

# **Investigating the efficacy of consumer interventions on sales of zero-calorie beverages**

Christine Chuong (20124798)

Giuliano Cruz (79247979)

Joey Hotz (18461400)

2024-03-29

## Introduction

Strong evidence has linked habitual intake of sugar-sweetened beverages (SSBs) with weight gain and a higher risk of type 2 diabetes mellitus, cardiovascular diseases, and even cancer [2]. Multiple policy interventions to inhibit consumption of SSBs have been proposed [1], including taxation of SSBs, limiting access to SSBs in schools and healthcare facilities, and adding informative package labelling. As policymakers create potential interventions to limit SSB consumption, there is a need for data-driven methods of determining their efficacy.

This study aims to evaluate two groups of interventions which intend to shift consumers towards purchasing zero-calorie beverages instead of SSBs. Five separate interventions implemented at three hospitals are assessed. Three of these interventions utilize calorie-related messaging and the other two use price discounts to incentivize purchasing zero-calorie beverages.

The primary statistical question is whether these interventions are associated with differences in the average daily proportion of zero-calorie and sweetened beverage sales. Additional questions of interest include whether these interventions are equally effective at each hospital, and whether the effect of multiple simultaneous interventions is greater than the “sum of its parts”.

## Data Description and Summaries

The data come from an interrupted time-series multi-site quasi-experimental study on sales of sweetened and zero-calorie beverages at three United States hospitals over a 30-week period. The data consists of 631 observations aggregated across the three sites. Each observation represents the daily sales of bottled beverages among multiple storefronts per site. Sales are automatically recorded by point of sale terminals in one site, and manually by store employees at remaining sites. Beverage sales are recorded every day from the beginning of the experiment until the final day, including weekends and holidays.

The data includes the following variables: the day of the experiment, the corresponding week-day, the site, and the ongoing intervention. The day variable is a positive integer. The latter three are categorical variables which respectively have seven, three, and nine levels. The data also contains the daily sales of bottled zero-calorie and sweetened beverages sold per site, and the total daily quantity of beverages sold per hospital. These are all non-negative integer values. Some observations also include the daily sales of 100% juice, orange juice, and sports drinks, which are also non-negative integer values. These quantities are only consistently recorded at *CHOP* and hardly recorded at the other hospitals.

Quantities of bottled zero-calorie and sugared beverages are missing in nine observations, each containing missing data for every sales-related column. Of these nine observations, seven occur in the final week of the study at *HF*, and two occur mid-study at *NS*. One mid-study

observation from *HF* does not report the total quantity of beverages sold, though the quantity of bottled zero-calorie and sugared beverage sales are provided.

## Exploratory Analysis

Side-by-side boxplots of the proportion of total sales coming from zero-calorie drinks for each site for each day of the week were created. They showed a similar proportion for each site and day of the week across sites, though one of the sites differed slightly and seemed to have more variable sales. Boxplots of total drinks sold showed one site had a lot more sales than the other two sites. Sales were also much lower for all sites on weekends compared to weekdays

A plot of the proportion of total sales coming from zero-calorie drinks over experiment day (Figure 1), shows the trends of sales across each site over day. There was some difference at the beginning of the experiment, but over time the average proportion seemed relatively stable, which is not expected. A scatterplot matrix comparing sugary, zero-calorie and total drink sales showed that they are all highly correlated with each other. The matrix also shows a positive linear relationship between zero-calorie, sugary and total sales.

A missing data plot (Figure 4) showed that orange juice, 100% juice and sports drinks had many missing observations. A missing data table (Figure 5) also showed many missing juice or sport drink sales, additionally showing that some sites did not record some of these sales at all. Since juice and sport drink sales are not of interest, and are inconsistently recorded, they are not included in the analysis. The missing data plot also showed nine observations had no sale measurements, which are also omitted.

## Formal Analysis

To answer the statistical questions, we formally assessed the relative effects of the interventions on the average daily proportion of both zero calorie and sugary bevarage sales. The intervention groups had a statistically significant effect on the daily sales of both zero calorie and sugary bevarage sales – likelihood-ratio test (LRT)  $p < 0.0001$  in both cases. Table 1 shows the pairwise group comparisons for zero calorie sales. The same comparisons for sugary bevarage sales are shown in Table 2. In all cases, relative effects are represented as relative percentage changes alongsie 95% confidence intervals (CI) and Holm-adjusted p-values (Wald tests). Compared with the baseline periods, it is clear that the combined “Discount and messaging’’ intervention was associated with an increase in the proportion of both zero calorie and sugary bevarage sales. The combined messaging strategy was also associated with an increase zero calorie beverage sales.

