KickStart Cohort Study

Pre Analysis Plan

Busara Center for Behavioural Economics*

Abstract

This document describes the analysis plan for the impact evaluation assessing the effect of Kickstart International's irrigation pumps on farmer welfare in Kenya. The evaluation of the cohort study consists of a panel analysis of small scale farming households across 35 districts that bought the pump in 2009, 2011 and 2015. Baseline and midline data collection of the cohorts that bought pumps in 2009 and 2011 was carried out by IFPRI, while the 2015 endline was conducted by the Busara Center for Behavioral Economics.

The present document outlines the evaluation design, outcome variables and econometric methods Busara will use to assess the effect of the pump on income and assets, land management practices, food security and consumption, time allocation, as well as intra-household decision-making.

Keywords: irrigation pumps, smallholder farmers, impact evaluation.

*Jonathan Court, Ngong Road, Nairobi, Kenya. `www.busaracenter.org`

1 Introduction and Intervention

KickStart's mission is to lift millions of people out of poverty quickly, cost-effectively and sustainably. In pursuit of this mission, KickStart designs, promotes and mass markets simple moneymaking tools. In particular, KickStart has developed and distributed several models of human-powered irrigation pumps since 1998. KickStart's flagship products are the MoneyMaker Max and MoneyMaker Hip pumps, which allow farmers to pull and pressurize water, irrigate up to two acres of land and grow crops year-round.

While KickStart has conducted internal evaluations of the impact of their pumps from the beginning, funding for an external independent evaluation has been secured, including from 3iE. To this effect, we will conduct a panel analysis of small scale farming households across 35 districts that bought the pump in 2009, 2011 and 2015. Baseline and midline data collection of the cohorts that bought pumps in 2009 and 2011 was carried out by IFPRI, while the 2015 endline is currently being conducted by the Busara Center for Behavioral Economics. Final analysis of all data will be conducted by Busara as well.

The study aims to reliably establish the impact of the pump on the welfare of small-scale farmer households in Kenya. While intuitive direct effects such changes in land management practices, assets and income as are the focus, secondary effects such as food security and consumption, education, health and time allocation are also considered. As a large number of small-scale farmers tend to be women, the study further seeks to understand the pump's ability to shift meaningful dimensions of women's empowerment in terms of intra-household decision-making, female psychological wellbeing and levels of intimate partner violence. The long-term purpose of the evaluation is to contribute to knowledge about the role of small-scale irrigation technology in rural development.

2 Evaluation Questions

Our main questions are: (i) What is the overall effect of owning a pump vs. not owning a pump on various dimensions of household welfare? (ii) What are the long-term effects of owning a pump on household welfare? The study therefore compares three cohorts of farmers each that have purchased the pump at different time points – in 2009, 2011 and 2015. While this arm is not randomized and assumes the farmers across all three batches are similar to one another, the farmers who bought the pump in 2015 serve as quasi-control group in this design. This arm will allow for comparing the long-term effects of the pump on household welfare through econometric techniques such as propensity score matching.

3 Evaluation Design

3.1 Sampling and Identification Strategy

The Cohort arm follows three cohorts of farmers who have purchased Moneymaker pumps. Cohort 1 is made up of farmers who purchased pumps in 2009, Cohort 2 purchased pumps in 2011 and Cohort 3 purchased pumps in 2014/15. KickStart and IFPRI, who kicked off the cohort impact evaluation before Busara came on board, interviewed a total of 1230 farmers across 35 Kenyan districts: 585 in 2009 (Cohort 1) and an additional 645 in 2011 (Cohort 2). Busara agreed to follow up with a subset of the total Cohort 1 and Cohort 2 sample in addition to interviewing Cohort 3 households. A random sample, stratified by district, out of the existing dataset was drawn to decide on which Cohort 1 and 2 households to follow up with. Busara further obtained incoming sales data from KickStart on a monthly basis, which was used to randomly select the required number of Cohort 3 households. To qualify for Cohort 3 inclusion from the sales list, a household had to have bought a pump no longer than 6 months before the interview. Cohort 3 sample size per district was determined in proportion to the combined Cohort 1 and 2 sample in each location. To control for seasonal effects, the interviews of all three cohorts in each district were completed within a range of maximum two weeks.

