Welfare, Work, and Wellbeing:

Evidence from an Informal Settlement in Kenya

PRE-ANALYSIS PLAN

Syon P. Bhanot*, Alexandra De Filippo†, Jiyoung Han‡

Chaning Jang§, Claudia Newman-Martin¶

July 24, 2015

Abstract

Governments have long relied on cash transfers and vouchers as key elements of social welfare systems. Consumer theory offers predictions about the impact of such programs on wellbeing and spending behavior, but there is little real-world evidence. In this paper, we outline the pre-analysis plan for a randomized field experiment to explore the effect of welfare program design on these two key outcomes. We specifically test three design features: 1) workfare versus welfare; 2) restricted versus unrestricted vouchers; and 3) voucher messaging with a focus on spending on oneself or on others. The ten-day experiment involved 432 individuals living below the poverty line in the Kawangware settlement of Nairobi, Kenya. Each day, subjects either sorted lentils or waited in a room for benefit payments, provided as vouchers to a local grocery store. We either restricted the vouchers to certain basic food items or did not restrict them, and we also randomized voucher messaging. We outline the analysis plan of the impact of these experimental treatments on happiness, productivity, and consumption decisions.

---

*Swarthmore College, email: sbhanot1@swarthmore.edu
†Behavioural Insights Team, email: ale.defilippo@gmail.com
‡ideas42, email: jiyounghn@gmail.com
§Princeton University, email: cjang@princeton.edu
¶Boston Consulting Group, email: claudia.newman.martin@gmail.com

1 Introduction

It is important for governments, non-profits, and philanthropic organizations to understand how the structures of social welfare programs influence the wellbeing and decision making of their target populations. Recently, academic research in development has focused on the impact of low-cost interventions on wellbeing (Dupas and Cohen, 2010; Banerjee and Duflo, 2009), but there is little concrete evidence on how the design of social welfare program influences key behaviors. For example, while there is significant work in behavioral economics and psychology on the effects of framing on individual decision making (Tversky and Kahneman, 1981), the impact of welfare voucher framing on recipients' decision making remains largely unexplored. Additionally, work in behavioral economics on mental accounting has explored theories around how earned money may be spent differently than granted money, which has implications for the impacts of social welfare programs and the benefit they provide to recipients. Finally, while there is a large body of research on conditional cash transfers in development (Lagarde, Haines, Palmer, 2007; Gertler, 2004; Blattman and Niehaus, 2014), there is a lack of experimental evidence on how conditionality in voucher payments might influence consumption decisions. The lack of research in these areas stems largely from the fact that publicly available randomized evaluations are uncommon amongst bodies that oversee welfare programs.

In this study, we conduct a randomized field experiment that seeks to mimic a variety of possible real-world social welfare programs to provide insights on the impacts of different ways to design and implement such programs for low-income beneficiaries. Conducted over a ten-day period in a low-income area in Nairobi's Kawangware slum,[1] the experiment was designed to contribute to our understanding of the effects of three specific elements of welfare program design on decision making: 1) the differences in spending behavior among those "working" for their welfare, versus those simply receiving it; 2) the effect of voucher messaging and framing on spending decisions; and 3) the effect of conditional (restrictions to certain categories of spending) vs. unconditional vouchers on subsequent purchasing behavior.

The experiment was conducted in partnership with the Busara Center for Behavioral Economics and the PBK Nonic Supermarket in Nairobi, Kenya. Recruited subjects were drawn from the Busara subject pool. The sample consisted of 432 individuals residing almost exclusively in the Kawangware slum where the experiment was conducted. Subjects were instructed to show up every day for one hour, for ten consecutive weekdays, at one of three research sites (one site per treatment). There were four possible timeslots in the day from which subjects could choose at the start of the experiment. Of the 432 subjects recruited, 379 showed up for at least one of the ten sessions, and 263 showed up for all ten sessions as instructed.

---

[1]The authors partnered with the Busara Center for Behavioral Economics, which has over three years of experience in conducting randomized lab and field studies amongst very poor populations in East Africa.

