Public Works and Cash Transfers in Urban Ethiopia

World Bank

2019

Public works and cash transfers in Urban Ethiopia.

Pre-Analysis Plan[1]

Date: March 7, 2019

Public works and cash transfers in urban Ethiopia: Evaluating the Urban Productive Safety Net Program

Girum Abebe[*], Simon Franklin[†] and Carolina Mejia-Mantilla[‡]

1 MOTIVATION

This document outlines the pre-analysis plan to study the effects of the Urban Productive Safety Net Program (UPSNP) in Addis Ababa, Ethiopia, one year after it started. The evaluation takes advantage of the randomization of the program at the woreda level during its first year of implementation. Of the 90 eligible woredas in Addis, 35 were randomly assigned to start implementation right away (in year 1) while the remaining 55 were assigned to start implementation approximately 12 months later (in year 2). We look at household and individual effects of the program on direct recipients of both of the two main intervention arms of the program, as well as the effect of the program on non-beneficiary households living in urban neighbourhoods where the program is rolled out.

The UPSNP is a comprehensive and large-scale social protection program that is designed to enhance the income of households living in the urban areas of Ethiopia, with the aim to reduce poverty and vulnerability. The program is envisioned as a three-phase graduation process. We study the first phase, in which beneficiaries are assigned either to engage in a public works (PW) program or, if they are unable to work, they receive an unconditional direct support transfer - DS. While the Government of Ethiopia plans to implement the UPSNP in all urban areas of the country, it is rolling it out in a gradual manner.

In its first stage, from 2016 until 2021, the program is being implemented in the largest 11 cities of the country (Addis Ababa, Harar, Gambela, Semera, Asosa, Mekelle, Dessie, Hawassa, Jigjiga, Dire Dawa and Adama), targeted at the poorest 12 percent[1] and with an emphasis on the public works component. Given that urban poverty is concentrated in Addis Ababa, around 70 percent of the beneficiaries of the program reside in this city. Thus, through the

[1]In Addis Ababa the program is targeted at the 18 percent poorest households in the city.

randomization in Addis we are able to analyze the short term direct effects of the first phase of the program, comprising two of its key components -public works and direct unconditional transfers- for the majority of beneficiaries.

The main questions this study aims to address are:

1. What is the effect of workfare on total income taking into account changes in total labour supply?

2. Does non-public-works labour force participation change in response to the program? Are changes in labour supply driven by the time requirement of the public works relative to the income effect of the program?

3. What are the effects of the program on household welfare, in terms of consumption and subjective well-being (comprising life-satisfaction and mental health)?

4. What are the effects of the workfare program on gender roles, norms and beliefs, given that the large majority of the work is being undertaken by women?

5. What are the effects of the public works program on local public goods and communities?

6. How does the presence of this large welfare program in a local area affect how people think about the state, redistribution and the status of welfare recipients, both among beneficaries and their non-beneficiary neighbours?

This pre-analysis plan is not intended to be exhaustive nor to prevent additional analysis that stems from results obtained, and we anticipate conducting supplementary analysis.

2 PROGRAM

The main objective of the UPSNP program is to enhance the income households living in the urban areas of Ethiopia, with the aim to reduce poverty and vulnerability.

2.1 TARGETING

Targeting was conducted at the household level by local community groups, the Ketena Targeting Committees (KTC), who were trained before the selection process started in treated woredas. The KTC listed all households in the community that were considered to be poor (it was reported that in all cases the listing was larger than the allotted number of beneficiaries per ketena), and then ranked households according to their level of income (if available), asset ownership (quality/quantity), health/disability status and other factors. After the ranking was completed, the KTC determined whether each household would be classified as a DS or PW beneficiary.

The KTC submitted the final list of targeted households to the city administration, which ratified the KTC nomination depending on the total number of beneficiaries per sub-city and making sure that the DS beneficiaries did not exceeded 16 percent of total beneficiaries in all cities. Households were then randomly selected and surveyed for a Proxy Means Testing. If the PMT at the woreda level indicated that there was a targeting inclusion error of more than 20%, then the targeting for that woreda was revised. After passing the PMT test, the final lists were posted in public places of the community for review, and any complaints were to be addressed to the Appeals Committee, who made a final decision.

