Active Labour Market Policies in Addis Ababa: Pre-Analysis Plan

Drafted by
Stefano Caria, Marcel Fafchamps, Simon Franklin and Simon Quinn

Fieldwork location:	Addis Ababa, Ethiopia
Fieldwork dates:	March 2014 – August 2015
Date of Pre-Analysis Plan:	12 October 2015

1 Introduction

This document outlines our pre-analysis plan for a series of related labour market interventions in Addis Ababa. The document summarises our experiment, our data and our plan of regressions. We intend to report the results of these interventions in more than one academic paper; this Pre-Analysis Plan therefore serves as an overall summary of the set of main regressions that we intend to run across several papers. We intend to submit this Pre-Analysis Plan to the AEA RCT Registry.

2 Sampling

2.1 Geographic Sampling

We randomised at two levels: the level of geographic cluster, and the level of the individual. To do this, we used the Ethiopian Central Statistical Agency (CSA) Enumeration Areas within Addis Ababa as a sampling frame. We defined geographic clusters in our sample as three adjacent Enumeration Areas.[1] Clusters were selected at random from our sampling frame, with the condition that no directly adjacent clusters could be selected. (This minimises potential spill-over effects across clusters, and ensures that the same streets were not surveyed twice.)

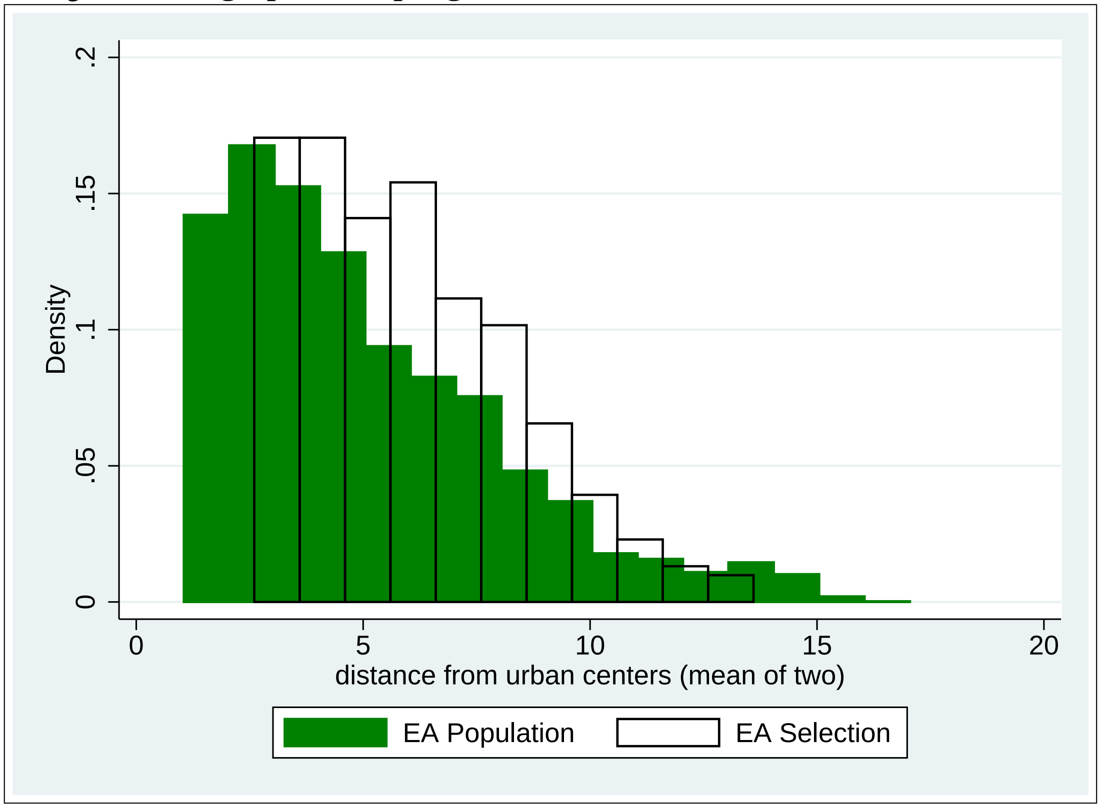
We implemented geographic rules to the sampling. The evaluation aimed to assess the impact of search costs on employment outcomes, so we omitted individuals living within 2.5 kilometres of the city centre. We also excluded from the sampling frame some of the more remote Enumeration Areas of the city, since they were largely undeveloped. Because the centre of the city is relatively dense, and so central clusters areas were more numerous, we also employed a sample probability weighting proportional to distance from the centre of the city to ensure that enumeration areas further away were well represented.

Figure 1 shows (i) the distribution of distances of sampled Enumeration Areas from the centre of the city, and (ii) the distribution of distances from the centre of the city to the Enumeration Areas selected in the sample.[2]

[1] Piloting of the survey listing revealed that individual Enumeration Areas are not populous enough for us to find enough eligible respondents in each.

[2] No distance is equal to zero because we used an average distance from two key central areas in the city.

Figure 1: **Geographic Sampling of Enumeration Areas within Addis Ababa**



2.2 Eligibility and stratification

We used door-to-door sampling in selected clusters, to construct a full listing of individuals fitting our eligibility criteria. Specifically, we required respondents:

(i). To be aged 18 or older, but younger than 30;

(ii). To have completed high school;

(iii). To be available to take work in the next three months; and

(iv). Not to be currently working in a permanent job.

We did not impose any restriction as to current job search behaviour. We listed both men and women.

From this listing, we randomly selected individuals to be involved in the sample. To do this, we used the following stratification rule:

(i). We selected all eligible individuals with university degrees;

(ii). We selected 75% of all individuals with TVET training or diplomas; and

(iii). We selected 25% of all individuals with high school certificates (but without any post-secondary education).

These selection rules enured a stratified sample in which individuals of all education levels, above high school finishers, were well represented.[3] These rules were implemented in the field by enumerator team supervisors, by drawing numbers out of a bag at random. Then all selected respondents were contacted for an interview. If we could not contact the respondent directly, we recorded their details from the head of household and returned to find the person at a later date.

3 Randomisation

3.1 Sample selection and eligibility

We completed baseline questionnaires for 4425 respondents. We then refined the sample in several ways. First, we dropped individuals who violated the inclusion criteria (but who had slipped past the screening procedures). This left 4388 eligible respondents. We attempted to contact individuals by phone for at least a month (three months, on average); we dropped individuals who could not be reached after at least three attempted callls. We also dropped any individual who had found a permanent jobs and who retained the job for at least six weeks. Finally, we dropped individuals who had migrated away from Addis Ababa during the phone survey. Table 1 provides an overview of how many individuals were dropped from the

[3] It was difficult to find better-educated individuals by sampling door-to-door; such individuals were less likely to be at home, and were less easy to approach to in the sample.

sample at each point and the reasons for them being dropped. In all we were left with 4059 individuals who were are included in our experimental study.