Subjects were randomized into three main experimental treatment arms, hereafter referred to as the "Work," "Wait-Unconditional," and "Wait-Hybrid" treatments. Subjects in the "Work" treatment sorted rice and lentils in order to receive two unconditional PBK vouchers per daily session, worth KES 100 each (KES 200 in total). Subjects in the "Wait − Unconditional" treatment waited for the same period of time (i.e. they sat idle, but did not work) in order to receive two unconditional PBK vouchers per day, worth KES 100 each (KES 200 in total). Finally, subjects in the "Wait − Hybrid" treatment received the same treatment as those in the "Wait − Unconditional" treatment, but received one unconditional PBK voucher (worth 100 KES) and one conditional PBK voucher (worth 100 KES, but only to be used on "basic goods," namely maize/wheat flour, rice, sugar, and cooking oil) per day.

In addition, two forms of vouchers were used − one version that emphasized spending money on oneself ("Self" vouchers), and one version that emphasized spending money on one's family ("Family" vouchers). Within each treatment, subjects were randomly assigned to receive "Self" vouchers either in the first week of the study (the first five days) or the second week of the study (the last five days). When the "Self" voucher was not received, the "Family" voucher was received. This means that we observe behavior for each subject receiving each of the two voucher types during the experiment.

The experimental conditions were designed to provide insights on three possible effects of social welfare program design on outcomes. First, the conditions sought to investigate the effect of working for payment rather than receiving a welfare "handout" on subject expenditures. Notably, a direct comparison of behavior under the "Work" and "Wait-Unconditional" treatments provides insight on this question. Second, the experimental setup also enables us to test the effect of voucher conditionality on spending decisions, by comparing outcomes in the "Wait-Unconditional" and "Wait-Hybrid" treatments. Third, varied messaging on the vouchers, encouraging subjects to spend money on themselves or to spend money on their families, was designed to test the effect of framing on decision making.

We measured three outcomes for each individual in the experiment. First, we tracked participants' spending behavior as they redeemed their grocery vouchers. Second, we measured happiness in two ways − through daily surveys on satisfaction and an end-of-study survey with broader happiness measures. Third, we tracked effort and engagement with the program through measures of attendance and timeliness for all treatments, as well as work effort exerted (for the Work treatment only).

This document serves as a pre-analysis plan for the experiment desscribed above. In the sections to follow we will outline the details and timeline of the experiment, outline the major questions, describe the econometric approach, and identify the main outcome variables.

2 Experiment Overview

2.1 Experiment Design

Partners

The authors partnered with two organizations: the Busara Center for Behavioral Economics and the PBK Nonic Supermarket. The Busara Center for Behavioral Economics has over three years of experience in conducting randomized lab and field studies amongst very poor populations in East Africa. PBK Nonic supermarket is a privately held local store in Kawangare that sells a large range of household goods and groceries at prices accessible to the local population (and subjects of this study) .

Subjects

The subjects in this experiment were 432 individuals living in the Kawangware area of Nairobi, Kenya, selected at random for recruitment to the study from the subject pool maintained by the Busara Center for Behavioral Economics.[2] Subjects were initially contacted by phone and asked if they were available and willing to participate in a study.

Study Design

The field experiment was conducted in Kawangware, an informal settlement located approximately 15 kilometers from the Nairobi city center. Participants were randomly assigned to three treatment groups, described in detail below, located in three different community halls. Each community hall hosted four sessions per day at the same times each day: 9:30 AM, 11:30 AM, 1:30 PM, and 3:30 PM. Each hall was within 5-10 minutes walking distance to the PBK Nonic Supermarket, but separate from one another to limit the potential for contamination between treatment groups. The experiment took place over ten consecutive business days during the two week period from September 22, 2014 through October 3, 2014.

After subjects were recruited by phone, they were asked to select one of four time slots during the day for their participation in the study. Once participants were confirmed for a given time slot, they were randomly assigned to one of the three treatment groups and notified of the location where they had to report every day. The assignment to one of three treatment groups were not known to the participants prior to arrival

---

[2]Busara recruits study participants periodically throughout the year from different areas of Nairobi, including some university students, residents of Kibera, and residents of Kwarangare. Busara and local community liaison officers engaged in a large recruitment effort in September 2014 to recruit participants for this study. Of those registered into the database, Busara randomly selected a subset to call to invite to the study

at the location. This design helped avoid any selection bias stemming the time of day that people chose for their participation in the study.

