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 weekday, 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 2) showed that orange juice, 100% juice and sports drinks had many missing observations. A missing data table (Figure 3) 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

In this study, an effective intervention will increase, relative to the baseline, the proportion of daily sales of zero-calorie beverages among sales that are either zero-calorie beverages or sugary beverages. This proportion increases when there are relatively more sales of zero-calorie beverages than of sugary beverages. This avoids the need to perform two separate analyses (one for zero-calorie and another one for sugary beverages) and any variation due to sales of other unrelated products.

To answer the statistical questions, we recommend using a generalized linear mixed effects model (GLMM) with a Poisson distribution. This type of model uses count data and allows for different effects across sites. It can also handle the longitudinal nature of the data. The response variable is the counts of zero-calorie beverage sales. Including the log-transformed total sales of zero-calorie or sugary beverages as an offset, the Poisson GLMM models the proportion of zero-calorie beverage sales as defined previously. The main independent variables
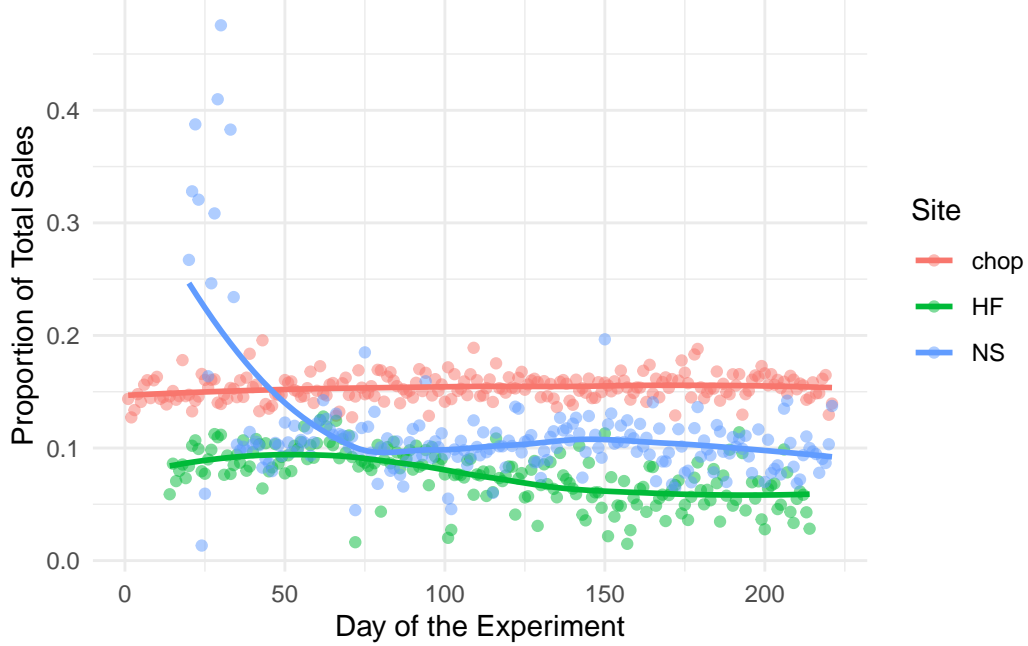
Figure 1: Proportion of total sugary and zero-calorie drink sales coming from zero-calorie drinks over time

of interest are intervention indicators, treating the baseline periods as the reference. Other covariates in the model include the site, the day of the week, and the study day.

To assess whether intervention effects differ by site, we recommend testing interaction effects between sites and intervention groups. This assessment can be implemented as a Likelihood Ratio Test. We recommend including random effects for intervention groups, which would allow for different effects between groups.

A major assumption of the GLMM with Poisson model is the correct specification of the Poisson distribution. To avoid this assumption, a robust version of the above model is recommended. In particular, the same model specification can be used, while employing Poisson estimating equations with a so-called sandwich covariance estimator. This model is known as robust Poisson as it remains valid even if the data do not support a Poisson distribution [3]. Once exponentiated, coefficients from the GLMM and the robust Poisson model have the same interpretation. For instance, if the exponentiated coefficient for a given intervention is equal to 1.5, then the proportion of zero-calorie beverage sales increases by 50% under the intervention, as compared to the baseline.

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.

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. 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. URL: http://dx.doi.org/10.1038/s41574-021-00627-6.

[3] Denis Talbot et al. "An Alternative Perspective on the Robust Poisson Method for Estimating Risk or Prevalence Ratios". In: *Epidemiology* 34.1 (Sept. 2022), pp. 1–7. ISSN: 1044-3983. DOI: 10.1097/ede.0000000000001544. URL: http://dx.doi.org/10.1097/EDE.0000000000001544.
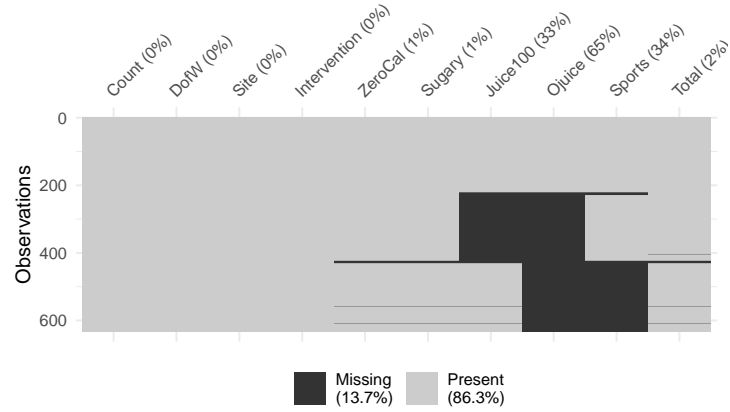
Statistical Appendix

Missing Data

Figure 2: Missing data by variable and observation

Site	Type of Sale	Number Missing	Percentage Missing
HF	Juice100	208	100
HF	Ojuice	208	100
NS	Ojuice	202	100
NS	Sports	202	100
HF	Sports	15	7.21
HF	Total	8	3.85
HF	ZeroCal	7	3.37
HF	Sugary	7	3.37
NS	ZeroCal	2	0.990
NS	Sugary	2	0.990
NS	Juice100	2	0.990
NS	Total	2	0.990

Figure 3: Number of missing observations and percentage missing by variable