In all, we estimate that roughly 18% of all households in treated districts (woredas) in Addis are enrolled into the program. We anticipate some variation in the percentage of the sample treated in each woreda. Of those targeted by the program, 30% are expected to receive direct support (DS), while 70% will be offered public works.

Due to the significant differences in the design and selection of beneficiaries of these two treatment arms and, we will separately identify their effects, by comparing households targeted for each arm across treatment and control areas.

2.2 PUBLIC WORKS COMPONENT

Beneficiary households enrolled in the public works component received a monthly transfer conditional on their participation in public projects close to their place of residence (woreda/ketena), which included urban greenery development, watershed management activities and environmental cleaning activities. During the first year, participation in public works entitled beneficiaries to a transfer of ETB 60 per person per day (around U\$ 2), with a cap of 5 days per month per household member, for up to four persons per household. This resulted in a maximum of 240 days per year for each beneficiary household. While the maximum of 20 days of public works per month could be distributed among up to four household members, it was up to the household to choose who did the work. Qualitative work conducted when the program had already begun suggests that the vast majority of participants are female.

In sum, households (with four members enrolled) can receive up to ETB 1,200 per month, or ETB 14,400 per year, deposited into a designated bank account. The transfer still represents a significant amount for a poor household in Addis Ababa. For the bottom quintile (target population) average annual household consumption in 2016 was around ETB 27,000.

2.3 DIRECT SUPPORT COMPONENT

Beneficiary households with no members able to participate in the public works, (due to chronic illness, age or disabilities) received ETB 170 per person a month as an unconditional cash transfer.

3   Randomization and Data Collection

The UPSNP program was randomized at the woreda (urban district) level in Addis Ababa. This study focusses on Addis Ababa alone, since that is where randomization was implemented. In year 1 of the program, only households residing in woredas with poverty rates above 20% were eligible for the program: specifically, 90 out of 116 woreda in the city. Randomization was conducted by a public lottery draw of woreda names on November 2016, and it was stratified by 10 sub-cities (urban sectors) within the city. Of these 90 eligible woredas, 35 were selected for the program in year 1 and the remaining 55 woredas were scheduled to receive the program in year 2.

The sample of households was selected both before the start of the program and before the randomization of woredas into the program (See Timeline below). The baseline survey was conducted in January 2017. Households were selected for the baseline survey using a PMT (proxy means test) calculation for poverty, with the understanding that selection of beneficiaries for the program would be based on a similar set of a criteria. More specifically, a random sample of all households was conducted, using a random walk method starting in randomly selected points within each eligible woredas. The short proxy-means survey, used to derive a predicted poverty score, was implemented to the sample. We then conducted the longform baseline survey with those households with poverty scores below the 18th percentile, chosen to reflect the fact that the program was to be targeted to the poorest 18% of households in each woreda. Because only 35 woredas were selected for year 1, and we surveyed before the targeting, our baseline sample includes more households in treated areas than in control areas.

Timeline:

Months	Year	Event
Oct-Nov	2016	Screening survey
Nov	2016	Woreda randomization
Nov-Jan	2016- 2017	Baseline survey collection
February	2017	Randomization announcement, beneficiary targeting and selection in year 1 districts
May	2017	Start of program in year 1 districts
July	2017	Start of payments
Dec	2018	Increase in the required work hours for public works beneficiaries.
March	2018	Endline survey.
July	2018	Beneficiary selection in year 2 of the program (control woredas)
August	2018	Start of the program in year 2 woredas.
August	2018	Survey of year 2 woredas to determine treatment status.