3.2 Data Collection Methods and Instruments

During the 2015 endline that Busara was responsible for, trained interviewers visited the households and both the primary male and the primary female of the household were interviewed (separately). Surveys were administered on tablets using the SurveyCTO survey software. To ensure data quality, Busara performed backchecks consisting of 10% of the survey, with a focus on non-changing information, on 10% of all interviews. This procedure was known to field officers ex ante.

3.3 Risk and Treatment of Attrition

Attrition was a concern in the cohort arm, since some of the tracking information, especially of the Cohort 1, was not accurate anymore and did not involve much detail. To control for attrition, Busara used three approaches. First, the data collection team used all pieces of information available to find respondents for follow-up surveys: phone number, GPS co-ordinates and the farmer's name. The first method Busara used is to contact farmers a week prior to the field team visiting a region to confirm their location and arrange a convenient time for the survey to be administered. The standard protocol when there is difficulty contacting farmers was to call three times a day for the week prior to the field team visiting. In the event that it is impossible to make contact by phone, or Busara does not have a phone number, GPS coordinates were used to locate therespondent's residence. When using both methods, phone or GPS, the field team confirmed this information by asking for the respondent by name in the nearest village. Farmers that are tracable but not available to participate in the survey at the time the field team is visiting their district are revisited at another time. Secondly, survey completion was incentivized through a small appreciation gift (spare pump parts, 2 kg of maize flour). Finally, attrition is controlled for econometrically in the analysis.

	Reschedule	Refused	Untraceable	Mteja	Relocated	Other
Cohort 1:	58	33	80	15	22	23
Cohort 2:	48	32	98	15	22	16
Cohort 3:	76	41	46	52	23	42
All Cohorts:	182	106	224	82	67	81
	Target	Completed	Success Rate			
Cohort 1	432	306	0.708			
Cohort 2	426	352	0.826			
Cohort 3	439	351	0.780			
Total:	1297	1009	0.778			

4 Econometric Specifications and Outcomes

4.1 Basic Specification

For the cohort study, the specifications used to identify differences between the cohort groups are

$$y_{\text{ivE}} = \beta_0 + \alpha_v + \beta_1 C1 + \beta_2 C2 + \epsilon_{iv} \tag{1}$$
$$y_{\text{ivE}} = \beta_0 + \alpha_v + \beta_1 C12 + \epsilon_{ive} \tag{2}$$
$$y_{\text{ivE}} = \beta_0 + \alpha_v + \beta_1 C12 + \delta y_{ivB} + X_{iv}\epsilon_{ive} \tag{3}$$

where *C1* is a dummy for whether a household belongs to Cohort 1, *C2* a dummy for whether a household belong to Cohort 2, and *C12* a pooled dummy for the two older cohorts. Here, y_{ivB} represents time-varying controls (i.e a vector of control variables taken in the taken from the baseline for each cohort, at $t = B$) and X_iv is a vector of time-invariant controls taken from the data collected by Busara.

4.1.1 Propensity Matching

We will supplement the regression results from the specifications above with a propensity matching approach. In using the propensity matching approach, we make the assumption that – conditional on observed characteristics – treatment assignment is independent of potential outcomes, and the comparison of treated and untreated groups can be interpreted as the treatment effect with no selection bias. The potential outcomes are also then independent of treatment status conditional on the propensity score $p(X_i)$, where the propensity score is the probability of being assigned to treatment conditional on observables: $p(X_i) = E\left[T_i|X_i\right]$.

To implement this, we then match treated individuals with non-treated individuals based on their propensity scores, to create comparable groups. This is implemented in STATA using the `teffects` command. See Abadie and Imbens 2009 for details.

4.2 Accounting for Multiple Inference

As household water pumps are likely to impact a large number of economic behaviors and dimensions of welfare, and given that our survey instrument often included several questions related to a single behavior or dimension, we will take account of the issues of multiple inference with a large number of outcomes well documented in the literature (see Romano and Wolf 2005). We will account for multiple hypotheses by using outcome variable indices and family-wise p-value adjustment.