Treatment Groups

Subjects were randomized into three main treatment groups, namely:

1. Work: People who worked in the treatment location for one hour each day in exchange for a daily payment of two unconditional PBK Nonic Supermarket vouchers worth KES 100 each (KES 200 in total).

2. Wait – Unconditional: People who waited in the treatment location for one hour each day in exchange for a daily payment of two unconditional PBK Nonic Supermarket vouchers worth KES 100 each (KES 200 in total).

3. Wait – Hybrid: People who waited in the treatment location for one hour each day in exchange for a daily payment of one unconditional PBK Nonic Supermarket voucher worth KES 100, and one conditional PBK Nonic Supermarket voucher worth KES 100 that could be used for "basic" food items only (Maize/wheat flour, rice, sugar, or cooking oil).

The total face value of the vouchers provided each day to the subjects was therefore constant across treatments (200 KES), and was an amount slightly higher than an average daily wage for most participants.

The on-the-ground experimental procedure for each treatment is outlined below.

Treatment 1: Work

Individuals were asked to arrive at the designated community hall for one hour every day for ten weekdays over a two-week period. After checking in with field officers, they were required to sort and separate a mix of rice and lentils into small cups. Field officers distributed large plastic cups filled with the rice and lentil mix to all participants at the beginning of every session and two smaller plastic cups for participants to collect the separated grains.

All participants in the sessions had space to work. Past studies with this population suggested that men might find doing such work in the presence of females embarrassing. To avoid such an distortionary effect on men's motivation or productivity, the men and women in the sample were separated in the hall. Field officers recorded the time when each participant started and stopped sorting grains. At the end of the work period, field officers would note the session conclusion time and measure the weight of rice and lentils that were sorted. This allowed field officers to calculate grams sorted per minute as a measure of productivity. Field officers were present at all times during the work sessions and participants knew that their output

would be weighed. At the end of each day's work, participants were given two unconditional PBK Nonic Supermarket vouchers worth KES 100 each (regardless of how much they sorted).

Treatment 2: Wait – Unconditional

Individuals were asked to arrive at the designated community hall to collect vouchers every day for ten workdays over a two-week period. Upon arriving at the hall, field officers asked participants to wait for their vouchers without providing any concrete justification for the delay. Occasionally, field officers would ask subjects to come to the front of the room to confirm their name, or write their name/signature on a sheet of paper. This was done to mimic the tedium of a traditional welfare "ordeal" process, as outlined in Nichols and Zeckhauser. Field officers were present at all times during the wait session. After an hour wait, participants were called to the front by name and given two unconditional PBK Nonic Supermarket vouchers worth KES 100 each.

Treatment 3: Wait – Hybrid

Treatment 3 was identical to Treatment 2 in every way, except that instead of receiving two unconditional PBK Nonic Supermarket vouchers, participants received one 100 KSH unconditional voucher and one 100 KSH conditional voucher, restricted to the purchase of maize/wheat flour, rice, sugar, or cooking oil only.

Cross randomization of framing effects

To test framing effects, participants were randomly assigned different messaging on their vouchers. There were two types of voucher messaging. The first version, the "Self" voucher, had an image of a woman with money and a message in text beneath the image that read, "Treat yourself to something nice!" The second version, the "Family" voucher, had an image of a family eating together and a message in text beneath the image that read, "Be responsible! Bring home something for the family!".

Subjects were randomized into two groups within treatments - one half of each treatment group received the "Self" voucher for the first five days of the study and the "Family" voucher for the last five days, and vice versa for the second group. The messaging on the vouchers was briefly reinforced by field officers, particularly for illiterate subjects. However, the field officers clarified that subjects were in no way required to purchase any items or types of items in particular.