In all, we listed 30,951 households with a short listing questionnaire at baseline. And from that group we interviewed 6,026 households with a long-form questionnaire at baseline. We

had budget to interview roughly 6,000 households at endline. To increase power to detect the effects of the program (based on information about the randomization of woredas and the types of hosueholds selected for the program) we changed the composition of the end-line sample from the 6,026 households surveyed at baseline. We did this in three ways, in each case making sure to use identical selection rules across treatment and control areas in order to maintain fidelity of our randomization design. First, we randomly dropped (in equal proportion across treatment and control) 739 households with high measured consumption per adult equivalent as measured in the long-form baseline survey. These were households deemed unlikely to be included in the program. Second, after the selection of treatment areas was announced, we randomly 922 dropped households from control areas only. We did this completely at random, without regard to household characteristics, to keep treatment and control woredas comparable. Finally we added 1,728 households from the list of 30,951 listed households with whom we did not originally conduct a longform baseline survey, randomly selected from small households with low predicted consumption scores. We did this to over sample smaller households: we learnt from government administrative data that small households were over-represented in the group of beneficiaries relative to our baseline sample.

The endline survey was conducted with a total of 6,093 households, from March to July 2018. Table 3 shows the composition of the baseline and endline samples, by treatment and control areas.

		Control	Treatment	Total
	Dropped from Baseline			
1	Baseline only (high consumption)	460	279	739
2	Baseline only (control areas)	922	0	922
	Final Sample			
3	Screening and Endline	836	892	1,728
4	Longform baseline and Endline	2,226	2,139	4,365
	Total Endline (3+4)	3,062	3,031	6,093
	Original longform baseline (1+2+4)	3,608	2,418	6,026

In addition to the baseline and endline surveys, we conducted a short phone survey with all households in the sample, on a rolling basis throughout the year. We randomized the order in which households were called. In December of 2018 we learnt of a change in the implementation of the public works component, whereby the number of hours work required by beneficiaries increased considerably (from 2 hours to between 6 and 8 hours per day). Therefore, we divide our phone survey data into two halves, one from before and one after the chance. Due to the rolling nature of the phone survey, we do not observe all households on both sides of the change. But because of the randomized order of the survey (stratified by

treatment and control) we are able to separately identify the effects of the program on either side of the change in the implementation.

Finally, after the endline survey in mid-2018, program roll out begun in the "year 2" woredas. Targeting begun immediately after our main endline survey, after which a short survey was conducted with all members of the year 2 woredas to determine their treatment status. This allows us to compare treated households in year 1 treated woredas to treated households in year 2 treated woredas (not treated at the time of the endline survey) to identify the effects of the program. We can do this for the public works and direct support components separately, by comparing year 1 PW and DS, to year 2 PW and DS beneficiaries, respectively.

In all, we have five data collection periods:

1. Listing and Baseline (Nov 2016 - Jan 2017)

2. Phone survey before (Jan 2018)

3. Phone survey after (Jan 2018)

4. Endline (March 2018)

5. Phone survey - treatment status year 2 (August 2018)

The main respondent of our household surveys was usually the senior female household member. In cases where a male and female couple resided together, we always interviewed the woman. In these cases, we also conducted a short follow-up survey with her male spouse, on his labour supply outcomes, control over household resources and attitudes to gender issues. In cases where a male household member resides alone without a female spouse, we interviewed him.

4 EMPIRICAL STRATEGY

We look at the impact of the program at endline (March 2018). We will use the phone survey data to explore mechanisms.

4.1 HOUSEHOLD LEVEL REGRESSIONS

The following equation estimates the effect at the woreda (district) level of receiving the program.

$$Y_{hwt} = \alpha_0 + \alpha_1 Y_{hw,t-1} + \eta \cdot X_{hw,t-1} + \beta \cdot T_w + \gamma_s + \mu_{wt}. \tag{1}$$

Here Y_{hwt} is the outcome of interest for household h, in woreda (district) w at time t. Following McKenzie (2012), we control for the baseline measure of the outcome of interest at

the household level, $Y_{hwt,t-1}$ (when possible)[2] and for a full set of household characteristics at baseline $X_{iw,t-1}$, and include fixed-effects for the 10 sub-cities into which woredas are organized and by which treatment was stratified (Bruhn and McKenzie, 2009).[3] T_w indicates that the program was implemented in the woreda in year 1, and captures its effects. We cluster standard errors at the level of woreda w (the level of randomization).

Our preferred specification for the estimate of the effects of the program at the household level, separately identifies the effect for those in the public works (PW) component and for those receiving direct cash support (DS).