Table 1: Pairwise group comparisons on average proportion of zero calorie sales.

---

Comparison	Relative change (95% CI)	Adjusted p-value
Discount vs. Baseline	9.4% (-2.11% — 20.88%)	0.516
Discount and messaging vs. Baseline	38.6% (24.02% — 53.19%)	0.0001
Exercise messaging vs. Baseline	-7.8% (-17.61% — 2.01%)	0.516
Caloric content messaging vs. Baseline	-8% (-17.68% — 1.63%)	0.516
Combination messaging vs. Baseline	-13.2% (-22.27% — -4.03%)	0.028
Discount and messaging vs. Discount	26.7% (12.95% — 40.48%)	0.001
Combination messaging vs. Exercise messaging	-5.8% (-16.19% — 4.58%)	0.546
Combination messaging vs. Caloric content messaging	-5.6% (-15.85% — 4.71%)	0.546

Table 2: Pairwise group comparisons on average proportion sugary sales.

Comparison	Relative change (95% CI)	Adjusted p-value
Discount vs. Baseline	-4.2% (-13.24% — 4.93%)	1
Discount and messaging vs. Baseline	24.2% (12.39% — 36.02%)	0.00042
Exercise messaging vs. Baseline	5.5% (-4.69% — 15.63%)	1
Caloric content messaging vs. Baseline	0.4% (-9.12% — 9.88%)	1
Combination messaging vs. Baseline	-1.3% (-10.63% — 8.09%)	1
Discount and messaging vs. Discount	29.6% (16.87% — 42.31%)	0.0001
Combination messaging vs. Exercise messaging	-6.4% (-15.72% — 2.95%)	1
Combination messaging vs. Caloric content messaging	-1.6% (-11.29% — 8.01%)	1

Further, the intervention effects varied significantly across study sites, for both zero-calorie and sugary beverage sales (LRT  $p < 0.0001$  in both cases). The variability in proportion of daily sales from each type of beverage can be visualized in Figure 2 and Figure 3.

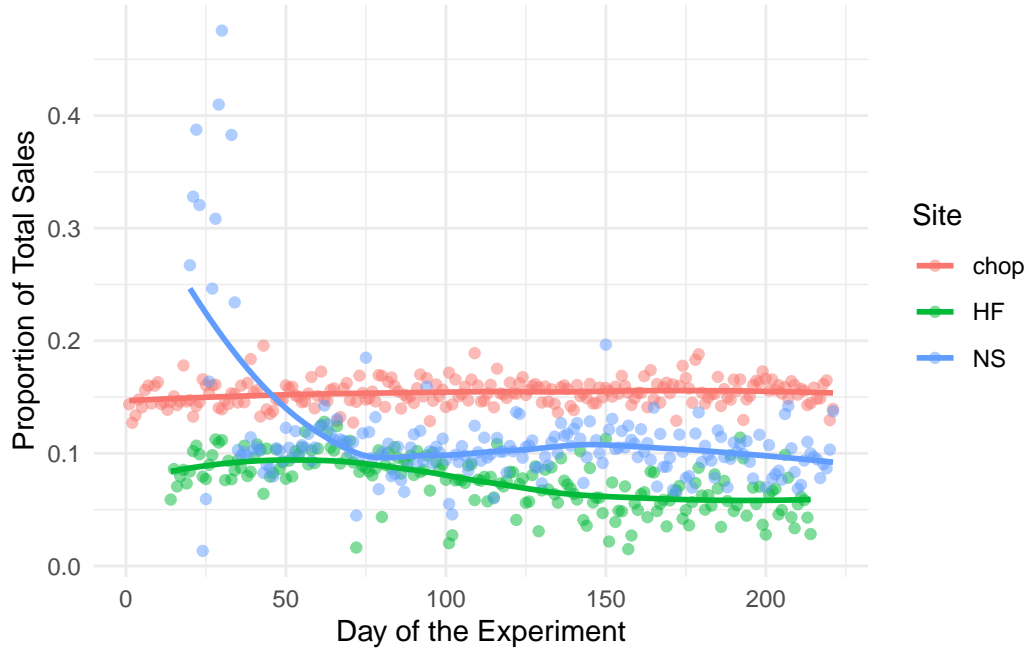


Figure 1: Proportion of zero-calorie beverages sold per day

## Conclusion

To accurately assess the efficacy of the interventions, it is best to create a statistical model which explains how these interventions affect the proportion of zero-calorie beverages. This response variable is less prone to variation than the overall sales of zero-calorie drinks, and it still provides an easily-interpretable metric for consumer purchasing behaviours.