We have catalogued below our primary outcomes, and the primary groups of outcomes that we intend to consider in the analysis outlined above. For each of these outcome groups, Busara will construct indices (where possible) and for each of the components of these indices, and will report both unadjusted p-values as well as p-values corrected for multiple comparisons using the Family-Wise Error Rate.

4.2.1 Construction of Indices

To keep the number of outcome variables low, allowing for greater statistical power even after adjusting p-values to control for multiple inference, we will construct indices for several of our groups of outcome variables. To this end, we will use the following procedure outlined in Anderson (2008). First, for each outcome variable y_{jk}, where j indexes the outcome group and k indexes variables within outcome groups, we re-code the variable such that high values correspond to positive outcomes. We then compute the covariance matrix Σ_j for outcomes in outcome group j, which consists of elements:

$$\hat{\Sigma}_{jmn} = \sum_{i=1}^{N_{jmn}} \frac{y_{ijm} - \bar{y}_{jm}}{\sigma^y_{jm}} \frac{y_{ijn} - \bar{y}_{jn}}{\sigma^y_{jn}} \tag{4}$$

Here, N_{jmn} is the number of non-missing observations for outcomes m and n in group j, $\bar{y}_{jm}$ and $\bar{y}_{jn}$ are the means for outcomes m and n in outcome group j,and $\sigma^y_j m$ and $\sigma^j_j n$ are the standard deviations in the control group for the same outcomes.

We then invert the covariance matrix, defining weight w_{jk} for each outcome k in outcome group j by summing the entries in the row of the inverted covariance matrix corresponding to that outcome:

$$\hat{\boldsymbol{\Sigma}}_j^{-1} = \begin{bmatrix} c_{j11} & c_{j12} & \cdots & c_{j1K} \\ c_{j21} & c_{j22} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots \\ c_{jK1} & \vdots & \ddots & c_{jKK} \end{bmatrix} \tag{5}$$

$$w_{jk} = \sum_{l=1}^{K_j} c_{jkl} \tag{6}$$

Here, K_j is the total number of outcome variables in outcome group j. Finally, we transform each of the outcome variables by subtracting its mean and dividing by the control group standard deviation, and then weighting it with the weights obtained using the method above. We denote the resulting transformed variable as $\hat{y}_{ij}$ it yields a generalized least squares estimator (Anderson 2008).

$$\hat{y}_{ij} = \left(\sum_{k \in \mathbb{K}_{ij}} \right)^{-1} \sum_{k \in \mathbb{K}_{ij}} w_{jk} \frac{y_{ijk} - \bar{y}_{jk}}{\sigma^y_{jk}} \tag{7}$$

Here, $\mathbb{K}_{ij}$ denotes the set of non-missing outcomes for observation i in outcome group j.

4.2.2 Family-Wise Error Rate

Because combining individual outcome variables in indices as described above still leaves us with multiple outcome variables (viz. separate index variables for health, education, etc.), we additionally adjust the p-values of our coefficients of interest for multiple statistical inference. These coefficients are those on the treatment dummies in the basic specifications, or those on the dummies for individual treatment arms. The procedure for this adjustment, from Anderson 2008, is as follows:

1. We compute naïve p-values for all index variables $\hat{y}_j$ of our j main outcome groups, and sort these in order of decreasing significance, i.e in order of increasing p-values such that $p_1 < p_2 < \cdots < p_J$.

2. We follow Anderson's (2008) variant of Efron & Tibshirani's (1993) non-parametric permutation test for each of the indices representing our main outcome groups. This permutation test is used in place of the standard t test, allowing us to calculate p-values that do not rely on assumptions about the distribution of the test statistic. This resampling involves random draws of treatment assignment, in order to sample the data under the assumption of no treatment effect. We then estimate the simulated p-value for difference in means for treatment and control.