To do this we compare households selected for the program in year 1 (treated) woredas to households that were selected for the program in year 2 (control) woredas, but who had not been targeted at the time of the endline survey. To do this we exploit the fact that we have gathered data on year 2 (control) households after both the main endline survey, and after targeting took place in those area. We do this to seperately identify public works and direct support beneficiaries in the treatment and control areas. Under the assumption that targeting followed the same proceedures in year 1 and year 2, we identify the direct effect of each arm of the program by estimating the following saturated model:

$$
\begin{aligned}
Y_{hwt} = \alpha_0 + \alpha_1 Y_{hw,t-1} + \delta_1 \cdot \text{PW}_h + \delta_2 \cdot \text{DS}_h \\
\beta_1 \cdot \text{PW}_h \cdot \text{T}_w + \beta_2 \cdot \text{DS}_h \cdot \text{T}_w + \beta_3 \cdot \text{C}_h \cdot \text{T}_w \\
+ \gamma_s + \eta \cdot X_{hw,t-1} + \mu_{wt}.
\end{aligned}
\tag{2}
$$

Here Y_{hwt} is the outcome of interest for household h, in woreda (district) w at time t. The key coefficients of interest are β_1, β_2, β_3, which estimate the effects of the public works component, the direct cash transfers component, and individuals who don't receive the program in treated woredas, respectively. Untreated households in year 2 woredas (C_h) are the omitted category.[4]

---

[2]For a small subset of outcomes, we did not measure $Y_{hwt,t-1}$ at baseline. Further, for a subset of households in our baseline we collected data only with a short-form questionnaire, so we do not measure $Y_{hwt,t-1}$ for certain outcomes. In these cases we will impute these missing baseline values to the mean.

[3]We include in our list of baseline controls the following characteristics for the household head: age, gender, labor status (employed), dummies for primary school and high school completion. We also include the maximum years of education in the household, total food expenditure per adult equivalent, and dummies for: household rents from local government, floor is made out of good materials and household has improved sanitation.

[4]We have heard that some households were added to the program in the course of the year. Therefore, there may be more households enrolled in the program at the time of the endline in year 1 woredas, compared to the number of households enrolled in year 2 woredas at the time of the phone survey immediately after targeting took place for these areas. Therefore, in order to measure the treatment status of households consistently, we will rely on the phone survey data collected throughout the year, to measure treatment status among households year 1 woredas, captured at a similar time relative to targeting as the data that we use for year 2 woredas.

4.2 INDIVIDUAL LEVEL REGRESSIONS

We will estimate the effect of the program (varying at the household level) on individual household members' labour, time use, and other individual outcomes using the following specification:

$$
\begin{aligned}
Y_{hiwt} = {} & \beta_1 \cdot \text{PW}_h \cdot \text{T}_w + \beta_2 \cdot \text{DS}_h \cdot \text{T}_w + \beta_3 \cdot \text{C}_h \cdot \text{T}_w \\
& + \delta_1 \cdot \text{PW}_h + \delta_2 \cdot \text{DS}_h \\
& + \gamma_s + \eta_1 \cdot X_{iw,t-1} + \eta_2 \cdot X_{hw,t-1} + \mu_{iwt}.
\end{aligned}
\tag{3}
$$

Where Y_{hiwt} is the outcome of interest. For labour outcomes we will restrict the sample to working-age individuals i in household h. We continue to cluster at the level the woreda w.

Using the specification above, we can introduce interactions between individual characteristics at baseline (for example, gender) with the each of the treatment categories at the household level. This will allow us to investigate the characteristics of individual household members driving the main household effects on items such as labour supply. We will estimate whether changes in household earnings are driven by changes among male or female members, or both.

