Pre-Analysis Plan

Consumer Responses to Food Safety Risk Information

Vivian Hoffmann
Sarah Kariuki
Mike Murphy

Date: August 24th, 2022
IRB Approval Kenya: AMREF #P1192/2022
IRB Approval US: IFPRI #MTID-22-0421

1 Background

The objective of the study is to test whether providing information on a known food safety hazard causes consumers to update their beliefs about the relative risks of different food products, and whether this belief updating leads them to change their purchasing behaviour.

1.1 Intervention

The intervention is provision of information on the relative risk of aflatoxin contamination in different types of maize flour. All participating individuals will take part in a baseline household survey, and a follow-up surveys approximately one month and 3-4 months later. Following the baseline survey, treated individuals will be read a short script by the enumerator, and presented with a poster summarizing some of the information they have received. This will include a recommendation to purchase formally milled flour rather than informally milled ("posho") flour to reduce their risk of aflatoxin exposure. A subset of treated individuals will receive additional information on the rate of aflatoxin contamination in posho flour. Respondents will then be given the poster which they may choose to display in their home. The goal of the research is to test whether treated households have different risk perceptions than control households at follow-up and whether they are more likely to have recently purchased formally milled flour.

1.2 Timeline

Baseline interviews (and information treatments where applicable) are intended to be carried out in September-October, 2022. Follow-up visits will be conducted approximately 1 month later, and again, conditional on observing impact during the first follow-up, in January 2023.

2 Design

2.1 Study population

The population of interest is low-income urban consumers who purchase informally milled maize flour. Low-income people are major consumers of informally processed foods, and also most likely to face difficulties accessing healthcare. Since the effects of aflatoxin contamination are most pernicious in early childhood, we target our sample to households with children aged five or younger.

2.2 Sample selection

Our study area will comprise households from (up to) three informal settlement areas in urban Nairobi: Kangemi, Kawangware, and Kibera, and (up to) 2 communities in Machakos county (Athi River and Machakos Town). Our sampling frame is taken from lists of parents of young children maintained by community health volunteers (CHVs) in each neighbourhood (unit of local administration). To be eligible for selection, households must have at least one child aged 5 years or younger at the time of sampling and have purchased posho flour (or purchased maize and had it ground at a posho mill) within the last 14 days.

Eligible households within each area will be randomly assigned to treatment as described below. We will attempt to interview all eligible households within an area. Households will be contacted up to three times in person or by phone to arrange a visit. After completing interviewing eligible households in a settlement area, we repeat the process in the next area, until reaching a sample size of approximately 1500 households. We will only include areas from Machakos county if the sample is not sufficiently large after completing all interviews in the three areas in urban Nairobi.

2.3 Treatment assignment

Treatment status will be randomly assigned at the household level, after stratifying by CHV (CHVs are allocated to particular areas within each neighborhood in settlements) prior to the baseline interview.

2.4 Power

In Table 1 below, we calculate the minimum detectable effect (MDE) size for our primary outcome of interest – purchase of formally milled flour – for a range of baseline values in the control group.

Table 1: Minimum Detectable Effect Size

Baseline Share (Control Households)	***Power level***			
	Control vs. Pooled		***Between Groups***	
	0.8	***0.9***	***0.8***	***0.9***
1%	2.29%	2.74%	2.67%	3.28%
5%	3.97%	4.66%	4.60%	5.46%
10%	5.14%	6.00%	5.95%	6.99%
15%	5.93%	6.91%	6.86%	8.03%
20%	6.52%	7.57%	7.53%	8.79%
25%	6.95%	8.07%	8.03%	9.35%
30%	7.27%	8.42%	8.40%	9.75%
40%	7.61%	8.81%	8.79%	10.18%
50%	7.64%	8.82%	8.81%	10.18%

The first comparison ”Control vs. Pooled” calculates the MDE comparing the control proportion to the pooled treatment, while the second comparison ”Between Groups” calculates the MDE for the control vs. either of the two treatments individually. Both calculations are carried out for power levels of 0.8 & 0.9 respectively. The rows present values for different assumed proportions in the control group at baseline, ranging from 1% to 50%.