Table 1: **Sample selection before randomisation**

	Sample Size	No. Dropped	% dropped
Original baseline	4425		
Eligible at baseline	4388	37	0.84%
Found on phone	4314	74	1.69%
Stayed in phone survey	4254	60	1.39%
Without permanent work	4076	178	4.18%
Stayed in Addis	4059	17	0.42%
Total Dropped		366	8.27%
Final Sample	4059		

3.2 Randomisation

We use a two level randomisation design in order to measure spill over effects within clusters. We randomise at the cluster level first (*i.e.* enumeration areas in Addis Ababa), then among individuals within clusters. (Additionally, we partitioned the data according to baseline interview date. Because phone calls started one month after individuals were surveyed at baseline and the baseline took longer than one month, at the time of randomisation some individuals had not been called. We waited three weeks after randomisation of the first partition to randomise over the second partition.)

3.2.1 Cluster randomisation

Our sample was drawn from a total of 234 clusters in Addis Ababa (117 in each of our time partitions). We began by blocking the clusters according to baseline observables (Bruhn and McKenzie, 2009). We divided each partition into 13 blocks of nine clusters each. This blocking was done by minimising the Mahalanobis distance among clusters within blocks, over the following variables:

(i). Distance of cluster centroid from city centre;

(ii). Total sample size surveyed in the cluster;

(iii). Total number of individuals with degrees;

(iv). Total number of individuals with vocational qualifications;

(v). Total number of individuals working (doing any work in the last 7 days);

(vi). Total number of individuals searching for work (took active steps to find work in the last 7 days);

(vii). Total number of individuals of Oromo ethnicity;

(viii). Average age of individuals in the cluster.

(We worked with *total* numbers in clusters — rather than cluster *averages* — in order to maximize balance between treatment and control among the individual sample.)

Having assigned each cluster to a block, we randomly assigned individuals within clusters to different treatments, with the following division:

(i). Three clusters to the Transport Treatment;

(ii). Two clusters to the Screening Treatment (only screening);

(iii). One cluster to the Fairs Treatment (only fairs);

(iv). One cluster to the Fairs plus Screening treatment; and

(v). Two clusters to no treatment (control clusters).

Table 3 (Column 4) shows the breakdown of the number of clusters assigned to each treatment group at the end of the randomisation.

3.2.2 Individual randomisation and saturation

For those in the transport and pure screening treatment, some individuals within treated clusters were designated not to receive treatment. Among those in the transport clusters, we implemented a randomised saturation design: we randomly assigned saturation levels of 20%, 40%, 75% and 90%. Table 2 shows respectively the number of clusters (Column 3) and the number of individuals (Column 4) who were assigned to those saturation levels. Columns 1 and 2 give an overview of how many individuals were assigned to treatment and control in the final randomisation, by saturation level.

Table 2: **Randomised Saturation Levels for the Transport Treatment**

	(1)	(2)	(3)	(4)
	Individuals		Clusters	
Proportion Treated	Controls	Treated	Count	Proportion
20 %	256	65	18	24.32 %
40 %	150	96	15	20.27 %
75 %	56	191	15	20.27 %
90 %	38	422	26	35.14 %

We did not randomly saturate the screening treatment clusters; however, we did designate some respondents in those clusters not to receive treatment — in order to measure spill-overs. Specifically, 80% of all individuals in screening clusters were assigned to receive the treatment. For the other treatments, all individuals in treated clusters were assigned to receive the treatment.

Having set these cluster saturation levels, we assigned individuals within clusters to treatment and controls to reflect those proportions. This was done by blocking individuals within clusters by their education level, and implementing a simple randomisation rule. The final assignment to treatment is outlined in Table 3.

Table 3: **Final Assignment to Treatment for Individuals and Clusters**

	(1)	(2)	(3)	(4)
	Individuals			Clusters
Cluster Treatment	Control	Treated	Total	Total
Transport	500	774	1,274	74
Screening	187	768	955	56
Fairs	0	493	493	28
Fairs + Screening	0	514	514	28
Control	823	0	823	48
Total	1,510	2,549	4,059	234

3.3 Re-randomisation

In order to ensure balance over treatment and control, we repeated our randomisation algorithm until we had no imbalance over key baseline covariates. We follow Imbens (2011) by re-randomising with a defined objective. Specifically, we re-randomised with the following procedure:

(i). For any given assignment to treatment, we run a regression of each key outcome on the main treatment group dummies.

(ii). We calculate the p-value for a Wald test (F statistic) of the joint null hypothesis that all treatment group dummies are equal.

(iii). Further, we record the p-value for the individual t-test of balance between each treatment and the control group.

(iv). We re-randomised until we found an assignment for which no p-value from any Wald test was below 0.3, *and* for which no p-value from any pairwise t-statistic was below 0.1.

(We ran this procedure separately for each time partition. For the second partition, we ran balance tests using the combination of the first and second samples.)

We performed these tests of balance in the re-randomisation procedure over the following individual-level outcomes:

Table 4: **Variables Used for Re-Randomisation**

VARIABLE	DEFINITION	SOURCE (QUESTION NUMBER)
`degree`	Dummy: Individual has finished a degree (bachelors or above) at a recognised university	Dummy: b5=20 or b5=21
`vocational`	Dummy: Individual has finished a course or vocational training at an official vocational college or TVET	Dummy: b5 $\in \{9, \ldots, 16\}$
`work`	Individual has had any work for pay in the last 7 days	Dummy: j1_1 = 1
`search`	Individual has taken any active step to find work in the last 7 days	Dummy: s0_2 = 1
`post_secondary`	Individual has any kind of non-vocational post-secondary education (degree or diploma)	Dummy: b5 $\in$ $\{17, \ldots, 21\}$.
`female`	Respondent is female	Dummy: respondent_gender = 2
`migrant_birth`	Respondent was born outside of Addis Ababa and migrated since birth	Dummy: b14!=10
`amhara`	Respondent is ethnically Amhara	Dummy: b21=1
`oromo`	Respondent is ethnically Oromo	Dummy: b21=2
`work_wage_6months`	Individual has worked for a wage at any point in the last 6 months	Dummy: j2_1 =1
`married`	Individual is married	Dummy: b1 = 1
`live_parents`	Respondents lives with his/her mother or father	Dummy: b22= 3 or b22= 4
`experience_perm`	Respondent has work experience at a permanent job	Dummy: b22= 3 or b22=4
`search_6months`	Respondent has searched for work any time in the last 6 months	Dummy: s0_1 = 1
`age`	Respondent age	respondent_age

`years_since_school`	Years since the respondent finished school (any school including university)	Constructed from j0_3 ($= 2006 - j0\_3$)
`search_freq`	Proportion of weeks that individual searched for work (from the phone surveys)	Mean (over first 3 months of calls) of Dummy: $p1\_14 = 1$
`work_freq`	Proportion of weeks that the individuals worked (from the phone surveys)	Mean (over first 3 months of calls) of Dummy: $p1\_3 \neq 0$

Conditional on being in the transport treatment, we randomly varied the start and end dates of the treatment within a window of 4–5 weeks on either end, such that individuals received a minimum of 13 and maximum of 20 weeks of access to the transport subsidies. The individuals invited to the screening were invited in a random order.

Below we present summary statistics from the key variables used in the blocking and re-randomisation (defined in the table above), as well as the p-value of the Wald test of the joint hypothesis.