Most importantly, we will look at heterogeneity of impacts on non-public works labour outcomes by whether the respondent was a public works participant or not. This will tell us whether any changes in household labour supply are due to overall household income effects or by participation in the program itself. Using our second endline of the control group after they were already engaged in the program, we will be able to compare individuals doing the public works in year 1 with households that would eventually do the public works when offered in year 2. Denoting PAR_{ih} as the indicator for whether individual i in household h did the public works, we will estimate:

$$
\begin{aligned}
Y_{hiwt} = {} & \alpha_1 \cdot \text{PAR}_{hi} \cdot \text{T}_w + \alpha_2 \cdot (1 - \text{PAR})_{hi} \cdot \text{T}_w + \delta_0 \cdot \text{PAR}_{hi} \\
& + \beta_2 \cdot \text{DS}_h \cdot \text{T}_w + \beta_3 \cdot \text{C}_h \cdot \text{T}_w \\
& + \delta_1 \cdot \text{PW}_h + \delta_2 \cdot \text{DS}_h + \delta_3 \cdot \text{C}_h \\
& + \gamma_s + \eta_1 \cdot X_{iw,t-1} + \eta_2 \cdot X_{hw,t-1} + \mu_{iwt}.
\end{aligned}
\tag{4}
$$

4.3 HETEROGENEITY

We will explore how the short term impact of the UPSNP at the household level varies along several dimensions, in order to determine if there are specific subgroups that are particularly benefited by the program. The main dimensions we will explore are:

1. Household head gender.

2. Household head employment at baseline.

3. Female respondents with no male spouse compared to female respondents with a male spouse in the household. (Excluding male headed households with no female spouse, which are approximately 5% of our sample.)

4. Total consumption per adult equivalent at baseline (split at the median).

5. Education level of household head at baseline: household head has any formal education other religious education and literacy programs.

In the case of individual level regressions: gender, work at baseline (self-employment or wage employment), education level, relationship to the household head, among others.

4.4 COMPARING PHONE AND ENDLINE LABOUR OUTCOMES

For a sub-sample of households we collect data on labour supply and earnings in a mid-line phone survey, at a time when they were enrolled into the program and were receiving money, but were only required to complete two hours of work per day in order to receive the money. For this sub-sample, we will compare the crowding-out effects of the program under the conditions of a two hour two requirement and the six to eight hour requirement in place at the time of the endline survey. Because the phone surveys took a long time to conduct, not all households were observed in this survey on either side of the increase in time requirements. However, due to the random ordering of the phone survey listing, our phone survey is comparable across treatment and control areas. It is worth noting that we were only able to collect information on the earnings and labour of the household head and his or her spouse, as we found that the household head was unable to reliably report on the earnings of other members of the household.[5] For that reason, when looking at the phone survey data, we will focus the analysis on the household head and spouse.

4.5 MULTIPLE HYPOTHESIS TESTING

In order to avoid problems of multiple inference we will focus on the primary outcomes listed in this analysis plan. In addition to reporting standard p-values, we will report False Discovery Rate q-values across our primary outcomes, and within families of outcomes Benjamini et al. (2006). We apply the adjustment separately for each coefficient of interest: first, the district (woreda)-level average treatment effect estimated by coefficient β in Equation 1, and then for each of the coefficients β_1, β_2, and β_3 in Equation 2.

---

[5]By contrast, in the main endline survey we were able to collect data on all household members by either ensuring that they were present at some time during the household interview, or calling them on the phone, when they were not available in person.

5 OUTCOMES OF INTEREST

5.1 PRIMARY OUTCOMES

We will study the effects of the UPSNP on seven primary outcome families of outcomes, for which we summarize the primary outcomes below.

1. Total household income: Total household income monthly income. Sum of income across public works income, non-public works labour income (including both wage and self-employment), direct support transfers, and other sources of income such as transfers, remittances, interest, and pensions.

2. Non-public works labour supply and earnings: Summary index (weighted standardized) of hours and earnings per working age adult in non-public works occupations. That is, monthly hours worked per adult (excluding public works) and monthly earnings per adult (excluding those from public works).

3. Consumption: Total household consumption over the last 12 months, which includes the sum of total food consumption in and out of the household, and total non-food consumption.

4. Subjective well being: Summary index (weighted standardized) of self-reported measures of well being at the individual (respondent) level, including life-satisfaction and mental health indicators, all signed so that positive values represent better well-being.