3. We then impose the original monotonicity (from ordering in step 1) in the resulting vector of p-values $[p^*_1, p^*_2, \cdots, p^*_J]'$ by computing $p^{**}_r = \min\{p^*_r, p^*_{r+1}, \cdots, p^*_J\}$,

where r is the position of the outcome in the vector of naïve p-values.

4. We then repeat steps 2-4 of the procedure 100,000 times and compute the fraction of iterations where the simulated p-value is lower than our observed p-value and define this as our non-parametric p-value, p_r^{fwer*}.

5. We enforce monotonicity again: $p_r^{fwer} = \min\{p_r^{fwer*}, p_{r+1}^{fwer*}, \cdots, p_J^{fwer*}\}$, to ensure that the largest unadjusted p-values correspond to the largest adjusted p-values.

This yields the final vector of family-wise error-rate corrected p-values. We will report both these p-values and the naïve p-values. Within outcome groups, we report naïve p-values for individual outcome variables other than the indices.

4.3 Heterogeneous Effects

We will examine several potential sources of heterogeneity in the effect of pumps. Most of these sources of heterogeneity will be time invariant characteristics from the data collected by Busara, some are time-varying sources of heterogeneity taken from the IFPRI dataset (which we take as roughly being a baseline for our study) or secondary data.

Time-invariant:

- Gender (of person that bought the pump)
- Infrastructure & Market Access (determined by survey questions on distance to markets, roads, etc.)
- Education levels of adult household members
- Source of water

From baseline:

- Asset and consumption indices
- Primary activity of HH members
- Number of plots farmed & total area farmed
- Plot quality (fertility & slope)

Secondary data:

- Regional rainfall

The specifications we use to identify heterogeneous treatment effects in the cohort arm are:

$$
\begin{aligned}
y_{ivE} &= \beta_0 + \beta_1 C1_{iv} + \beta_2 C2_{iv} + \beta_3 X_{iv} + \beta_4 C1_{iv} \cdot X_{iv} + \beta_4 C2_{iv} \cdot X_{iv} + \epsilon_{iv} \quad (8)\\
y_{ivE} &= \beta_0 + \beta_1 C12_{iv} + \beta_2 X_{iv} + \beta_3 C12_{iv} \cdot X_{iv} + \epsilon_{iv} \quad (9)\\
y_{ivE} &= \beta_0 + \beta_1 C12_{iv} + \beta_2 X_{iv} + \beta_3 C12_{iv} \cdot X_{iv} + \delta y_{ivB} + \epsilon_{iv} \quad (10)
\end{aligned}
$$

where, again, X_{iv} is the dimension of heterogeneity.

4.3.1 Clustering Approach to Heterogeneous Effects

For some of the dimensions of heterogeneity listed above, such as gender, the groups for which we want to identify heterogeneous effects are obvious. However, for other dimensions of heterogeneity, such as asset ownership and consumption or the measures of psychological well-being and empowerment, it is less obvious which subgroups to use for the interaction effects specifications used above. For these dimensions of heterogeneity, we will instead use Latent Class Analysis (LCA). The intuition behind this analytical approach is that there are underlying clusters of similar individuals, but that we observe this indirectly through a number of discrete variables. If latent classes are present, the covariation we see in observed variables is driven by this latent variable. We use the covariation in the variables we do observe to predict latent class membership. (McCutcheon 1987)

Consider data grouped into K latent classes/clusters that we want to identify from L categorical/discrete questions, each with C_l categories and observed vlues Y_l, where $l \in [1, L]$, and that the variable X, with $x \in [1, K]$, is the variable representing membership in one of the K latent clusters. We can write the probability of observing the vector of responses $\vec{y}$ as

$$P(\vec{Y} = \vec{y}) = \sum_{x=1}^{K} P(X = x) \cdot P(\vec{Y} = \vec{y} | X = x)$$

which is simply the weighted sum, over clusters from 1 to K, of the probability of observing $\vec{Y} = \vec{y}$ given membership in cluster x. If we make a local independence assumption, the joint probability of observing a vector $\vec{Y} = \vec{y}$ given $X = x$ is simply the product of observing each $Y_l = y_l$ given $X = x$:

$$P(\vec{Y} = \vec{y}) = \sum_{x=1}^{K} \left(P(X = x) \cdot \prod_{l=1}^{L} P(Y_l = y_l | X = x) \right)$$