Therefore, we recommend fitting a Poisson GLMM to the log-transformed sales to predict the proportion of zero-calorie beverage sales, and determining which effects are most relevant using the likelihood ratio test. However, the Poisson GLMM assumes that the underlying data follows a Poisson distribution. If this assumption is not met, we recommend utilizing robust Poisson regression, as the robust Poisson model does not require this distributional assumption. Both analyses employ the zero-calorie beverage sales as the primary outcome with the log of the sum of zero-calorie beverage sales and sugary beverage sales as an offset.

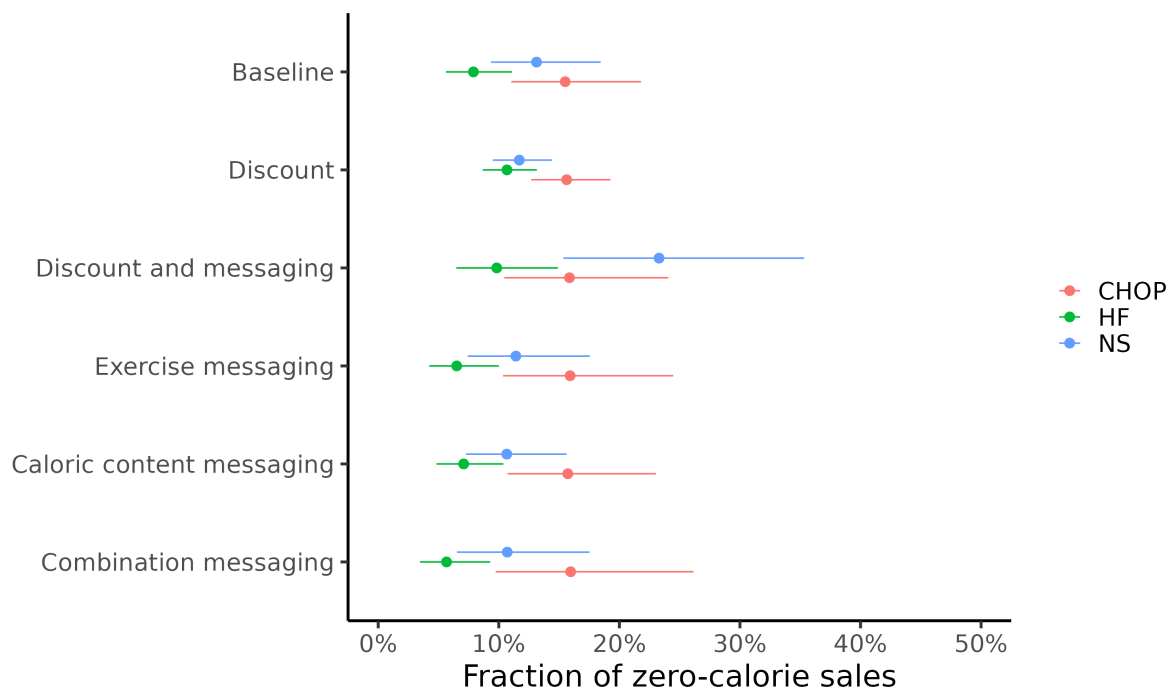


Figure 2: Expected daily proportion of zero-calorie beverage sales across study sites and intervention groups.