Table 5: **Summary for Blocking/Re-Randomisation Variables**

	N	**Mean**	**S.Dev.**	**1st Q.**	**Median**	**3rd Q.**	**Min.**	**Max.**	**p (F)**
`degree`	4059	0.18	0.39	0.00	0.00	0.00	0.0	1.0	0.813
`vocational`	4059	0.43	0.49	0.00	0.00	1.00	0.0	1.0	0.970
`work`	4059	0.30	0.46	0.00	0.00	1.00	0.0	1.0	0.658
`search`	4059	0.50	0.50	0.00	0.00	1.00	0.0	1.0	0.983
`post_secondary`	4059	0.25	0.43	0.00	0.00	1.00	0.0	1.0	0.981
`female`	4059	0.53	0.50	0.00	1.00	1.00	0.0	1.0	0.999
`migrant_birth`	4059	0.36	0.48	0.00	0.00	1.00	0.0	1.0	0.820
`amhara`	4059	0.44	0.50	0.00	0.00	1.00	0.0	1.0	0.435
`oromo`	4059	0.25	0.43	0.00	0.00	1.00	0.0	1.0	0.389

work_wage_6months	4059	0.45	0.50	0.00	0.00	1.00	0.0	1.0	0.538
married	4059	0.20	0.40	0.00	0.00	0.00	0.0	1.0	0.604
live_parents	4059	0.53	0.50	0.00	1.00	1.00	0.0	1.0	0.875
experience_perm	4059	0.13	0.33	0.00	0.00	0.00	0.0	1.0	0.590
search_6months	4059	0.75	0.43	0.00	1.00	1.00	0.0	1.0	0.716
age	4059	23.53	3.00	21.00	23.00	26.00	18.0	29.0	0.742
years_since_school	4057	3.57	2.96	1.00	3.00	5.00	0.0	16.0	0.515
search_freq	4059	0.57	0.32	0.33	0.60	0.83	0.0	1.0	0.868
work_freq	4059	0.33	0.38	0.00	0.20	0.67	0.0	1.0	0.953

For ease of notation, we will generate copies in Stata of every variable reported in Table 5 (*i.e.* being the same variables reported in Table 4); for each copy, we will use the prefix '`balance_`'. For example, we will create a copy of the dummy variable for having a degree as:

```
gen balance_degree = degree
```

4 Description of the interventions

4.1 Transport subsidy

Individuals in this treatment group were offered a reimbursement for travel expenses. The reimbursement was available up to three times a week. The amount of the reimbursement that could be collected on any of these days was calibrated to the cost of a return bus fare from the area of residence of the participant at baseline to a central location in the city. This location is close to where many of the city's firms are based, and close to the major public job vacancy boards. It is also near a central bus station, from where participants coul reach virtually any area of the city with a direct bus ride.

The median subsidy was 20 ETB (1 USD at the exchange rate at the beginning of the intervention), the minimum was 15 ETB (0.75 USD) and the maximum was 30 ETB (1.5 USD).

We staggered the start time and the end of time of the subsidy, randomly. Respondents were assigned to a start weeks and end weeks in the following way:

Table 6: **Assignment to Transport Start and End Weeks**

	End Week (2014-2015)						
Start Week (2014)	22-Dec	29-Dec	05-Jan	12-Jan	19-Jan	26-Jan	Total
01-Sep	12	11	14	13	0	0	50
08-Sep	12	21	38	29	0	0	100
15-Sep	6	10	12	22	0	0	50
22-Sep	10	15	27	24	0	0	76
29-Sep	16	23	29	78	25	29	200
06-Oct	0	0	0	53	51	46	150
13-Oct	0	0	0	59	44	45	148
Total	56	80	120	278	120	120	774

4.2 Screening intervention

Individuals in this treatment group were invited to take part in a series of personnel selection tests. The results of the tests were presented in a certificate, which participant can use in support of their job applications. We administered four tests: (i) a Raven matrices test, (ii) a test of Amharic language skills, (iii) a test of mathematical ability and (iv) a 'work-sample' test (divided in three parts). The certificates report the relative grade of the test-taker for each test, and on an aggregate measure of performance.

The intervention took place over two days. On the first day, participants took the tests. On the second day, they attended a training event where qualified human resource professionals discussed how the information from these tests can be used to support participants' job applications. The tests were administered by the School of Commerce of Addis Ababa University, who also helped with the design of some of the components of the intervention.

4.3 Job fairs

Individuals in this treatment group were invited to attend two job fairs. In each job fair, jobseekers had the opportunity to interact with recruiters from several firms in Addis Ababa. We did not put any constraint on the nature of these short meetings. Typically, however, jobseekers used the meetings to learn about the firm and the available vacancies. If firms were interested in hiring the jobseeker, they invited her or him for a formal interview shortly after the job fair.

The firms that took part in this exercise were selected from a sample of 500 firms. This sample was randomly draw from a comprehensive lists of firms operating in the major sectors of Addis Ababa's economy.

The first fair took place on October 25 and 26, 2014. The second fair took place on 14 and 15 February 2015. Each job fair lasted two days in a single weekend. Individuals in the "fairs only" treatment were invited to attend the fair on one of these days, while individuals in the "fairs + screening" treatment were invited to attend on the other day.

5 Data

5.1 Construction of variables: Phone surveys

We conducted a novel phone survey where we called respondents every two weeks for an entire year (54 weeks in all). Respondents were called within three weeks of their first face-to-face interview. The phone surveys were conducted by trained phone enumerators from our dedicated phone survey centre in Addis Ababa. Respondents were asked a short series of questions covering the main outcomes of interest to the study. Since the phone calls were made every two weeks, we ask about activities that took place in the last two weeks, but then also about the previous seven days, as well as the preceding seven days (the seven days of the previous week), in order to construct a weekly panel for key outcomes of interest.

Table 7: **Phone Survey Constructed Variables**

VARIABLE	DEFINITION	SOURCE (QUESTION NUMBER)
	EMPLOYMENT OUTCOMES (PHONE SURVEYS)	
`work`	Dummy: Respondent has done any work for pay in the last 7 days	Dummy: (p1_3 > 0 or p1_6 > 0)
`days_worked`	Number of days worked in the last 7 days	p1_3 or p1_6
`hours_worked`	Number of hours worked in the last 7 days	p1_4 or p1_7
`earnings`	Earnings in the last 7 days	p1_5 or p1_8
`permanent_work`	Dummy: Respondent's main employment in the last 14 days has been permanent employment	Dummy: p1_9 = 1
	OUTCOMES ABOUT SEARCH (PHONE SURVEYS)	
`searching`	Dummy: Respondent has been searching for work in the last 7 days	p1_14
`days_searching`	Number of days searching for work in the last 7 days	p1_15 (= 0 if p1_14 = 0)

`searching_job_board`	Number of days visiting the job board in the last 7 days	p1_16
`job_applications`	Number of job applications made in the last month	p3_4
	SECONDARY OUTCOMES (PHONE SURVEYS)	
`temporary_work`	Dummy: Respondent's main employment in the last 14 days has been temporary employment	Dummy: p1_9 = 2 or p1_9 = 3 or p1_9 = 4
`self_employment`	Dummy: Respondent's main employment in the last 14 days has been self-employment	Dummy: (p1_9 = 5 or p1_9 = 6)
`work_satisfaction`	Dummy: Respondent is satisfied with work done	Dummy: (p2_3 = 1 or p2_3 = 2) (= 0 if no work)
`life_satisfaction`	Step on an imagined 10-step ladder ('best possible life') to ('worst possible life')	p3_3
`total_savings`	Total money saved (including formal or informal, but not including informal lending)	p2_4 + p2_5
`reservation_wage`	Minimum monthly earnings to accept full-time permanent wage work	p3_2
`travel_count`	Number of days in an average week in the past month that the respondent has travelled to the centre of Addis Ababa	p2_6
`moved_home`	Dummy: Respondent has moved home in the last month	Dummy: p2_1 = 1

5.2 Construction of variables: Face-to-face interviews

We define the following outcomes.