From this, we can use Bayes' rule:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

to determine the probability of an observation being in a particular cluster given the observed $\vec{Y}$:

$$P(X = x|\vec{Y} = \vec{y}) = \frac{P(\vec{Y} = \vec{y}|X = x) \cdot P(X = x)}{P(\vec{Y} = \vec{y})}$$

We then assign individuals to the cluster with the highest probability of membership. (McCutcheon 1987; Jeroen K. Vermunt and Jay Magidson and Vermunt 2002)

To determine the optimal number of clusters to use, we will use the Krzanowski and Lai stopping rule. The intuition behind this stopping rule is to define an objective function describing what makes a "good" partition and then choose the number of clusters that optimises this critereon. The stopping rule proposed by Krzanowski and Lai uses a critereon based on the sum of squares $S = \text{trace}(W)$, where W is the the pooled within-group covariance matrix for a partition of the sample. This critereon uses the fact that, if there are k latent classes, S will decrease quickly as we increase the number of clusters up to k, and decrease slow down after this. (Krzanowski and Lai 1988)

To estimate the cluster heterogeneous effects, we will be using the specifications from Section 4.3, with cluster membership dummies as X_{iv}, the dimension of heterogeneity.

With clustering it is important to choose variables - based on economic theory - on which we will cluster prior to the analysis, to prevent subjective 'data mining' during analysis. We will therefore cluster over our major categories, for which we have created indices, to create clusters of overall socioeconomic status. We will therefore cluster over, education, baseline asset and consumption indices and baseline land usage.

5 Outcome Variables

Our primary outcome variables will be the following. Construction of outcome variables in more detail below.

1. Time spent working on agriculture by household and separately for female household members and children
2. Farm income and total income
3. Total consumption
4. Land management practices (investments, labour inputs and crops grown)
5. Food Security index for household and for children only
6. Psychological health: PPS, Scheier-Rosenberg and CES-D
7. Female empowerment and violence against women
8. Value of household assets

6 Covariates

Time-invariant from Busara data:

- District/Location fixed effects
- Respondent gender
- Household size
- Education and gender of HH head
- Infrastructure access index

Baseline:

- Asset and consumption index
- Primary activity of HH members
- Fertilizer use
- Crop marketing decisions
- Number of plots farmed & total area farmed
- Plot quality (fertility & slope)

6.1 Construction of Variables

6.1.1 Outcomes

I. Asset Dummies

(i) House ownership, size and materials

(ii) Pump

(iii) Hose pipe

(iv) Ox-Ploughs

(v) Oxen/work bulls

(vi) Knapsack sprayers

(vii) Refrigerator/cooler

(viii) Motor vehicle/pickup truck

(ix) Bicycle

(x) Motor cycle

(xi) Wheelbarrows

(xii) Ox-carts/donkey carts

(xiii) Hand carts

(xiv) Zero grazing unit

(xv) Boreholes/wells

(xvi) Storage structures (specify)

(xvii) Fishing equipment (boats, canoes, etc)

(xviii) Fish pond

(xix) Sewing machine

(xx) Cellular phone

(xxi) TV

(xxii) Radio

(xxiii) Other electronic equipment

(xxiv) Satellite dishes

(xxv) Electric generator

(xxvi) Solar panel

(xxvii) Car battery

(xxviii) Sofas

(xxix) Chairs

(xxx) Table

(xxxi) Clock / Watch

(xxxii) Beds

(xxxiii) Mattresses

(xxxiv) Cupboard

(xxxv) Kerosene Stove

(xxxvi) Land / Plot

(xxxvii) Greenhouse

(xxxviii) Water tank

(xxxix) Electricity

(xl) Posho Mill

(xli) Number of livestock owned

(xlii) Number of plots owned

(xliii) Loans and savings

Index: Total value of assets

II. Time use

Time allocation of household members (men, women, boys, girls) during rainy and dry season to:

