

This readme file was generated on [2023-02-28] by Bethany Simard

GENERAL INFORMATION

1. Title of Dataset: SSB_analysis.csv
2. Author Informaton
 - A. Principal Investigator Information
Name: Lynn Silver
Institution: Public Health Institute
Address: 2000 Center St, Suite 308, Berkeley, CA 94704
Email: lsilver@phi.org
 - B. Co-investigator Information
Name: Alisa Padon
Institution: Public Health Institute
Address: 2000 Center St, Suite 308, Berkeley, CA 94704
Email: apadon@phi.org
 - C. Co-investigator Information
Name: Libo Li
Institution: Alcohol Research Group
Address: 6001 Shellmound St #450, Emeryville, CA 94608
Email: lli@arg.org
 - D. Co-investigator Information
Name: Bethany Simard*
Institution: Public Health Institute
Address: 2000 Center St, Suite 308, Berkeley, CA 94704
Email: bsimard@phi.org
*Contact for data set questions
 - E. Co-investigator Information
Name: Tom Greenfield
Institution: Alcohol Research Group
Address: 6001 Shellmound St #450, Emeryville, CA 94608
Email: tgreenfield@arg.org
1. Date of data collection:
 - a. Baseline: 2017-12-21 to 2018-01-24
 - b. Follow-up 1: 2018-11-07 to 2019-04-16
 - c. Follow-up 2: 2019-11-16 to 2020-01-22
2. Geographic location of data collection:
 - a. San Jose, California: 95113, 95126, 95117, 95128, 95134, 95119, 95112, 95110, 95123, 95148, 95111, 95118, 95136, 95125
 - b. San Francisco, California; zip codes: 94115, 94102, 94103, 94110, 94124, 94134, 94112, 94128
3. Information about funding sources that supported the collection of the data:

This study was funded by the National Cancer Institute R01 CA230733-01. Additional funding was provided by the National Institute of Alcohol Abuse and Alcoholism P50 AA005595.

SHARING/ACCESS INFORMATION

1. Licenses/restrictions placed on the data: CC0 1.0 Universal (CC0 1.0), Public Domain Dedication
2. Links to publications that cite or use the data: <https://doi.org/10.1371/journal.pgph.0001219>
3. Recommended citation for this dataset:
Silver, LD, Padon, AA, Li, L, Simard, BJ, Greenfield, TK. (2023) Data set for: Changes in sugar-sweetened beverage consumption in the first two years (2018 – 2020) of San Francisco’s tax: A prospective longitudinal study, Dryad, Dataset.

DATA & FILE OVERVIEW

1. File List
 - a. `SSB_Sampling_and_Data_Collection_Methodology.pdf` – This document was written by our data collection partners ICF International and describes detailed methodology of data collection, including sample design, questionnaire content, data collection protocol, weighting, and response rates. There were 4 waves of data collection. The first follow-up (wave 2) was used as the baseline for our study and the second and third follow-up surveys (waves 3 and 4) were used as the 1st and 2nd follow-ups in our study. The weights calculation described in this document refers to the weights that would be used to analyze all 4 waves. Because we used the last 3 waves in our study, new baseline sample weights were constructed for the final sample and survival probabilities between our baseline (wave 2) and our first follow-up (wave 3) were estimated by logistic models. The baseline sample weights were then multiplied by the corresponding survival weights before post-stratification via raking to adjust the baseline sample weights and provide a closer match between the sample and the population across the post-strata defined by city, sex, race, age, and education groups.
 - b. `SSB_Analytic Sample Creation Flowchart.pdf` – This figure depicts how the analytic sample was created from 3 different sampling sources over 4 data collections.
 - c. `SSB_analysis.do` – This code conducts the final analyses in our study.
 - 0. Set up
 - 1a. Models - set up
 - 1b. Gamma Models, outcome = daily consumed taxed SSB (ounces)
 - 1c. Binomial Models, outcomes = dichotomous high versus low daily consumed taxed SSB (0/1)
 - 2. Prepare and export marginal estimates to Excel files
 - 3. Export model coefficients to RTFs
 - d. `SSB_analysis.csv` – This is the final processed data set for running the analyses in our study.
2. Additional related data collected that was not included in the current data package: There is unprocessed data from each wave of data collection not included in the package, including raw variables used to generate final constructs. Data on participants who were excluded from our analysis are also not included in the current package.

