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

Christine Chuong (20124798)

Giuliano Cruz (79247979)

Joey Hotz (18461400)

2024-03-01

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 [1]. Multiple policy interventions to inhibit consumption of SSBs have been proposed [2], 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 kinds of interventions which intend to shift consumers towards zero-calorie beverages: interventions with caloric messaging or price discounts. There are three caloric messaging interventions and two price discount interventions.

The primary statistical question is whether these interventions are associated with differences in the average proportion of daily sales of zero-calorie beverages. Secondary questions include whether there are significant interactions between interventions and site effects, and whether a combination of interventions are more effective than each individual intervention.

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. There are 631 observations across all three sites. Each observation represents the total daily sales of bottled beverages across multiple storefronts at a single site. Sales are recorded automatically by point of sale terminals in Site A, and manually by cashiers at Sites B and C. Sales are recorded every day from the start 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. Categorical variables include weekday with seven levels, site with three levels, and intervention with nine levels. The daily sales metrics include the number of bottled zero-calorie and sweetened beverages sold per site. These are each non-negative integer variables. The data also includes the total number of beverages sold per site, which is another non-negative integer value.

Some observations include the sales of 100% juice, orange juice, and sports drinks, which are recorded as integers. These counts are only consistently recorded at Site A, and rarely recorded for Site B and C. There are nine days which are missing counts for the bottled zero-calorie and sugared beverages. Counts for the bottles 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 Site B, and two occur mid-study at Site C.

Exploratory Analysis

Exploratory data analysis can visualize relationships among variables, check distributional assumptions, and informally give some insight into the results that may be expected. Based on the data and statistical questions, the following figures are recommended:

1. Side-by-side boxplots of zero-calorie or sugary drink sales by variables such as day of the week or site can help identify if there are differences in sales between different levels of the categorical variables.
2. Plots of the proportion of total sugary and zero-calorie sales coming from zero-calorie drinks over counts, such as Figure 1, can be used to observe and identify the patterns over time. These plots can also be stratified by other variables, such as intervention to identify differences between groups.
3. A scatterplot of zero-calorie vs sugary drink sales can be used to investigate the kind of relationship they may have (eg. linear, positive) and to potentially identify if they are highly correlated.

Additionally, a missing data table can summarize how many observations are missing for each variable. A missing data plot could be used to visualize the extent of the missing data.

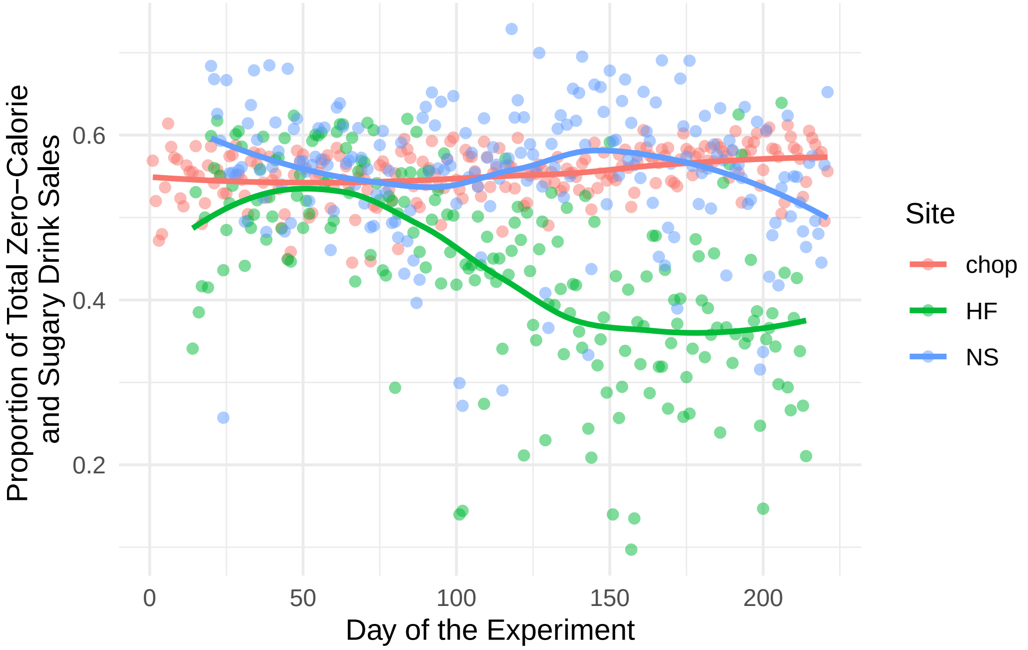

Figure 1: Proportion of total sugary and zero-calorie drink sales coming from zero-calorie drinks over time

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 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] Malik VS, Hu FB. The role of sugar-sweetened beverages in the global epidemics of obesity and chronic diseases. Nature Reviews Endocrinology 2022;18. https://doi.org/10.1038/s41574-021-00627-6.

[2] Krieger J, Bleich SN, Scarmo S, Ng SW. Sugar-sweetened beverage reduction policies: Progress and promise. Annual Review of Public Health 2021;42. https://doi.org/10.1146/annurev-publhealth-090419-103005.

[3] Talbot D, Mésidor M, Chiu Y, Simard M, Sirois C. An alternative perspective on the robust poisson method for estimating risk or prevalence ratios. Epidemiology 2022;34:1–7. https://doi.org/10.1097/ede.0000000000001544.

Statistical Appendix