in law in the household decision making, household assets and income, including labor income from the husband, remittances, government transfers, transfers from the family, as well as a set of dummy variables corresponding to the fact that the household i is in the catchment area of nursery c .

treatment and control groups for these services. We will therefore examine in each treatment and control groups the take-up for nursery services, placement services and their possible combinations. We consider five different treatments. Each offers different incentives to participate in three different combinations of services (nursery services alone, employment services alone, both employment and nursery services). This makes the usual analysis in Local Average Treatment effect and the usual notions of Compliers, Never-Takers and Always-Takers complex (see Behaghel et al. 2013). We will therefore limit our analysis of imperfect compliance to the description of the take-up.

5.3.2 Attrition

As a longitudinal study, one risk facing this research is attrition, particularly differential attrition depending on treatment status. To minimize attrition, the research team will work on developing very detailed guidelines for data collection, starting with hiring experienced enumerators, and providing high-quality training and technical support during the data collection process. As part of the data collection, the questionnaires will capture several types of contact data such as the respondent’s mobile phone number, landline, another mobile phone number within the household, and a very exact address. In addition, the enumerators will collect the GPS location using their electronic devices on SurveyCTO during the interview. At the same time, the enumerators will provide small incentives (less than two USD) during each interview in appreciation of the participant’s time and efforts. Enumerators will follow up between 3-4 times with each participant until they complete the survey interview. After following all the previous guidelines, it is expected to interview around 95 percent of the total sample during the baseline, follow-up, and end-line survey. We will report attrition rates for each stage of data collection. If attrition exceeds 5 percent in any treatment arm, we will analyze attrition by the same characteristics as for our LASSO model to identify whether there is significant differential attrition on any dimension across the treatment groups.

5.4 Issues related to multiple testing

Our analysis will be performed over many different variables and our basic specifications also involve several treatments. Multiple hypothesis testing is therefore an important issue. To deal with it we have (i) specified a primary set of hypotheses, and (ii) we will offer corrections for multiple comparisons. We will create an index variable for our primary family of hypotheses (Family 1). We will run four tests on this index (H1, H2, H3, H4). For each test, we will report both standard p values and sharpened q-values that control the False-Discovery-Rate. We will repeat this procedure for each subgroup defined in section 5.2. For all other families and outcomes we will report coefficients, standard errors, and 95 percent confidence intervals, but not undertake any specific hypothesis tests.

6 Power analysis

We estimate that the MDE for either of our main treatments is an effect of 3.2 percentage points. We expect that the treatment effects of both interventions in isolation will be 5 percentage points or more. We will have power of 99 percent to detect these effects. Further, we will have power of 60 percent to detect a 4 percentage points increase in the effectiveness of childcare subsidies due to the combination with employment services. These are conservative estimates, as they do not take into account the potential gains in power obtained by using a double-LASSO procedure to optimally select controls. For these calculations, we assumed a control employment rate of 7% and an autocorrelation in employment of .15. We made these assumptions on the basis of data from the latest Survey of Young People in Egypt (SYPE). The SYPE oversamples individuals in informal settlements and is thus an ideal source of information about the likely outcomes of the mothers in our sample.