5. Gender: Summary index (weighted standardized) of female empowerment, including attitudes towards female labor participation, towards female autonomy and towards participating in the community's public discussions, all signed so that positive values represent more favorable attitudes towards women.

6. Neighbourhoods: Summary index (weighted standardized) of the neighbourhood conditions, comprising the quality of sewerage and drainage systems, cleanliness of streets, availability and quality of public toilets, notices smell of drains or sewerage in local area (reverse coded), notices smell of solid waste or trash in local area (reserve coded). All components have been adjusted so that a higher level represents better conditions.

7. Attitude towards welfare: Summary index (weighted standardized) of two attitudes associated with the stigmatization of welfare recipients: people who participate in the UPSNP are respected members of the community and there is some yilunta (shame) associated with participation in the UPSNP (reverse coded). The components are signed in such a way that higher values represent less stigma.

5.2 Secondary outcomes

As secondary analysis we will look at the effects of the program on household assets and human capital investments.

8. Savings: Total household savings across all financial instruments, including: cash on hand, bank accounts, savings cooperatives, microfinance and building societies.

9. Asset ownership: Standardized index of household assets across.

10. Education: share of school-aged children in the household enrolled in school.

11. Health: Summary index (weighted standardized) of health, including the proportion of household members that suffered from ill health (preventing them to engage in normal activities) in the last 30 days, the number of days that the illness prevented households member from performing normal activities (normalized by household size) and the proportion of illness in which treatment was sought. All components have been adjusted so that a higher level represents better health conditions.

6 Detailed Description of Outcome Families

In this section, we provide a more detailed description of our primary outcomes of interests, the components of our primary outcome measures, and secondary outcomes and tests to be studied within the same family of outcomes. The appendix to this document provides more detail on how we code each of these measures, with the accompanying variable names.

All monetary values will be adjusted by monthly inflation, obtained from the Central Statistical Agency (CSA) and as is common practice, the monetary variables will be winsorized at the top 1 percent. Unless otherwise stated, we use the procedure proposed by Anderson (2008) to construct summary indexes for most primary outcomes. The summary index is a weighted mean of several standardized outcomes (by the control group standard deviation), where the weights are designed to maximize the information contained in the index.

6.1 Income

Primary measure: Total annualized income per capita from labour (public and non-public works), direct cash support from the UPSNP, remittances, investments, pensions, gifts and all other transfers.

6.2 Labor and Earnings

Primary measure: Index of hours and earnings per working age adult in non-public works occupations (items 1 and 2 below).

1. Earnings per adult (excluding public works).

2. Hours per adult (excluding public works).

3. Employment per adult (excluding public works).

4. Total earnings (public works and non-public works) per adult.

5. Total hours (public works and non-public works) per adult.

Using individual household member data, we will look at heterogeneity in the treatment effects on non-public works labour among individuals by:

- Whether household members engaged in the public works.

- Gender

Using individual data we will also look at the time-use of the household head and spouse (including for male and female members separately). We will look at time spent on:

- All work for pay.

- Work in the home.

- Leisure.

- Sleep.

6.3 Consumption

The increase in earnings due to the participation of households in the public works program will likely result in higher total consumption expenditure (per adult equivalent) for beneficiary households. If the program had a positive impact on total consumption, we will explore if the effects were larger in the case of food consumption, compared to non-food consumption, as poorer households usually devote a larger share of resources to food items.

  Primary measure: Total expenditure per adult equivalent converted to annual values. We calculate this as the sum of items 1 and 2 below:

1. Total food expenditure: The sum of household consumption of 27 food items (inside the house) in the last seven days -including cereals, vegetables, meat, sugar, oil, etc. - and of the total food consumption outside of all household members in the last seven days, converted to yearly values.

2. Total non-food expenditure: Sum of total medium term consumption in the last month, which includes rent, transport, utilities, fuel, etc. converted to yearly values, and total long term consumption in the last 12 months, which includes clothing, durable goods, education and health expenditure among other items.

We winsorize the top 1 percent of the monetary outcomes to deal with outliers.