3 Analysis

3.1 Outcomes of interest

Our primary outcome of interest is the share of participants at endline who have formally milled flour for household consumption present in the household at endline. This will be measured by direct observation by enumerators following completion of the endline survey. This will be a binary indicator which takes the value '1' if formally milled maize flour (or its packaging material) is observed in the household at endline and '0' otherwise.

Our secondary outcomes of interest are the following:

1) Respondents' subjective probabilities of aflatoxin contamination in formally and informally milled flour respectively. As part of the baseline and endline surveys, respondents will be asked to represent the probability, using 100 beans, that a bag of [FLOUR TYPE] is affected by a food safety problem, for packaged and informally milled flour respectively.

2) Total reported monetary value of non-maize starches consumed per adult equivalent over the past 7 days.

3.2 Estimation

Since our primary outcome of interest is a binary variable, our first specification will estimate the following logistic regression:

$$Y_{i,t=1} = \beta_0 + \beta_1 InfoAny_{i,t} + \gamma_1 z_i' + \phi_i + \epsilon_i \quad (1)$$

Where $Y_{i,t=1}$ takes the value 1 if individual i has formally milled flour present in the household at endline, and 0 otherwise, and *InfoAny* is an indicator that takes the value 1 if the household was assigned to one of the two information treatments, and 0 otherwise. We include a vector of control variables, z', which will be selected via post-double-selection LASSO [2], as well as randomization strata (CHV) fixed effects, ϕ_i.

The following variables from the baseline survey will be included as candidates for selection as controls: outcome measure at baseline; types of maize reported consumed by the household during the past 7 days (binary variables); mean unit price of maize consumed during the past 7 days; total value of maize consumed during the past 7 days; total value of other starches consumed during the past 7 days; household size and composition (binary variables for presence of members in age groups: 0-6 months, 6 months-2 years, 2-5 years, 5-12 years, 13-18 years, over 50 years); age of respondent; gender of respondent; education level of respondent (binary variables by category); marital status of respondent; monthly income; assets owned (binary variables per asset type, housing quality variables, index of these based on the method proposed by Anderson [1], implemented using the Stata command swindex [3]) value of regular expenditures, by type; respondent's relative risk preferences, as self-reported (0-10 scale) in the baseline survey; probability that each of posho and formally milled maize are affected by a food safety problem, difference in this probability across the two flour types; knowledge of aflatoxin (binary variables for each correct knowledge point). We will report the odds-ratio of the coefficients, as well as the average of the marginal effects.

We will also estimate a specification testing the impact of the two versions of the treatment separately:

$$Y_{i,t=1} = \beta_0 + \beta_1 InfoRelOnly_i + \beta_2 InfoRelPlus_i + \gamma_1 z_i' + \phi_i + \epsilon_i \quad (2)$$

Where *InfoRelOnly* is an indicator for being assigned to receive only the relative information (T1), and *InfoRelPlus* is an indicator for being assigned to receive, in addition to this, absolute information on the aflatoxin contamination rate in informally milled flour (T2). For secondary outcomes, we will estimate the same specifications using OLS.

To test for potential information spillovers, we will additionally estimate specifications (1) and (2) with a control for the proportion of study households within a set of radii (to be determined after completion of the baseline survey) assigned to either treatment group, $ShareInfoAny_{i,t}$, the interaction of this term with control treatment status, $I(InfoAny = 0)$, and an indicator for the absence of any study households within that radius $NoneInRadius_i$:

$$\begin{gathered} Y_{i,t=1} = \beta_0 + \beta_1 InfoAny_{i,t} + \beta_2 ShareInfoAny_{i,t} \\ + \beta_3 I(InfoAny = 0) * ShareInfoAny_{i,t} + ShareInfoAny_{i,t} \\ + \beta_4 NoneInRadius_i + \gamma_1 z_i' + \phi_i + \epsilon_i \quad (3) \end{gathered}$$