References

- [1] Abadie, A., Athey, S., Imbens, G., and Wooldridge, J. (2017) When should you adjust standard errors for clustering? *National Bureau of Economic Research*
- [2] Assaad, R., Krafft, C., and Selwaness, I. (2017). The Impact of Marriage on Women’s Employment in the Middle East and North Africa. Economic Research Forum Working Paper Series No. 1086. Cairo, Egypt.
- [3] Assaad, R., Krafft, C., and Yassin, S. (2018). Job Creation or Labor Absorption? An Analysis of Private Sector Job Growth in Egypt. Economic Research Forum Working Paper Series No. 1237. Cairo, Egypt.
- [4] Ballon, P., and Yalonetzky, G. (2018). Introduction to Special Section: Quantitative Approaches to the Measurement and Analysis of Female Empowerment and Agency. *The Journal of Development Studies*, 54(8): 1279–1283. <https://doi.org/10.1080/00220388.2017.1414191>
- [5] Bargain, O., Boutin, D., and Champeaux, H. (2019). Women’s Political Participation and Intrahousehold Empowerment: Evidence from the Egyptian Arab Spring. *Journal of Development Economics* 141: 102379.
- [6] Behaghel, L., Crepon, B., and Gurgand, M. (2013). Robustness of the Encouragement Design in a Two-Treatment Randomized Control Trial. IZA Discussion Paper No. 7447.
- [7] Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2), 608-650.
- [8] Bertrand, M., and Duflo, E. (2017). Field Experiments on Discrimination. In *Handbook of Economic Field Experiments*, vol. 1, edited by Abhijit Vinayak Banerjee and Esther Duflo, 309–393. The Hague: Elsevier.
- [9] Britto, Pia R., Stephen J. Lye, Kerrie Proulx, Aisha K. Yousafzai, Stephen G. Matthews, Tyler Vaivada, Rafael Perez-Escamilla. (2017). Nurturing care: promoting early childhood . *Lancet*, 389(10064): 91-102.
- [10] Bursztyn, L., Alessandra, L. Gonzalez, and Yanagizawa-Drott, D. (2018). Misperceived Social Norms: Female Labor Force Participation in Saudi Arabia. NBER Working Paper Series 24736, NBER, Cambridge, MA
- [11] Duflo, Esther. (2012). Women Empowerment and Economic Development. *Journal of Economic Literature*, 50 (4): 1051-79.
- [12] El Feki, S., Heilman, B., and Barker, G. (Eds.). (2017). Understanding Masculinities: Results from the International Men and Gender Equality Survey (IMAGES) - Middle East and North Africa. UN Women and Promundo-US.
- [13] Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University
- [14] Glennerster, R., Walsh, C., and Diaz-Martin, L. (2018). *A practical guide to measuring women’s and girls’ empowerment in impact evaluations*. JPAL.
- [15] Jayachandran, S. (2019). Social Norms as a Barrier to Women’s Employment in Developing Countries. WIDER Working Paper No. 2019/74.
- [16] Klasen, S. (2019). What Explains Uneven Female Labor Force Participation Levels and Trends in Developing Countries? *World Bank Research Observer*, 34(2), 161–197.
- [17] Krafft, C., Assaad, R., and Keo, C. (2019). The Evolution of Labor Supply in Egypt from 1988-2018: A Gendered Analysis. Economic Research Forum Working Paper Series No. 1358. Cairo, Egypt.

- [18] Kraff, C., Keo, C., and Fedi, L. (2019). Rural Women In Egypt: Opportunities And Vulnerabilities. Economic Research Forum Working Paper Series No. 1359. Cairo, Egypt.
- [19] McCoy, D. C., Fink, G., and Waldman, M. (2018). *CREDI Data Management and Scoring Manual*
- [20] McCoy, D. C., Waldman, M., CREDI Field Team, and Fink, G. (2018). Measuring early childhood development at a global scale: Evidence from the Caregiver-Reported Early Development Instruments. *Early Childhood Research Quarterly* 45: 58-68.
- [21] Petit, P. (2007). The Effects of Age and Family Constraints on Gender Hiring Discrimination: A Field Experiment in the French Financial Sector. *Labour Economics*. 14. 371-391. 10.1016/j.labeco.2006.01.006.
- [22] Raikes, A. (2018). "Results from Four MELQO Pilot Countries: Evidence of Within-Country Validity and Cross-Country Comparison of Child Development Data." *MELQO : Measuring Early Learning Quality and Outcomes Brief 1*.
- [23] Sieverding, M., and Hassan, R. (2019). Associations Between Economic Vulnerability And Health And Wellbeing In Egypt. *Economic Research Forum Working Paper Series No. 1364*.
- [24] Topp, C. W., Østergaard, S. D., Søndergaard, S., and Bech, P. (2015). The WHO-5 well-being index: A systematic review of the literature *Psychotherapy and Psychosomatics* 84(3):167-176.
- [25] UNESCO. (2017). *Overview: MELQO Measuring Early Learning Quality and Outcomes* UNESCO, UNICEF, Brookings Institution and the World Bank.
- [26] UNICEF Egypt. (2019). Children, Food and Nutrition: Growing Well in a Changing World. Cairo: UNICEF.
- [27] UN Women. (2020). The role of care economy in promoting gender equality. Cairo, Egypt
- [28] Walker, S. P., Wachs, T. D., Grantham-McGregor, S., Black, M. M., Nelson, C. A., Huffman, S. L., Baker-Henningham, H., Chang, S. M., Hamadani, J. D., Lozoff, B., Gardner, J. M. M., Powell, C. A., Rahman, A., and Richter, L. (2011). Inequality in Early Childhood: Risk and Protective Factors for Early Child Development. *Lancet*, 378(9799): 1325–38.
- [29] World Bank. (2012). World Development Report 2012: Gender Equality and Development. <https://openknowledge.worldbank.org/handle/10986/4391> License: CC BY 3.0 IGO.
- [30] World Bank. (2013). Opening Doors: Gender Equality and Development in the Middle East and North Africa. Washington, DC: World Bank.
- [31] World Bank. (2018). Women Economic Empowerment Study. Washington, DC: World Bank.