Attendance and Attrition Prevention

To discourage attrition and encourage regular attendance, subjects who completed the study received a bonus payment. Specifically, those who attended the sessions for ten, nine, or eight days were provided with an

unconditional bonus voucher for the PBK Nonic Supermarket worth 500, 200, and 100 KSH, respectively, at the end of the study. These vouchers were also randomly assigned to be either "Self" or "Family" vouchers. Given that only individuals who regularly attended sessions were given bonus vouchers, this sub-group may be non-representative of the broader study population. We classify people who attended frequently as "high compliers." An analysis of the differential impact of the "Self" and "Family" voucher versions for "high compliers" is conceptually interesting because it may help us understand if such individuals are naturally more inclined to prudent spending, despite labeling encouraging them to focus on themselves.

Over the course of the experiment, 379 subjects of the 432 subjects recruited (87.7%) attended at least one of the ten sessions, while 269 subjects (62.3%) had perfect attendance. In the following section we will discuss our approach to quantifying and adjusting for possible differential attrition. The figure below outlines the number of individuals who attended for each of the possible number of days:

Figure 1: Attendance

	DAYS OF ATTENDANCE										
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Number of Subjects	53	13	3	6	8	23	8	4	10	35	269
Percent of Subjects	12.3%	3.0%	0.7%	1.4%	1.9%	5.3%	1.9%	0.9%	2.3%	8.1%	62.3%
Cumulative Percent of Subjects	12.3%	15.3%	16%	17.4%	19.2%	24.5%	26.4%	27.3%	29.6%	37.7%	100%

Outcomes I: Expenditures

During the course of the experiment, PBK accepted the vouchers distributed during the experiment and redeemed the value of the vouchers for any item (or the set of restricted items), according to the instructions on each voucher. Vouchers were labeled with ID numbers reflecting the individual's unique identifier, the treatment group, and the date of issuance in a manner that was not transparent to subjects.

Every time a subject paid for his/her purchases with vouchers, PBK Nonic Supermarket staff stapled the receipt to the voucher. Next, a Busara research assistant with experience conducting randomized field experiments and familiarity with the local culture reviewed the vouchers and attached receipts for errors. The research assistant then collected all vouchers at the end of the day and returned them to the Busara office.

Participants could redeem their vouchers anytime from the start of the experiment until a week after the experiment concluded. This timeline gave participants up to three full weeks to redeem their vouchers. Through this partnership with PBK Nonic Supermarket, we were able to track the proportion of the vouchers spent on essential versus non-essential goods, along with the voucher type (Self or Family, conditional or unconditional), amount spent, date redeemed, and other information associated with the subject. This data served as basis for the analysis on consumption.

Outcomes II: Baseline, Daily, and Endline Surveys

Three survey types were administered during the study. First, on the initial day of the study, participants completed a Baseline survey. The Baseline survey asked questions related to people's current emotions (using a PANAS scale), weekly spending habits, employment, household characteristics, familiarity with the PBK Nonic Supermarket, and decision-making power within the household.

Second, every day of the study after the initial day, participants completed a Daily survey that asked the same "current emotions" questions from the Baseline survey, but did not include the additional questions in the Baseline survey.

Third, at the end of the study participants completed an Endline survey, which asked a series of questions on self-esteem, general happiness, and optimism. The Endline survey was designed to measure overall life satisfaction and wellbeing, rather than incidental happiness, enabling us to distinguish between "fleeting" effects of the treatments on wellbeing as opposed to longer run effects on disposition. The Endline survey also asked about family dynamics, income levels, how subjects approached spending the vouchers, and how they felt about their consumption decisions.

Outcomes III: Effort and Worker Productivity

Two measures of effort and productivity were collected. First, for all treatment groups, attendance and timeliness was tracked by on-site field officers. Second, for the Work treatment only, the productivity of each worker was measured by: 1) weighing of the amount of lentils and rice sorted in the work period; 2) keeping track of the precise time spent sorting for each subject; and 3) computing grams sorted per minute as a proxy for productivity/efficiency.