$$\begin{gathered} Y_{i,t=1} = \beta_0 + \beta_1 InfoRelOnly_{i,t} + \beta_2 InfoRelPlus_{i,t} + \beta_3 ShareInfoAny_{i,t} \\ + \beta_4 I(InfoAny = 0) * ShareInfoAny_{i,t} + \beta_5 NoneInRadius_i + \gamma_1 z_i' + \phi_i + \epsilon_i \quad (4) \end{gathered}$$

Lastly, we will conduct a heterogeneity analysis, to test whether there are differences in the effects of treatment on formal flour purchases, conditional on the respondent's belief on the level of aflatoxin present in informally milled maize at baseline. To do this, we will estimate specifications (1) and (2), (a) interacting the indicators for treatment status with a dummy variable that takes the value 1 if the respondent's baseline subjective probability of aflatoxin contamination in informally milled flour is at or below the median value in the sample, and zero otherwise:

$$Y_{i,t=1} = \beta_0 + \beta_1 InfoAny_{i,t} + \beta_2 BelowMedian_{i,t=0} + \beta_3 InfoAny * BelowMedian_{i,t=0} + \gamma_1 z_i' + \phi_i + \epsilon_i \quad (5)$$

$$\text{Y}_{i,t=1} = \beta_0 + \beta_1 InfoOnly_i + \beta_2 InfoSpec_i + \beta_3 BelowMedian_{i,t=0} + \beta_4 InfoOnly * BelowMedian_{i,t=0} + \beta_5 InfoSpec * BelowMedian_{i,t=0} + + \gamma_1 z_i' + \phi_i + \epsilon_i \quad (6)$$

and (b) interacting the indicators for treatment status with a dummy variable that takes the value 1 if the respondent believes that informally milled flour is either equally as risky or safer than formally milled flour at baseline:

$$Y_{i,t=1} = \beta_0 + \beta_1 InfoAny_{i,t} + \beta_2 InformalSafe_{i,t=0} + \beta_3 InfoAny * InformalSafe_{i,t=0} + \gamma_1 z_i' + \phi_i + \epsilon_i \quad (5)$$

$$\text{Y}_{i,t=1} = \beta_0 + \beta_1 InfoOnly_i + \beta_2 InfoSpec_i + \beta_3 InformalSafe_{i,t=0} + \beta_4 InfoOnly * InformalSafe_{i,t=0} + \beta_5 InfoSpec * InformalSafe_{i,t=0} + \beta_6 Y_{i,t=0} + \gamma_1 z_i' + \phi_i + \epsilon_i \quad (6)$$

For all of the specifications listed above, if there is more than 10% attrition in the sample between baseline and endline, or if there is statistically significant attrition across treatment groups we will estimate Lee bounds and report these as our primary results.

References

[1] Michael Anderson. "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 (2008), pp. 1481–1495. URL: https://EconPapers.repec.org/RePEc:bes:jnlasa:v:103:i:484:y:2008:p:1481-1495.

[2] Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. "Inference on Treatment Effects after Selection among High-Dimensional Controls†". In: *The Review of Economic Studies* 81.2 (Nov. 2013), pp. 608–650. ISSN: 0034-6527. DOI: 10.1093/restud/rdt044. eprint: https://academic.oup.com/restud/article-pdf/81/2/608/18394034/rdt044.pdf. URL: https://doi.org/10.1093/restud/rdt044.

[3] Benjamin Schwab et al. "Constructing a summary index using the standardized inverse-covariance weighted average of indicators". In: *The Stata Journal* 20.4 (2020), pp. 952–964. DOI: 10.1177/1536867X20976325. eprint: https://doi.org/10.1177/1536867X20976325. URL: https://doi.org/10.1177/1536867X20976325.