Table 8: **Face-to-face Survey Constructed Variables**

VARIABLE	DEFINITION	SOURCE (QUESTION NUMBER)

MAIN OUTCOMES ABOUT EMPLOYMENT

`work`	Dummy: Respondent has done any work for pay in the last 7 days	Dummy: j1_1= 1
`permanent_work`	Dummy: Respondent's main job in the last 7 days has been permanent work	Dummy: j1_6 = 1
`hours_worked`	Number of hours worked in the last 7 days	j1_4
`earnings`	Monthly earnings from main occupation	j1_10
`work_satisfaction`	Dummy: Respondent is satisfied with work done	Dummy: (j1_12 = 1 or j1_12 = 2) (= 0 if no work)
`written_agreement`	Does the respondent have a formal written agreement for the job?	Dummy = 1 if j1_17 = 1 or j1_17 = 2 (= 0 if no work)

OUTCOMES ABOUT EFFECTIVENESS OF JOB SEARCH

`apply_temp`	Number of applications for temporary jobs the respondent has made in the last 12 months	s3_6
`apply_perm`	Number of applications for permanent jobs the respondent has made in the last 12 months	s3_10
`interview_apply_all`	Ratio: number of interviews ÷ number of applications (all jobs combined)	(s3_11 + s3_7) ÷ (s3_10 + s3_6)
`offer_apply_all`	Ratio: number of job offers ÷ number of applications (all jobs combined)	(s3_12 + s3_8) ÷ (s3_10 + s3_6)
`interview_apply_perm`	Ratio: number of interviews ÷ number of applications for permanent jobs	s3_11 ÷ s3_10
`offer_apply_perm`	Ratio: number of job offers ÷ number of applications for permanent jobs	s3_12 ÷ s3_10
`interview_apply_temp`	Ratio: number of interviews ÷ number of applications for temporary jobs	s3_7 ÷ s3_6

`offer_apply_temp`	Ratio: number of job offers ÷ number of applications for temporary jobs	s3_8 ÷ s3_6
`cv_application`	Dummy: respondent has CV that he/she uses for job applications	Dummy: s3_2=1
`cert_application`	Dummy: respondent has certificates that he/she uses for job applications	Dummy: s3_4= 1

OUTCOMES ABOUT THE QUALITY OF THE JOB

`skills_match`	Does the respondent feel (s)he has the right skills for the position? (+)	Dummy = 1 if j1_22Enew = 3 (= 0 if no work)
`over_qualified`	Does the respondent feel underqualified for the position? (-)	Dummy = 1 if j1_22Enew = 1 (= 0 if no work)
`under_qualified`	Does the respondent feel overqualified for the position? (-)	Dummy = 1 if j1_22Enew = 2 (= 0 if no work)
`job_by_interview`	Did the respondent do a formal interview before getting the job? (+)	Dummy = 1 if j1_21 = 1
`office_work`	Does the respondent work in an office? (+)	Dummy = 1 if j1_11 = 1 or j1_11 = 2

FINANCIAL OUTCOMES

`expenditure`	Total expenditure in last 7 days (including rent) (+)	e1_11 + (e1_12 ÷ 4)
`saving`	Total savings available (including cash at hand and deposit in government housing scheme) (+)	e2_10 + e2_12 + e2_14 + e2_15
`assets`	Asset index (+)	weighted average of h1-h16 using variance covariance matrix

EXPECTATIONS, RESERVATION WAGES AND ASPIRATIONS

`expect_job`	If you do not have a permanent job, for how long do you expect to be without a permanent job? (-)	a1_3 new
`expect_offer`	How many job offers do you expect to receive in the next 4 months? (+)	a1_3
`res_wage`	Reservation wage (+)	a1_22
`aspiration`	What after tax monthly wage do you aspire (hope) to be earning in 5 years from now? (+)	a1_23

OUTCOMES ABOUT SPATIAL MOBILITY

`travel`	Number of trips to central Addis Ababa in last 7 days (+)	t2
`work_away`	Respondent works at a job that is not within walking distance of their home (+)	j1_15 !=0
`moved_occup`	Respondent moved location of main occupation since first interview (+)	b11a_new
`moved_in_addis`	Dummy: Respondent has moved within Addis (+)	Dummy: b12 $\in$ {1, 2, 3, 4}
`moved_out_of_addis`	Dummy: Respondent has moved out of Addis (+)	Dummy: b12 $\in$ {5, 6}

OUTCOMES ABOUT EDUCATION AND TRAINING

`fulltime_education`	Dummy: Respondent is in full-time education (including vocational training) (+)	Dummy: j0_3 = 22
`parttime_education`	Dummy: Respondent is in formal part-time education (including vocational training) (+)	j0_5
`informal_training`	Dummy: Respondent is in informal training (e.g. apprenticeship, on-the-job training) (+)	j0_6
`graduated`	Dummy: Respondent completed some education or training in the past 12 months (+)	Dummy: B4a_new = 1
`graduated_vocational`	Dummy: Respondent graduated from vocational education in the past 12 months (+)	Dummy: b5 $\in$ {9, 10, 11, 12, 13, 14, 15, 16}

`graduated_training`	Dummy: Respondent graduated from employment training in the past 12 months (+)	Dummy: b5 = 23

PSYCHOLOGICAL OUTCOMES

`life_satisfaction`	10 points ladder for life satisfaction (+)	a1_2
`locus_control`	10 points ladder for locus of control (+)	a1_2new
`oneness`	7 points scale for feeling of oneness with society (+)	a1_3new
`trust`	4 points scale for trust in others (+)	a1_5new

SOCIAL AND JOB NETWORKS

`network_size`	Number of ties in job contact network (+)	n11
`network_quality`	Number of ties in job contact network who are employed (+)	n12
`guarantor`	Would you be able to access a guarantor for a job if you needed one in the next month? (+)	n14
`associations`	How many meetings of voluntary associations have you attended in the last 30 days? (+)	n17

6 Identification strategy

6.1 Testing balance

We will begin our analysis of balance by reporting the balance statistics of Table 5. We will complement this by testing balance for all of the outcome variables listed in the previous two tables (*i.e.* outcome variables from the phone surveys and the interview data respectively).

Table 9 shows the different types of treatment; Table 10 shows the different types of control.

Table 9: **Assignment to Treatment**

TREATMENT	DESCRIPTION	DUMMY VARIABLE
group 1	transport intervention only	`treat1`
group 2	screening intervention only	`treat2`
group 3	job fair intervention only	`treat3`
group 4	screening intervention and job fair intervention	`treat4`

Table 10: **Assignment to Control**

TREATMENT	DESCRIPTION	DUMMY VARIABLE
group 5	'pure' control respondent	·
group 6	control respondent in cluster assigned to treatment 1	`spillover1`
group 7	control respondent in cluster assigned to treatment 2	`spillover2`

We will test balance by estimating the following regression, in which $y_{i,pre}$ denotes the first measure that exists for each outcome (*i.e.* depending on the variable, either the first phone interview with the respondent, or the baseline face-to-face interview):

$$
\begin{aligned}
y_{i,pre} = \beta_0 + \beta_1 \cdot \texttt{treat1} + \beta_2 \cdot \texttt{treat2} + \beta_3 \cdot \texttt{treat3} + \beta_4 \cdot \texttt{treat4} \\
+ \gamma_1 \cdot \texttt{spillover1} + \gamma_2 \cdot \texttt{spillover2} + \mu_{ic}. \qquad (1)
\end{aligned}
$$

```
ivreg2 y_pre treat1 treat2 treat3 treat4 spillover1 spillover2,
 cluster(ClusterID)                                              (2)
```

We will cluster errors by geographical cluster (in the sense discussed earlier, in Section 3.2.1); this is the variable `ClusterID` used in the example Stata code.