3 Econometric approach

3.1 Attrition

To assess whether attrition potentially confounds our results, we proceed as follows. First, we define $attrit_{it} = 1$ if individual i was absent on day t, and zero otherwise. We then assess the severity of attrition using three approaches. First, equation 1 estimates whether the magnitude of attrition is different for treatment groups and the control groups:

$$attrit_{it} = \beta_0 + \beta_1 WORK_i + \beta_2 COND_i + \varepsilon_{it} \tag{1}$$

Second, equation 2 assesses whether attrition individuals for a given day t are different in terms of a comprehensive range of baseline characteristics:

$$y_{i,t=0} = \beta_0 + \sum_{k=1}^{10} \beta_1^k (attrit \times [t = k]) + \varepsilon_{i,t=0} \tag{2}$$

And third, equation 3 measures whether the baseline characteristics of attrition individuals in the treatment group are significantly different from those in the control group. The sample for regression will be restricted to attrition individuals:

$$(y_{i,t=0} \mid attrit_{it} = 1) = \beta_0 + \beta_1 WORK_i + \beta_2 COND_i + \varepsilon_{i,t=0} \tag{3}$$

We will also seek to measure how attrition manifests in days attended. This will be measured by examining the average number of days completed by people in each treatment group as well as the the distribution of days completed across the treatment groups.

If worrying levels of attrition are found, we will adjust for the potential effect of such attrition by bounding the treatment effects, modeling the selection process, or using matching techniques.

3.2 Basic Specification

Our basic treatment effects specification estimates the following equation:

$$y_{it} = \beta_0 + \beta_1 WORK_i + \beta_2 COND_i + \varepsilon_{it} \tag{4}$$

where y_{it} is the outcome of interest for individual i on day t. $WORK_i$ and $COND_i$ are dummy variables equal to 1 if the participant was randomly assigned to the work and wait (conditional) condition, respectively, and 0 otherwise. Note that a dummy variable for wait (unconditional) is not included in the model to avoid multicollinearity. ε_{it} is the unobserved error component, which is assumed to be serially uncorrelated. Where possible, we will control for baseline levels of the outcome variables, $y_{i,t=0}$ to improve statistical power. The modified estimating equation is:

$$y_{it} = \beta_0 + \beta_1 WORK_i + \beta_2 COND_i + \delta y_{i,t=0} + \varepsilon_{it}. \tag{5}$$

We will also pool across "wait" conditions to improve statistical power. We will also control for time-of-day effects where necessary. We also have random variation in the voucher messaging as respondents either received vouchers encouraging "self-spending" or "family-spending" in the first week, and vice-versa in the second week. To analyze the effect of messaging on spending and subjective well-being, we propose the following specification:

$$y_{i,k \leq t < j} = \beta_0 + \beta_1 SELF_{g,k \leq t < j} + \delta y_{i,t=0} + \varepsilon_i \tag{6}$$

Where $SELF_{g,k \leq t < j}$ is a dummy variable equal to 1 if the participant was assigned to the "self-spending" vouchers during days $k \leq t < j$, and 0 if the participant was assigned to the "family-spending" vouchers. $k \leq t < j$ refers to the outcomes measured in the first $j - k$ days of surveys, or expenditures captured using vouchers from the first $j - k$ days. Restricting $j = 5, k = 0$ allows for between-group comparison to identify the effect on messaging and controls for the fact that there may be differential response to within-group order effects. For measures of self-reported SWB, we will use $j = 6, k = 1$ to account for the fact that SWB measures were taken prior to the receipt of the voucher on any given, thus any measures of SWB should be attributed to vouchers from the previous day.

We can also use a within-individual design that looks at how each participant responds differently to "self-spending" versus "family-spending" vouchers. To analyze the effect of messaging on spending and subjective wellbeing within the individual, we propose the following specification:

$$y_{igt} = \beta_0 + \beta_1 SELF_g + \beta_2 DAY_t + \beta_3 SELFDAY_{gt} + \eta_i + \delta y_{i,t=0} + \varepsilon_i \tag{7}$$

Where $SELF_g$ is a dummy variable equal to 1 for observations based on the self-spending vouchers and 0 for observations based on the family-spending voucher, DAY_t is a dummy variable for each study day t (1-10),

$SELFDAY_{gt}$ is an indicator equal to 1 for each day/voucher version combination (20 dummy variables - 10 days and 2 voucher versions, interacted), and η_i is a term representing individual fixed effects. This allows us to assess the average difference between the two observations per subject, namely spending when receiving the self-spending vouchers and spending when receiving the family-spending vouchers. Note we may also aggregate the outcomes by individual for each week, for simplicity, and run a version of the above regression with week fixed effects instead of day fixed effects.