As part of the secondary analysis we will look at expenditure on the following sub-groups of expenditure: health, education, housing and durable goods. Also, we will also study the impact of the program on food security. Here we will consider a weighted standardized index of the following food security outcomes at the household level (all of which have been properly adjusted such that higher values represent greater food security:

1. Number of dietary groups consumed by the household in the last seven days. We classified all food items purchased in the last seven day into one of eight dietary groups, except for spices, coffee and soft drinks.

    (a) Cereals, grains, maize grain/flour, millet, sorghum, flour, bread, pasta, roots, tubers, and plantains.
    (b) Nuts and pulses
    (c) Vegetables
    (d) Meat, fish, other meat, eggs
    (e) Fruits
    (f) Milk Products
    (g) Fats and oils
    (h) Sugar, sugar products and honey

2. Following binary variables (0, -1 if action was taken). Over the last seven days (30 days in the case of the first item), the households reports:

    (a) Worrying that household would not have enough food,
    (b) Relying on less preferred and/or less expensive foods,
    (c) Limiting portion size at meal-times/reducing number of meals eaten in a day,
    (d) Restricting consumption by adults so that small children may eat,
    (e) Go a whole day and night without eating anything.

As a robustness check and in order to have comparable results with similar programs in the region, we will estimate the effect on the World Food Program (WPF) food security score. [6]

---

[6] The food security score follows the WFP guidelines and takes a value of- 1,- 2,- 3, or- 4 (a lower absolute value indicates greater security). The food security score is - 1 if in the past seven days, the household reports not worrying about having enough food and reports no days in which they: (a) rely on less preferred and/or less expensive foods, (b) limit portion size at meal-times/reduce number of meals eaten in a day, (c) restrict consumption by adults so that small children may eat, or (d) Go a whole day and night without eating anything. The food security score is - 2 if the household reports that it worried about having enough food and reports no days in which they resort to actions a-d. The food security score is - 3 if the household reports that it relied on less preferred and/or less expensive foods and but not on b-d . The food security score is - 4 if the household reports relying on b-d. Larger values indicate improved food security.

6.4 Self-reported Well-being

We measure well-being for the main survey respondent, the household head or main household decision maker.

Primary measure: weighted standardized average of life-satisfaction, depression and anxiety, signed so that positive values represent better well-being (items 1-3 below).

1. Life-satisfaction on a ladder scale.

2. Anxiety index.

3. Depression index.

4. Change in social ties in the last year (self-reported)

5. Membership of local community group.

6.5 Women's Empowerment

We understand that around 85 percent of the people undertaking the public works are women, which likely translated into an increase in the earnings that female members are contributing to the household. In turn, this likely affected the beneficiaries' attitudes towards female labor participation, female autonomy and participation in the community.

Primary measure: a weighted standardized index, comprising the following variables 3 sub-indices:

1. Attitudes towards female labor participation.

2. Attitude towards female autonomy, and

3. Attitudes towards participating in the community's public discussions.

In the secondary analysis we will also look at beliefs on gender norms/roles, perception of female safety, and female control of household assets.

Also as part of the secondary analysis, we will explore the effects of the program on intra-household allocation. Although the large majority of our sample are households conformed by non-married females (widowed, divorced, separated and single) and their relatives, around 40 percent of households are comprised by the traditional married couple. We will explore if the program increased the women's bargaining power, and thus female decision making power, in these households, using an index summarizing the women's decision making power within the household (see appendix).

We will explore if the program affects gendered consumption (spending on girls relative to boys, and spending on women relative to men) and on the spouse (male) gender norms and attitudes towards women. The latter is possible as we collected data for a subset of male spouses on this particular topic.

6.6 Neighborhood Quality and Public Goods

Primary measure: index of items 1-5 below:

1. Quality of sewerage and drainage systems

2. Cleanliness of streets

3. Availability and quality of public toilets

4. Respondent notices smells of drains or sewerage in local area (negatively signed)

5. Respondent notices smell of solid waste or trash (negatively signed) in local area

Here we place particular emphasis on households not selected for the program but living in treated areas, to assess whether they, as well as treated households perceive changes in neighbourhood outcomes.

Secondary outcomes

- Willingness of respondent to contribute to community and public goods: a) picking up litter, and b) spending time doing community maintenance.