We will then run a Wald test of the joint hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \gamma_1 = \gamma_2 = 0$. We will report the p-value from this test in an extended table of descriptive statistics. Given the number of variables being tested, we will not be surprised if some variables reject this null hypothesis; we will include these variables as controls in robustness tests (discussed shortly).

6.2 Effects of the interventions

6.2.1 Treatment effects at endline

ANCOVA affords greater power than diff-in-diff when outcomes have low autocorrelation (McKenzie, 2012). As all participants do not have permanent employment at the beginning of the project, and search intensity is known to vary widely in the course of an unemployment spell, we have reason to expect relatively low autocorrelation in our key outcomes of interest. In the early weeks of our phone call survey, we find that the autocorrelation, week-on-week, of employment is only 0.48, the autocorrelation of search, week-on-week, is only 0.24. Using similar survey data from the same context in Franklin (2015), combined with phone call survey data, we anticipate that auto-correlation between baseline and endline employment (a period of over 50 weeks) to be lower than 0.25, and lower than 0.10 for the job search outcome. For this reason, we will use an ANCOVA specification.

Additionally, we assigned treatment using a re-randomisation method. Following the recommendations of Bruhn and McKenzie (2009), we will control in our estimations for the baseline covariates used for re-randomisation (that is, the set of variables described in Table 4) and for the baseline covariates used to construct the randomisation blocks.[4]

[4] Most of these are included in the list of Table 4. The two variables used for blocking that are not included in that list are: distance from the city centre and total number of individuals surveyed in the cluster.

We will estimate effects on a variety of outcome variables (discussed in detail shortly). For each outcome, we will estimate the following (where y_{ic} is the endline outcome for individual i in cluster c and $\boldsymbol{x}_{i0}$ is the vector of baseline covariate values that were used for re-randomisation and blocking):

$$y_{ic} = \beta_0 + \beta_1 \cdot \texttt{treat1}_i + \beta_2 \cdot \texttt{treat2}_i + \beta_3 \cdot \texttt{treat3}_i + \beta_4 \cdot \texttt{treat4}_i + \gamma_1 \cdot \texttt{spillover1}_i + \gamma_2 \cdot \texttt{spillover2}_i + \alpha \cdot y_{ic,pre} + \boldsymbol{\delta} \cdot \boldsymbol{x}_{i0} + \mu_{ic} \tag{3}$$

```
ivreg2 y treat1 treat2 treat3 treat4 spillover1 spillover2
y_pre balance_*, partial(balance_*) cluster(ClusterID)
```

As in the balance tests, we will cluster errors by geographical cluster, represented by `ClusterID`.

We will then run the following hypothesis tests:

(i). $H_0 : \beta_1 = 0$: The transport intervention had no effect;

(ii). $H_0 : \beta_2 = 0$: The screening intervention had no effect;

(iii). $H_0 : \beta_3 = 0$: The job fair intervention had no effect;

(iv). $H_0 : \beta_4 = 0$: The combination of screening intervention and job fair intervention had no effect;

(v). $H_0 : \gamma_1 = 0$: The transport intervention had no spill-over effect;

(vi). $H_0 : \gamma_2 = 0$: The screening intervention had no spill-over effect;

(vii). $H_0 : \beta_2 + \beta_3 = \beta_4$: The effects of the screening intervention and the job fair intervention were additively separable.

The first six hypothesis tests will naturally be reported as part of the regression output; the remaining hypothesis test will be reported as separate post-estimation test.

For outcomes where variable definitions differ slightly at baseline and endline, we will use as $y_{ic,pre}$ the value of the baseline covariate that is closest to y_{ic}. The following table lists variables that differ between baseline and endline; note that all of these variables are defined earlier in this pre-analysis plan. The table specifies which variable will be used in each case as $y_{ic,pre}$. (If 'baseline code' is blank in the following table, we will omit $y_{ic,pre}$ entirely.)

Table 11: **Variables that differ between baseline and endline**

VARIABLE	DIFFERENCE	ENDLINE CODE	BASELINE CODE
`fulltime_education`	Endline dummy: j0_5 = 3. Baseline dummy: j0_3 = 22	j0_5	j0_3
`skills_match`, `under_qualified`, `over_qualified`	We did not ask this question at baseline.	j1_22Enew	
`apply_temp`	Endline: we ask for the number of applications; baseline: we ask whether any application was made.	s3_6	s3_6
`apply_perm`	Endline: we ask for the number of applications; baseline: we ask whether any application was made.	s3_10	s3_10
`interview_apply_perm`, `offer_interview_perm`, `interview_apply_temp`, `offer_interview_temp`	We cannot calculate these ratios as baseline, as we do not have numbers of interviews and applications.		
`locus_control`, `oneness`, `trust`	We did not ask these questions at baseline.	a1_2new, a1_3new, a1_5new	
`network_quality`, `guarantor`, `associations`	We did not ask these questions at baseline.	n12, n14, n17	
`moved_occup`	We did not ask these questions at baseline.	b11a_new	

moved_in_addis	We did not ask this question at baseline.	Dummy: b12 $\in \{1, 2, 3, 4\}$	
moved_out_of_addis	We did not ask this question at baseline.	Dummy: b12 $\in \{5, 6\}$	
graduated	We did not ask this question at baseline.	Dummy: B4a_new = 1	
graduated_vocational	We did not ask this question at baseline.	Dummy: b5 $\in \{9, 10, 11, 12, 13, 14, 15, 16\}$	
graduated_training	We did not ask this question at baseline.	Dummy: b5 =23	

6.2.2 Bounding effects for outcomes only observed among the employed

Several of our outcome variables are only observed among those respondents who are employed; for example, a respondent cannot have 'monthly earnings from main occupation' if he or she does not have a main occupation. We will deal with this in two ways:

(i). In our primary estimations, we will treat such observations as being zero by definition; thus, for example, we will code zero earnings for those who do not have a main occupation. Using this approach, our main specifications will provide consistent estimates of the Average Treatment Effect on the *unconditional* value of our outcome variables.

(ii). Further, we will be interested to know the effect of our treatment on outcome variables *conditional* on a respondent being employed. To calculate this, we will report Lee (2009) bounds, justified by the assumption that our treatments have a monotonic effect on employment.

We plan to report such bounds for the following outcomes:

(i). `earnings`,

(ii). `hours_worked`,

(iii). `work_satisfaction`,

(iv). `written_agreement`,

(v). `interview_apply_all`, `offer_apply_all`, `interview_apply_perm`, `offer_apply_perm`, `interview_apply_temp`, `offer_apply_temp`,

(vi). `skills_match`,

(vii). `over_qualified`,

(viii). `under_qualified`,

(ix). `job_by_interview`.