3.3 Heterogeneous Treatment Effects

We will further test whether the impact of the psychological intervention varies with pre-determined individual characteristics, measured at baseline and denoted by $\mathbf{X}_{i,t=0}$. The estimating equation for differential effect of treatment for a particular characteristic is given by:

$$y_i = \beta_0 + \beta_1 WORK_i + \beta_2 COND_i + \beta_4 X_{i,t=0} + \beta_5 T_i \times X_{i,t=0} + \delta y_{i,t=0} + \varepsilon_i. \tag{8}$$

where β_5 captures the additional effect that treatment has for individuals with characteristic X. We also test heterogeneous effects on messaging order using a modified equation 6 which follows:

$$y_i = \beta_0 + \beta_1 SELF_g + \beta_2 X_{i,t=0} + \beta_3 T_g \times X_{i,t=0} + \delta y_{i,t=0} + \varepsilon_i. \tag{9}$$

The dimensions of heterogeneous effects are:

1. Respondent gender

2. Respondent age (above & below median)

3. Respondent baseline employment status

4. Family size (median number of children)

5. Baseline psychological well-being (index of the first daily well-being measure PANAS)

6. Respondent baseline education level

3.4 Temporal Dynamics of the Treatment Effect

With daily data on psychological affect, we will be able to explore the temporal dynamics of the treatment effect. That is, we can observe how the intervention impacts psychological well-being, spending, and at-

tendance over the two week period. In the daily data, we have outcome measures y_{it} for individual i for $t = 1, \ldots, 10$, where $t = 1$ is the measure after the first day of the intervention, and $t = 10$ is the measure at endline. We will estimate the following specification:

$$y_{it} = \beta_0 + \sum_{k=1}^{10} \beta_1^k (WORK_i \times [t = k]) + \sum_{k=1}^{10} \beta_2^k (COND_i \times [t = k]) + \delta y_{i,t=0} + \varepsilon_{it} \tag{10}$$

where $[t = k]$ is a dummy indicator for the kth day of the intervention. The standard errors will be corrected for auto-correlation by clustering at the individual level. As before $y_{i,t=0}$ is the measure of the outcome variable at baseline, prior to the start of the intervention, and is included as a control to improve precision. As before, we will also pool across "wait" conditions to improve statstical power.

3.5 Outcome Variables

Below we list the outcome variables and indices which we will consider by group. Outcomes marked by an asterisk (*) were measured daily.

1. Daily Wellbeing (PANAS)

    (a) Happy*

    (b) Excited*

    (c) Upset*

    (d) Ashamed*

    (e) Proud*

    (f) Alert*

    (g) Index: PCA/weighted standardized average of variables (a)-(f)

2. Attrition

    (a) Total number of days absent

    (b) Daily absence

3. Endline Wellbeing

    (a) Self-reported happiness

    (b) Comparative happiness

(c) Self-reported decision-making power in household

(d) Self-reported self-esteem

(e) Happiness regarding voucher purchases

(f) Mental Health Status

(g) Optimism about life prospects

(h) Women only: Life security and hardship

(i) Coping and Self-Efficacy

(j) Index: PCA/weighted standardized average of variables (a)-(g)

4. Consumption

(a) Amount spent on predetermined categories (we will look into specific types of products in the event we find significance in the following broader categories):

    i. Amount spent on "staples" vs. "superfluous/luxuries"

    ii. Amount spent on food vs. non-food items

    iii. Amount spent on healthy vs. unhealthy food items

(b) Endline Survey: self-reported spending at PBK

(c) Index: PCA/weighted standardized average of variables (b)-(e)

5. Voucher saving

(a) Number of vouchers for a given purchase (Vouchers per Receipt)

(b) Average number of items purchased per 100 KSH

References