METHODOLOGICAL INFORMATION

1. Description of methods used for collection/generation of data:

- a. Our data collection partners ICF International collected four waves of data, the first wave via telephone, and the 2nd-fourth waves by web survey. The first follow-up was used as the baseline for our study and the second and third follow-up surveys were used as the 1st and 2nd follow-ups in our study. The study used three sampling techniques due to challenges with low response rates: telephone using random digit dialing techniques applied to sampling frames of landline and cell telephone exchanges, address-based sampling, and a non-probability web panel. Questionnaires were completed in English, Spanish, or Chinese.
2. Methods for processing the data: Data from each survey were appended together—each row/observation is unique to participant ID and wave. Variables for the study were constructed using Stata.
3. People involved with sample collection, processing, analysis and/or submission:
 - a. ICF International - sample collection
 - b. Libo Li - processing, analysis
 - c. Bethany Simard - processing, analysis

DATA-SPECIFIC INFORMATION FOR: SSB_analysis.csv

Number of variables: 20

Number of cases/rows: 4,130

Variable List:

Variable	Variable labels	Value definitions/labels	Analysis
masterid	Unique participant identification	Text	Main
analytic_sample7	Analytic sample indicator - meet criteria to be included in analysis	1= Yes, 0=No. Inclusion criteria included: 1) Took surveys B2 and F1 2) Didn't report daily beverage consumption in excess of 400 oz 3) Not missing information on covariates.	Main
repeatsB2F1OnlyF2_WPool_rk	Weights for final analytic sample	Continuous. To correct for respondent attrition and disproportionate sampling probabilities introduced by the sampling design, baseline sample weights were constructed for the final sample and survival probabilities over F1 were estimated by logistic models. The baseline sample weights were then multiplied by the corresponding survival weights before post-stratification via raking to adjust the baseline sample weights and provide a closer match between the sample and the population across the post-strata defined by city, sex, race, age, and education groups	Main
time01	Data collection wave - 1st follow-up	1=1st follow-up survey; 0=otherwise	Main
time02	Data collection wave - 2nd follow-up	1=2nd follow-up survey; 0=otherwise	Main
whatacity3_LF	City of residence	0=San Jose, 1=San Francisco	Main
ageb	Baseline age	Continuous	Main
raceNew_di	Race/ethnicity	1 = non-Hispanic (NH) white or NH Asian; 0=NH Black, Hispanic/Latinx, Other NH race	Main
eduG_di	Highest educational attainment at baseline	0=college degree or more; 1= less than a college degree	Main
sex	Sex	1=Male, 2=Female	Main

srcB2F1_LF_Hi_taxozgte4	Sampling source	Categorical variable: 2=Random digit dialing, 4= Address based, 6= Non-probability	Main
povertyfplR01_LF	Federal Poverty Level based on 2016 thresholds	0=Federal Poverty Level <200%, 1=Federal Poverty Level >= 200%	Main
taxoz	Average daily consumption of taxed sugar-sweetened beverages (ounces)	Continuous. Taxed beverages include: sweetened juice, regular soda, sweet tea, energy drinks, tea/coffee with added sugar.	Main
Hi_taxozgte4	Average daily consumption of taxed sugar-sweetened beverages >= 4 oz	1=Yes, 0 = No	Main
Hi_taxozgte6	Average daily consumption of taxed sugar-sweetened beverages >= 6 oz	1=Yes, 0 = No	Main
Hi_taxozgte8	Average daily consumption of taxed sugar-sweetened beverages >= 8 oz	1=Yes, 0 = No	Main
Hi_taxozgte85thB	Average daily consumption of taxed sugar-sweetened beverages >= 12 oz	1=Yes, 0=No (Note that the 85th percentile of consumed taxed sugar-sweetened beverages at baseline was 12 oz.)	Main
b2_pretax	In the final analytic sample who took baseline survey before the tax was implemented (1/1/2018)	1=Yes, 0=No	Sensitivity
dayssfNewgte16F1eqF2_r	In final analytic sample and spend few days (<16 days) in San Francisco in both follow-up surveys OR spend many days (>=16 days) in San Francisco in both follow-up surveys.	1=Yes, 0=No	Sensitivity
dayssfNewgte16_r_LF	Spent >= 16 days in San Francisco in past month	1=Yes, 0=No	Sensitivity

Missing data codes: None