(x). `office_work`.

6.2.3 Differences between saturation levels

To measure the spill-over effects of the treatments, we ensured that some respondents living in treated clusters were not offered the treatments. We also randomly varied the proportion of individuals treated in the clusters that received the *transport* intervention. To test for spill-overs of the transport intervention, we will run a regression of the form:

$$
\begin{aligned}
y_{ic} = \kappa &+ \beta_{20} \cdot S_{20c} \cdot \mathtt{C}_i + \beta_{40} \cdot S_{40c} \cdot \mathtt{C}_i + \beta_{75} \cdot S_{75c} \cdot \mathtt{C}_i + \beta_{90} \cdot S_{90c} \cdot \mathtt{C}_i \\
&+ \gamma_{20} \cdot S_{20c} \cdot \mathtt{T}_i + \gamma_{40} \cdot S_{40c} \cdot \mathtt{T}_i + \gamma_{75} \cdot S_{75c} \cdot \mathtt{T}_i + \gamma_{90} \cdot S_{90c} \cdot \mathtt{T}_i \\
&+ \alpha \cdot y_{ic,pre} + \boldsymbol{\delta} \cdot \boldsymbol{x}_{i0} + \mu_{ic}. \qquad (4)
\end{aligned}
$$

```
ivreg2 y s20Ic s40Ic s75Ic s90Ic s20It s40It s75It s90It
 y_pre balance_*, partial(balance_*) cluster(ClusterID)
```

where the sample is restricted to individuals in groups 1, 5 and 6 from table 9. The reference group is group 5, the "pure" control group. T_i identifies individuals who have been assigned to the transport treatment, while C_i identifies individuals who have not been assigned to the transport treatment. S_{20c} is a dummy variable for individuals living in a cluster where 20% of individuals were offered the transport treatment. Thus, β_{20} captures the difference in outcomes between untreated individuals in these clusters and untreated individuals in clusters where nobody was treated. Further, γ_{20} measures the difference in outcomes between treated individuals in S_{20c} clusters and untreated individuals in untreated clusters. S_{40c}, S_{75c}, S_{90c}, and the remaining β and γ coefficients have a similar interpretation.

If the treatment has no external effect on the untreated, untreated individuals in treated clusters have similar outcomes to untreated individuals in treated clusters. We will thus test the hypothesis that the treatment has external effects on untreated individuals who live in the same cluster as treated individuals with an F-test of the null hypothesis that $\beta_{20} = \beta_{40} = \beta_{75} = \beta_{90} = 0$. We will also test $\beta_{20} = \beta_{40} = \beta_{75} = \beta_{90}$, to check whether untreated individuals have different outcomes depending on the treatment saturation of their cluster.

If the treatment has no external effect on the treated, treated individuals have the same outcomes irrespective of the proportion of treated individuals in their areas.We will test the hypothesis that the treatment has external effects on treated individuals who live in the same cluster as other treated individuals with an F-test of the null hypothesis that $\gamma_{20} = \gamma_{40} = \gamma_{75} = \gamma_{90}$.

6.2.4 Differences in timing of the transport treatment

Different individuals have received the transport treatment for different amounts of time. This varied from a minimum of 13 weeks to a maximum of 20 weeks. To test for the effect of duration of treatment we will run a regression of the following form, using only observations from individuals offered the transport intervention ('Group 1' in Table 9):

$$y_{ib} = \beta_0 + \beta_1 \cdot \text{weeks1315} + \beta_2 \cdot \text{weeks1720} + \alpha \cdot y_{ic,pre} + \boldsymbol{\delta} \cdot \boldsymbol{x}_{i0} + \mu_{ic}. \tag{5}$$

```
ivreg2 y weeks1315 weeks1720
  y_pre balance_*, partial(balance_*) cluster(ClusterID)
```

where "weeks1315", and "weeks1720" are dummy variables that identify individuals who have received the transport treatment for 13 to 15 weeks, and 17 to 20 weeks, respectively. The residual category is individuals who have received the treatment for 16 weeks.

6.2.5 Differences in trajectories: Exploiting the phone data

We will pool the data from the phone calls across all weeks to estimate the trajectory of the treatment effects across the weeks of the study.

By including week specific dummy variables, and interacting treatment dummies with dummy variables indicating for how long the respondent had been receiving treatment, we will estimate the impact of each treatment g and the spillover effect on each spillover group s, for each week before (or after) that treatment began.

Define w as a variable indicating the number of weeks since each treated individual began receiving his/her treatment. $w = 0$ in the week that the treatment started, and is negative for weeks before that. Define the dummy d_{wit} as a dummy variable equal to 1 in period t if an individual started receiving their treatment w periods ago. For an individual assigned to receive the transport treatment from week 15 of the study onwards, the dummy d_{0it} will take the value 1 in week 15 (and is equal to 0 for all other weeks). Similarly we will assign $d_{-1i14} = 1$, and $d_{5i20} = 1$, and so on. Individuals in the control group have all such dummy variables set to 0. Let η_t represent the usual set of time dummies (unrelated to the timing of treatment). We then run:

$$\begin{aligned} y_{itc} = \eta_t &+ \sum_{g=1}^{4} \sum_{w=S_g}^{E_g} \beta_{gw} \cdot \texttt{treat}_{gi} \cdot d_{wit} \\ &+ \sum_{s=1}^{2} \sum_{w=S_s}^{E_s} \gamma_{sw} \cdot \texttt{spillover}_{si} \cdot d_{wit} + \alpha_t \cdot y_{itc,pre} + \boldsymbol{\delta} \cdot \boldsymbol{x}_{i0} + \mu_{itc}. \end{aligned} \tag{6}$$

Here, η_t is a time-specific intercept term. We allow the effect of the baseline control term $y_{ic,pre}$ to vary over time by estimating α_t for each time period, while we estimate time-invariant effects of individual covariates x_{i0}.

Note that because interventions ran for different lengths of time, the number of weeks for which we will be able to estimate the treatment effect relative to the start week of the treatment will differ by treatment. In the notation above S_g denotes the earliest week for which we will be able to estimate a treatment effect for treatment or spillover group g. E_g denotes the final week. If, for example, a treatment began in week 15 of the study, then $S_g = -15$ and $E_g = 39$. For this treatment, we will use data from week 10 of the study to estimate the coefficient β_{g-5}.

Thus, for each time period before and after treatment g began, we will get an estimate β_{gw} of the impact of the treatment in that time period. Similarly γ_{sw} estimates the spillover effect in spillover group s, in week w after treatment began.

We will be able to plot graphically the trajectory of these estimators over time, in a graphs similar to those in Franklin (2015), Figure 3. We will be able to estimate 54 different coefficients for each intervention for the weekly data, 27 coefficients for data that we have at fortnightly interviews, and 13 coefficients for outcomes that we have data on only every 4 weeks.

