Official Title: Preventing Non-communicable Diseases in Guatemala Through Sugary Drink Reduction

NCT number: NCT04022694

Document name: Analysis Plan

Document date: April 30, 2020

Analysis Plan: Preventing non-communicable diseases in Guatemala through sugary drink reduction

Aims:

The primary aim of this study is to test the association between SSB warning/non-SSB promotion posters compared to control posters and 1) changes in mean fluid ounces of SSBs purchased and 2) changes in total beverage calories purchased by adolescents in school cafeterias in Guatemala City.

Hypotheses:

Compared to the counterfactual (what would have happened to purchased fluid ounces and total beverage calories if the SSB warning posters had not been implemented, taking into account pre-treatment differences),

1) SSB warning/non-SSB promotion posters are associated with greater reduction in mean fluid ounces of SSB purchased
2) SSB warning/non-SSB promotion posters are associated with greater reduction in total beverage calories purchased

Primary outcomes:

1. Fluid ounces of SSBs purchased per sales transaction
    a. SSBs= any beverage with more than 5g/100mL of sugar reported in the NFP
2. Total beverage calories purchased per sales transaction
    a. Beverage calories= any calories coming from beverages (SSBs and non-SSBs)

Secondary outcomes (from survey data):

1. Percent of all surveyed transactions that include an SSB
2. Percent of all surveyed beverage transactions that include an SSB
3. Total calories per surveyed transaction

Description of data:

The first is a longitudinal data set (at the school level) of store purchases from July – August 2019 obtained by trained research assistants in three schools located in Guatemala City. Data are daily purchases of food items and beverages by students. Sample is total purchases obtained from scanning products' barcodes or using an application (Square app) to assign unique codes to products without a barcode. Data include food purchases, brand, flavor, size, and quantity. Additionally, an inventory of beverages' brand, flavor, price, size (mL), serving size (mL) as reported in the Nutrition Facts Panel, calories (kcal), total fat (g), saturated fat (g), total carbohydrates (g), sugar (g), sodium (mg), protein (g), ingredients, total sugar, sodium, and protein per container (g).

The second dataset is an intercept survey conducted with students after they made a purchase at the school store. Data include total expenditure, number and type of items purchased (i.e., beverage, food, or non beverage/food item), perception of healthfulness of

specific beverages (i.e., water, fruit drink, soft drink, light soft drink), and the label awareness, recall, use, and trust. Usual SSB consumption, demographics, height, and weight are also included in the survey.

For our primary analysis, we will look only at the 2 schools which were randomly assigned to condition (one treatment and one control school).

Statistical analysis:

- Consistency of data: We will conduct analyses to check the consistency of categorical data, including frequency tables and chi-square tests (i.e., purchase type, beverage type, healthfulness perception, label awareness, and demographic variables). To check the consistency of continuous variables, we will create box plots of mean size by beverage categories (SSBs, non-SSBs), purchased quantity, price, volume purchased, serving size, calories, and expenditure.
- Outliers: Within each beverage category (SSBs, non-SSBs), we will examine purchases in the 99th percentile and above (by volume and total calories) to identify any beverage purchases that seem implausible. For the survey data, we will exclude any expenditures that appear implausible.
    - Data will be pooled (across conditions) when assessing for quality so that we are blind to which condition outliers are in. We will ask the school stores and research assistants to clarify any values that appear out of range prior to excluding any data. Any data that are excluded will be done prior to re-matching data points to their condition.
    - We will exclude any individual-level data from our customer purchase assessments if no school cafeteria purchase was recorded.
    - We will exclude any customer purchase assessments if they include implausible purchases (e.g., items not sold at the store). Although we expect to have minimal missing data, we will consider other analytic techniques to handle missing data if the amount is greater than 10%.
- Descriptive analysis: We will summarize (i.e., mean, SE, trends) the primary and secondary outcome variables by condition and time period (pre- or post-). We will also examine changes in primary and secondary variables.
- Statistical design: We will use a difference-in-differences approach to examine the effect of the posters taking into account pre-treatment differences between the treatment and control groups and time trends (pre- and post-). The basic equation for this approach is:

$$Y_{it} = \eta + \gamma d_i + \delta T_t + \alpha d_i * T_t + u_{it}$$

  Where d_i is a binary variable equal to 1 if they were exposed to the poster, 0 otherwise, T_t is a binary variable equal to 1 in the post period and 0 in the pre period, d_i*T_t is an interaction term between the two, and α is the difference-in-differences estimator.
- The effectiveness of the intervention will be tested by evaluating the null hypothesis that the coefficient of the interaction of time period and intervention is 0, indicating no change in the fluid ounces of SSBs purchased and beverage calories purchased between conditions.
- Regarding the secondary outcomes, we will present the percentage of all transactions and all beverage transactions that include an SSB by condition and time period. To

calculate the total calories per transaction, we will use the Nutrition Facts Panel data of each food or beverage product.