(i) Leisure time
(ii) Non-leisure social activities (e.g. faith-based activities)
(iii) Rainfed crops
(iv) Irrigated crops
(v) Livestock production
(vi) Non-farm economic activities
(vii) School activities
(viii) Household chores

Index: Household - Time working in agriculture (all members)
Children – Time working in agriculture (non-adult members)
Females – Time working in agriculture (female members)

III. Land management practices

For each planting season during the dry season, short and long rains:

(i) Investment (fertilizer, labor inputs)
(ii) Output (crops grown, harvest volume)

IV. Income

(i) Income from livestock
(ii) Income from harvest
(iii) Income from other productive activities
(iv) Casual labor/ wage jobs
(v) Forest and agro-forestry products
(vi) Fisheries
(vii) Entrepreneurial activities
(viii) Income from other sources (remittances, rental income, pension, charitable organizations)

Index: Total income

V. Consumption

  (i) Consumption of livestock
  (ii) Consumption of harvest
  (iii) Consumption of drinking water
  (iv) Consumption of food groups
  (v) Cereals
  (vi) Bread and pasta
  (vii) Roots and tubers
  (viii) Vegetables
  (ix) Meat
  (x) Fish
  (xi) Dairy products
  (xii) Oils and fats
  (xiii) Fruits
  (xiv) Sweets
  (xv) Beverages
  (xvi) Alcoholic beverages
  (xvii) Other household-related expenditure
  (xviii) Expenditure on health-related medicines and services
  (xix) Expenditure on preventative healthcare services

Index: Total consumption

VI. Food security

  (i) Meals skipped (adults)
  (ii) Whole days without food (adults)
  (iii) Meals skipped (children)
  (iv) Whole days without food (children)

(v) Less preferred foods

(vi) Purchase of food on credit

Index: Children - Weighted standardized average of variables iii), iv)
Household – Weighted standardized average of variables i)-vi)

VII. Psychological health (only asked to females)

(i) Locus of control (PPS)

(ii) Optimism and self-esteem (Schleier-Rosenberg)

(iii) Depression (CESD)

VIII. baseline consumption

IX. baseline assets **Index**: Weighted standardized average of variables ii)-iii)

X. Female empowerment (only asked to females)

(i) Actual female decision-making

(ii) Potential female decision-making

(iii) Physical violence dummy

(iv) Sexual violence dummy

(v) Emotional violence dummy

(vi) Justifiability of violence score

Index: Violence - Weighted standardized average of variables iii)-v)
Empowerment – Weighted standardized average of variables i),ii), v), vi)

References

Abadie, Alberto and Guido W. Imbens (2009). "Matching on the Estimated Propensity Score". In: *National Bureau of Economic Research Working Paper Series* No. 15301. DOI: 10.3386/w15301. URL: http://www.nber.org/papers/w15301.

Anderson, Michael L. (2008). "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects". In: *Journal of the American Statistical Association* 103.484, pp. 1481–1495. ISSN: 0162-1459. DOI: 10.1198/016214508000000841.

Jeroen K. Vermunt and Jay Magidson and Jeroen K. and Jay Magidson. Vermunt (2002). "Latent Class Cluster Analysis". In: *Applied Latent Cluster Analysis*. Cambridge University Press. URL: http://dx.doi.org/10.1017/CBO9780511499531.004.

Krzanowski, W. J. and Y. T. Lai (1988). "A Criterion for Determining the Number of Groups in a Data Set Using Sum-of-Squares Clustering". In: *Biometrics* 44.1, pp. 23–34. ISSN: 0006341X, 15410420. DOI: 10.2307/2531893. URL: http://www.jstor.org/stable/2531893.

McCutcheon, A.L. (1987). *Latent Class Analysis*. Latent Class Analysis. SAGE Publications. ISBN: 978-0-8039-2752-0.

Romano, Joseph P. and Michael Wolf (2005). "Stepwise Multiple Testing as Formalized Data Snooping". In: *Econometrica* 73.4, pp. 1237–1282. ISSN: 0012-9682. DOI: 10.1111/j.1468-0262.2005.00615.x.