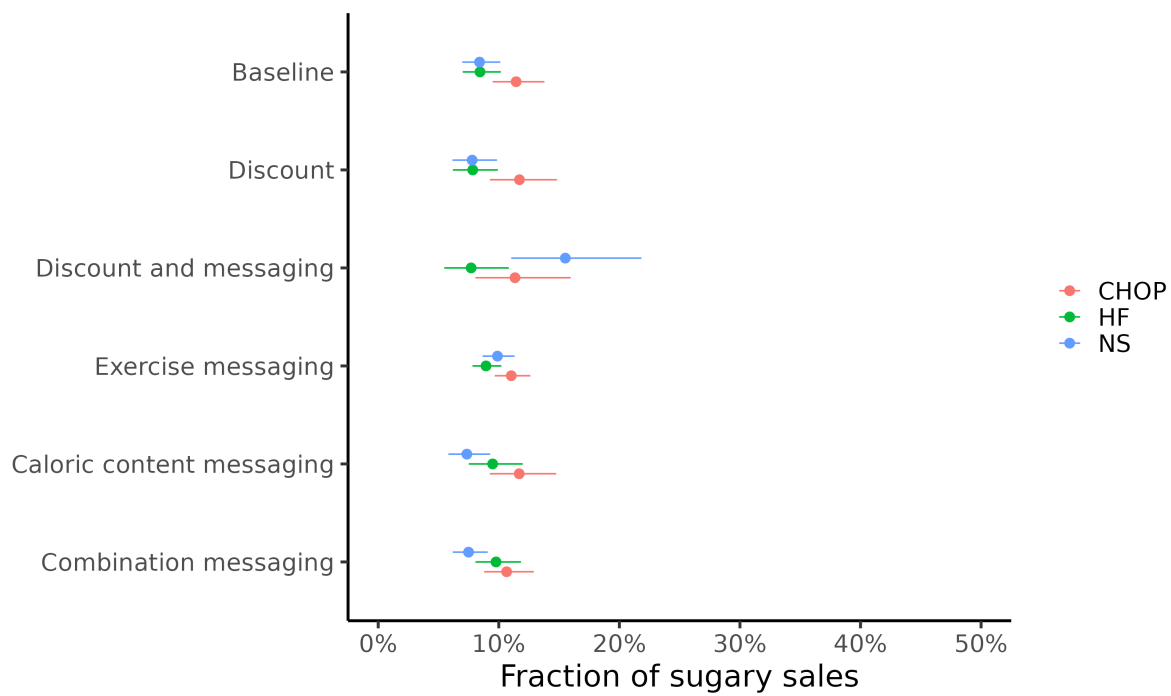


Figure 3: Expected daily proportion of zero-calorie beverage sales across study sites and intervention groups.

## References

- [1] James Krieger et al. “Sugar-Sweetened Beverage Reduction Policies: Progress and Promise”. In: *Annual Review of Public Health* 42.1 (2021). ISSN: 1545-2093. DOI: [10.1146/annurev-publhealth-090419-103005](https://doi.org/10.1146/annurev-publhealth-090419-103005). URL: <http://dx.doi.org/10.1146/annurev-publhealth-090419-103005>.
- [2] Vasanti S. Malik and Frank B. Hu. “The role of sugar-sweetened beverages in the global epidemics of obesity and chronic diseases”. In: *Nature Reviews Endocrinology* 18.4 (2022). ISSN: 1759-5037. DOI: [10.1038/s41574-021-00627-6](https://doi.org/10.1038/s41574-021-00627-6). URL: <http://dx.doi.org/10.1038/s41574-021-00627-6>.



## Statistical Appendix

### Missing Data

Below is the plot summarizing which observations have missing data and in which variables. A missing data table is also given, summarizing the number and percentage of missing observations for variables across sites. Site and variable combinations with no missing data points are omitted from the table

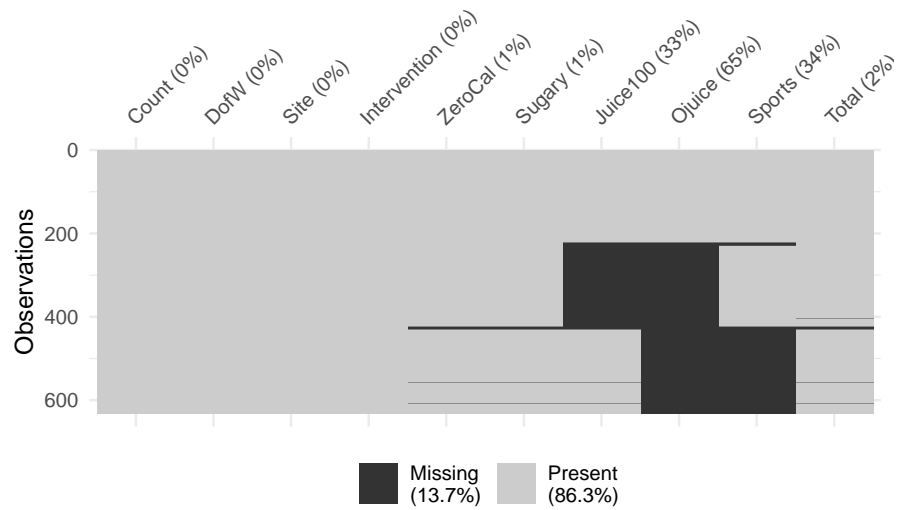


Figure 4: Missing data by variable and observation

Site	Type of Drink	Number of Missing Observations	Percentage Missing
HF	Juice100	208	100.00%
HF	Ojuice	208	100.00%
NS	Ojuice	202	100.00%
NS	Sports	202	100.00%
HF	Sports	15	7.21%
HF	Total	8	3.85%
HF	ZeroCal	7	3.37%
HF	Sugary	7	3.37%
NS	ZeroCal	2	0.99%
NS	Sugary	2	0.99%
NS	Juice100	2	0.99%
NS	Total	2	0.99%

Figure 5: Number and proportion of missing observations, grouped by site and by variable