For some of the weekly outcomes, we will follow McKenzie (2012) and pool weekly time observations into groups of 4 weeks (months) and estimate the effect of the treatment on average outcomes for that month for each intervention. Averaging observations in this way should improve the precision of our estimates by reducing the variance of high-frequency outcomes, which may be very volatile. We will run the same specification, but using the monthly averages of the weekly data, to run the regression for each month m. Here w_{iw}, S_g and E_g are defined as above, but now in terms of *months* before or after a treatment began.

$$
\begin{aligned}
y_{imc} = \eta_m &+ \sum_{g=1}^{4} \sum_{w=-S_g}^{E_g} \beta_{gw} \cdot \texttt{treat}_{gi} \cdot d_{wim} \\
&+ \sum_{s=1}^{2} \sum_{w=-S_s}^{E_s} \gamma_{sw} \cdot \texttt{spillover}_{si} \cdot d_{wim} + \alpha_m \cdot y_{ic,pre} + \boldsymbol{\delta} \cdot \boldsymbol{x}_{i0} + \mu_{imc}.
\end{aligned} \tag{7}
$$

For this estimation, we define

$$y_{imc} \equiv 0.25 \times \sum_{t=(4*(m-1))+1}^{t=(4*m)} y_{it}. \tag{8}$$

Further, we can estimate the trajectory of treatment effects by pooling all post treatment ($w \geq 0$) time observations together and estimating quadratic trends over time of the treatment effect for each main intervention. To do this, we take the specification in equation 6, and impose that no treatment has an effect before it commences (an assumption that is testable through equation 6). The easiest way to express this is by estimating equation 6, but subject to quadratic constraints on β_{gw} and γ_{sw}:

$$\begin{aligned} y_{itc} = \eta_t &+ \sum_{g=1}^{4} \sum_{w=S_g}^{E_g} \beta_{gw} \cdot \texttt{treat}_{gi} \cdot d_{wit} \\ &+ \sum_{s=1}^{2} \sum_{w=S_s}^{E_s} \gamma_{sw} \cdot \texttt{spillover}_{si} \cdot d_{wit} + \alpha_t \cdot y_{itc,pre} + \boldsymbol{\delta} \cdot \boldsymbol{x}_{i0} + \mu_{itc} \end{aligned} \tag{6}$$

subject to:

$$\beta_{gw} = \begin{cases} 0 & \text{if } w < 0; \\ \phi_{g0} + \phi_{g1} \cdot w + \phi_{g2} \cdot w^2 & \text{if } w \geq 0; \end{cases} \tag{9}$$

$$\text{and } \gamma_{sw} = \begin{cases} 0 & \text{if } w < 0; \\ \theta_{s0} + \theta_{s1} \cdot w + \theta_{s2} \cdot w^2 & \text{if } w \geq 0. \end{cases} \tag{10}$$

That is, instead of estimating parameters β_{gw} and γ_{sw}, we will estimate ϕ_{g0}, ϕ_{g1}, ϕ_{g2}, θ_{s0}, θ_{s1} and θ_{s2}.

7 Structure of analysis and correcting for multiple testing

The previous section explains the different estimations that we plan to run. In this section, we outline the structure of those estimations; in particular, we highlight which outcome variables we consider to be primary to our analysis, and we discuss our intended correction for multiple hypothesis testing.

7.1 Primary outcome at endline: Employment

Following Olken (2015), we begin by defining our primary outcomes of interest. Our key hypothesis is that our treatments affect respondents' employment outcomes; this forms our primary set of dependent variables. Our key hypothesis is that the treatments do this via increased and more effective job search; this forms our primary set of mechanisms.

To test the effect of our treatments on employment, we will construct a family of outcome variables, comprising those outcomes earlier defined in the group 'MAIN OUTCOMES ABOUT EMPLOYMENT' (face-to-face interviews). For each of these outcomes, we will run the estimation and hypothesis tests outlined earlier in section 6.2.1. For each hypothesis test, we will report two values:

(i). The usual p-value from a Wald test; and

(ii). We will report False Discovery Rate q-values, taken across the family of outcomes (Benjamini, Krieger, and Yekutieli, 2006). (That is, for each type of test, we will construct a q-value for that test across outcomes. For example, we will construct a set of q-values using all p-values for the null hypothesis 'The transport intervention had no effect'; we will then construct a set of q-values using all p-values for the null hypothesis 'The screening intervention had no effect', and so on.)

7.2 Primary outcome trajectory: Employment

We will then test the effect of our treatments over time, by exploiting our mobile phone data. Specifically, we will use the identification strategy outlined in section 6.2.5, and we will present our results graphically (following Franklin (2015)). To do this, we will construct a family of outcome variables using the group

'**EMPLOYMENT OUTCOMES (PHONE SURVEYS)**'.

For each outcome variable, we will report the following tests:

(i). Estimate equations 6, 7 and equation 6 subject to constraints (9) and (10)

(ii). For each estimated equation and each treatment g, report an F-test of the hypothesis '$\beta_1 = 0$ in all post-treatment periods', a separate p-value from the F-test of the hypothesis '$\beta_2 = 0$ for all post-treatment time periods', and so on

(iii). For each of these tests, we will construct a q-value, using the False Discovery Rate, applied to the family of outcome variables described earlier.

7.3 Primary mechanism: Job search trajectories

To test the effect on job search, we will create a family of outcome variables, by combining the variables defined earlier as '**OUTCOMES ABOUT SEARCH**' (fortnightly phone surveys). For this family of four outcomes, we will repeat the estimation exercise described in the previous section (including the method of constructing p-values and q-values).

7.4 Primary mechanism: Job search recall at endline

The previous analysis will reveal the effect of our treatments on trajectories of search and of employment. To link these trajectories, we will use endline measures recording the total stock of search activity, to test whether our treatments increased the efficacy of search.

To do this, we will form a family of outcomes, using the outcome variables earlier described as '**OUTCOMES ABOUT EFFECTIVENESS OF JOB SEARCH**' (face-to-face interviews). To analyse these outcomes, we will use the same approach as outlined earlier in section 7.1. Namely, for each of these outcomes, we will run the estimation and hypothesis tests outlined earlier in section 6.2.1. For each hypothesis test, we will report two values:

(i). The usual p-value from a Wald test; and

(ii). We will report False Discovery Rate q-values, taken across the family of outcomes (Benjamini, Krieger, and Yekutieli, 2006). (That is, for each type of test, we will construct a q-value for that test across outcomes. For example, we will construct a set of q-values using all p-values for the null hypothesis 'The transport intervention had no effect'; we will then construct a set of q-values using all p-values for the null hypothesis 'The screening intervention had no effect', and so on.)

7.5 Secondary outcomes: Endline

We have a wide range of secondary outcomes. To deal with these outcomes, we will use a standard 'omnibus' approach: namely, we will use families of outcomes, construct an index for each family, and test whether each index is affected by our treatments.

Specifically, we will construct the following combined hypotheses:

(i). **H1: Our treatment improved other aspects of job quality (separate from those tested in the primary outcomes).** We test this using the family 'OUTCOMES ABOUT THE QUALITY OF THE JOB'.

(ii). **H2: Our treatment improved the financial position of programme recipients.** We test this using the family 'FINANCIAL OUTCOMES'.

(iii). **H3: Our treatment improved recipients' expectations and aspirations for future employment.** We test this using the family 'EXPECTATIONS, RESERVATION WAGES AND ASPIRATIONS'.

(iv). **H4: Our treatment improved recipients' spatial mobility.** We test this using the family 'OUTCOMES ABOUT SPATIAL MOBILITY'.

(v). **H5: Treated recipients obtained more education and training.** We test this using the family 'OUTCOMES ABOUT EDUCATION AND TRAINING'.