- Respondents' belief about neighbours' willingness to spend time doing community maintenance.

- Neighbourhood goodwill: a) how often neighbours are caring or concerned, b) how often can you turn to neighbours for advice, c) neighbourhood conflict (negatively signed).

- Neighbourhood crime index: regularity of theft in home, regularity of theft in the street, feeling physically safe in home (negatively signed), feeling physically safe in the street (negatively signed).

6.7 Attitudes to Welfare and Welfare Recipients

Primary measure: Level of stigma associated with welfare recipients: index of two questions: (1) Respondent thinks that people who do the UPSNP are respected members of the community. (2) Respondent thinks there is some yilunta (shame) associated with participation in the UPSNP.

Secondary outcomes:

1. Support for higher levels of redistribution to poor (index).

2. Willingness to pay taxes to support the poor.

3. Pride in vocation among welfare recipients.

4. Believes that welfare should come with a work requirement.

Here we place particular emphasis on households not selected for the program but living in treated areas, to assess whether they, as well as treated households, change their attitudes to welfare after observing the program in their neighbourhood.

6.8 SAVINGS (SECONDARY FAMILY)

Primary measure: Total household savings across the following savings instruments: Cash, Banks (any formal bank), Building Society (housing building asso, Government housing program savaings (20/80, 40/60, etc), Ekubs, Iddirs, Microfinance programs, Government Bonds (such as Hidase bond), Small Medium Enterprises saving programs.

6.9 ASSET OWNERSHIP (SECONDARY FAMILY)

Primary measure: Asset ownership index: weighted index of the number of assets owned by the households across 27 asset types.

We then look separately at set of an index assets that could be used for income generating activities: Car, Computer/laptop, Cell phones, Smartphones, Bajaj, Brewing equipment, Sewing machine, Storage facility, Small shops, Weighing machine, Hand tools (such as spade, shovel, fork), Wheelbarrow, Satellite dish, Chair, Weaving machine, Laundry machine, Livestock. Finally, we will look at an index of assets, from the list of above, that households report to use in income generating activites.

6.10 EDUCATION (SECONDARY FAMILY)

Primary measure: The share of school-aged children (from 7 years to 16 years of age) enrolled in school.

In Ethiopia, schooling consistents of a primary cycle until age 14 (grade 8), at which point students transition to the secondary cycle for ages 15 and 16 (grades 9 and 10). While an increasing number of students transition to secondary cycle, drop out rates around the age of 14 and 15 are still high. Futhermore, students intending to go on to tertiary education enroll in secondary preparatory school for ages 17 and 18 (grades 11 and 12). Therefore, we look primarily at school attendance rates from ages 7 to 16, when most students are expected to be in school, but also focus specifically on students from ages 14 to 18, when students make key decisions about whether to enter the labour force or to continue with tertiary education.

Secondary outcomes:

- the share of 14-18 years old enrolled in school or further education.

- The share of school-aged children (from 7 years to 16 years of age) in the right grade for their age [7]

---

[7]The correspondence is age 7 - Grade 1, age 8 - Grade 2, age 9 - Grade 3, age 10 - Grade 4, age 11 - Grade 5,

- Education expenditure per school-age child.

6.11 HEALTH (SECONDARY FAMILY)

In the case of health, as primary outcome we will use a weighted standardized index of three health outcomes:

1. Proportion of household members that suffered from ill health (preventing them to engage in normal activities) in the last 30 days. Signed so that a higher indicator signals better health outcomes.

2. The number of days that the illness prevented households member from performing normal activities in the last 2 months (normalized by household size). Signed so that a higher indicator signals better health outcomes.

3. Proportion of illness for which treatment was sought.

As a secondary indicator, we will explore the effects of the program in the per capita health expenditure of the household. We anticipate that participating in the program may increase the number of illness episodes (connected with the participation of in the public works program) but at the same may allow otherwise budget constrained households to seek medical treatment, and therefore improve health outcomes.

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.

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

---

age 12 - Grade 6, age 13 - Grade 7, age 14 - Grade 8, age 15 - Grade 9, age 16 - Grade 10. We do not consider secondary preparatory.