(vi). **H6: Our treatment improved recipients' psychological well-being.** We test this using the family 'PSYCHOLOGICAL OUTCOMES'.

(vii). **H7: Our treatment improved recipients' social and job networks.** We test this using the family 'SOCIAL AND JOB NETWORKS'.

For each family, we will use the following approach:

(i). We will construct a weighted index of outcomes, using the method of Anderson (2008). We will construct this index both for baseline and for endline. To construct the index, we will use the sign indicated against the relevant variable descriptions in the earlier variable definitions (*i.e.* '+' or '-').

(ii). We will test the effect of our treatments on the index, using the ANCOVA specification in equation 3.

(iii). We will report a p-value for each family separately. We will also construct FDR q-values by treating each index as a separate member of a 'super-family' of indices.

For completeness, we will separately report effects on individual variables. We will report these outcomes within the families defined earlier; for each individual variable, we will report a p-value and will report FDR q-values calculated within the family. We will pay particular attention to those families whose index is significant.

7.6 Secondary outcomes: Trajectories

We will use the identification strategy outlined in section 6.2.5 to study the trajectories of a set of secondary outcomes which we collected in the phone interviews. We will present our results graphically (following Franklin (2015)). To do this, we will construct a family of outcome variables using the group 'SECONDARY OUTCOMES (PHONE SURVEYS)'.

For each outcome in this family, we will proceed as outlined in section 7.2.

7.7 Heterogeneous effects

We plan to study treatment effects for a number of relevant sub-groups. Sub-groups are identified by categorical variables capturing characteristics at baseline. When characteristics are continuous, we create subgroups by separating individuals below and above the median level of the characteristic.

For each intervention, we will run the following specification:

$$y_{ic} = \sum_{v=0}^{m} \left[\beta_v + \sum_{f=1}^{4} \gamma_{vf} \cdot \texttt{treat}_{fi} \cdot I\left(x_{i,pre} = v\right) + \gamma_{v5} \cdot \texttt{spillover}_{1i} \cdot I\left(x_{i,pre} = v\right) \right.$$
$$\left. + \gamma_{v6} \cdot \texttt{spillover}_{2i} \cdot I\left(x_{i,pre} = v\right) \right] + \alpha \cdot y_{ic,pre} + \boldsymbol{\delta} \cdot \boldsymbol{x}_{i0} + \mu_{ic}, \tag{11}$$

where $\texttt{treat}_{ai}$ is a dummy for whether individual i received treatment a (see Table 9), and $\texttt{spillover}_{1i}$ and $\texttt{spillover}_{2i}$ respectively refer to group 6 and group 7 (see Table 10). $x_{i,pre}$ is a categorical variable with values $\{0, \ldots, m\}$, and $I\left(x_{i,pre} = v\right)$ is a indicator variable that takes the value of 1 when $x_{i,pre}$ is equal to v.

The coefficients γ_{vf} estimate the effect of treatment f for subgroup v. For each estimation, we will report six separate post-estimation tests (one for each value of $f \in \{1, \ldots, 6\}$), for the null hypothesis that the treatment effect does not vary across subgroups:

$$H_0 : \gamma_{0f} = \gamma_{1f} = \ldots = \gamma_{mf}.$$

We plan to study heterogeneity in impacts for the subgroups defined in Table 12. We will perform the subgroup analysis for the outcomes in the family '**MAIN OUTCOMES ABOUT EMPLOYMENT**'.

8 Attrition

By 'attrition', we mean both (i) being unable to interview respondents when scheduled for their regular phone interview, and (ii) being unable to interview respondents for the endline interview. We will deal with both kinds of attrition in the same basic way.

First, we will create a dummy variable for whether the individual's scheduled interview is missing. We will regress this dummy on the treatment dummies in an ANCOVA specification, clustering by `ClusterID`. That is, we will estimate equation 3 or 6, depending on whether the outcome is a dummy for the respondent being missing in the telephone interviews or the endline interview. We will interpret this as a test for whether we have differential attrition by treatment status. (In the case of the phone interviews, we will report coefficients separately by interview wave, by graphing.)

Second, if (and only if) we find significant differential attrition by treatment status, we will report Lee (2009) bounds for the outcome variables potentially affected by that differential attrition. (For example, if we find differential attrition in the endline face-to-face interview, we will report Lee (2009) bounds for outcomes in the endline face-to-face interview.)

Table 12: **Subgroups for Heterogenous Treatment Effects**

VARIABLE	DESCRIPTION	DEFINITION
`education`	Level of education attained	`education` = 0 if b5 $\in \{6,\ldots,8\}$; `education` = 1 if b5 $\in \{9,\ldots,16\}$; `education` = 2 if b5 $\in \{17,\ldots,19\}$; `education` = 3 if b5 $\in \{20,21\}$
`saving_dummy`	Total amount of savings is above median	`saving` = e2_10 + e2_12+ e2_14 + e2_15; `saving_dummy`: dummy if `saving` is above median
`experience_perm`	Individual has work experience in a permanent position	Dummy: b22= 3 or b22=4
`search_freq`	Proportion of weeks that individual searched for work (from the phone surveys)	Mean (over first 3 months of calls) of Dummy: $p1\_14 = 1$
`female`	Gender	`female` = 1 if `respondent_gender`= 2; `female`=0 if `respondent_gender`= 1
`migrant_birth`	Respondent was born outside of Addis Ababa and migrated since birth	`migrant_birth` = 1 if b14!=10
`distance_centre`	Distance between the centre of Addis Ababa and the place of residence	GPS coordinates
`certificate`	Uses certificates or CV in job applications	`certificate` = 1 if s3_2 = 1 or s3_4 = 1
`networksize`	Number of individuals with whom respondent exchanges information about jobs	Sum of n6 and n7 across all rows
`present_bias`	Individual revised allocation decision towards the immediate future in the behavioural game (a1_25, a1_26 and phone data)	Incentivised behavioural game
`bias_unsoph`	`present_bias` = 1 and individual does not anticipate revising allocation towards the immediate future	Incentivised behavioural game and a1_28 and a1_29

References

ANDERSON, M. L. (2008): "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American statistical Association*, 103(484).

BENJAMINI, Y., A. M. KRIEGER, AND D. YEKUTIELI (2006): "Adaptive Linear Step-up Procedures that Control the False Discovery Rate," *Biometrika*, 93(3), 491–507.

BRUHN, M., AND D. MCKENZIE (2009): "In Pursuit of Balance: Randomization in Practice in Development Field Experiments," *American Economic Journal: Applied Economics*, 1(4), 200–232.

FRANKLIN, S. (2015): "Location, Search Costs and Youth Unemployment: A Randomized Trial of Transport Subsidies in Ethiopia," *CSAE Working Paper WPS/2015-11*.

IMBENS, G. (2011): "Experimental Design for Unit and Cluster Randomized Trials," http://cyrussamii.com/wp-content/uploads/2011/06/Imbens_June_8_paper.pdf, accessed 29 July 2015.

LEE, D. S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *The Review of Economic Studies*, 76(3), 1071–1102.

MCKENZIE, D. (2012): "Beyond Baseline and Follow-up: The Case for More T in Experiments," *Journal of Development Economics*, 99(2), 210–221.

OLKEN, B. A. (2015): "Promises and Perils of Pre-Analysis Plans," *The Journal of Economic Perspectives*, 29(3), 61–80.