Food Labeling Policies: Aggregate Impacts and Heterogeneity Across Categories

Nano Barahona Cristóbal Otero Sebastián Otero Joshua Kim*

September 1, 2026

Abstract: We study the aggregate and heterogeneous effects of a front-of-package labeling policy implemented in Chile. We find that consumers reduced their sugar and caloric intake by 9% and 6%, respectively. On the demand side, labels prompt consumers to substitute within categories rather than switching between categories. Within-category responses are more pronounced when labels provide new information. On the supply side, we observe bunching at regulatory thresholds, with substantial heterogeneity across categories, consistent with differing costs of product reformulation. We conclude that considering policy-response heterogeneity is key for effective policy design.

JEL Codes: D12, D22, I12, I18, L11, L81

*An earlier version of this paper circulated under the title "On the Design of Food Labeling Policies." We want to thank Hunt Allcott, Liran Einav, Matthew Gentzkow, Dmitry Taubinsky, and seminar participants at the Haas School of Business and the Virtual Quantitative Marketing Seminar for valuable comments and suggestions. We also thank Camila Corvalán and Marcela Reyes for beneficial conversations on institutional details; Christine Von Dessauer, Roberto Cases, Romina Quagliotti, José Tomás Feliu, and Enrico Ruggieri for excellent research assistance; and Alejandro Guin-Po and Fernanda Mediano for their contributions to the data-collection process. Finally, we thank Walmart-Chile and Instituto de Nutrición y Tecnología de los Alimentos (INTA) and OK to Shop for sharing the data for the project, and the Stanford Center of Population Health Sciences (PHS) for providing a secure environment to store and analyze the data. Barahona: UC Berkeley and NBER, nanobk@berkeley.edu. C. Otero: Columbia Business School and NBER, c.otero@columbia.edu. S. Otero: Columbia University and NBER, so2699@columbia.edu. Kim: Uber, josh@joshkim.org.

1. Introduction

The average American adult weighed nearly 23 pounds more in 2018 than in 1975 (NCHS, 2018). This dramatic rise in obesity is not specific to the United States: Over the same period, obesity in the world has tripled, and today, roughly 40% of the world's adult population is considered to be either obese or overweight (WHO, 2018). In response to this health pandemic, governments around the world are grappling with how to design policies that effectively improve diet quality.

An increasingly popular policy is to provide simplified information about products' healthiness to consumers through front-of-package (FOP) warning labels. These labels are simple symbols that clearly signal to consumers when a product is considered unhealthy based on whether targeted critical nutrients—such as sugar and calories—exceed certain concentration thresholds. These are single-threshold labels in the sense that the regulatory thresholds are uniform across product categories and only make a distinction between solids and liquids. Chile was a pioneer country in implementing government-mandated FOP warning labels in the Chilean Food Act in 2016. Since then, more than 30 countries have approved or are considering similar regulations based on single-threshold binary labels.[1]

In this article, we review the Chilean experience and study the effects of food labeling policies across a wide range of product categories. In previous work, we focused on a single category—breakfast cereals—and constructed an equilibrium model showing that the Chilean Food Act reduced the consumption of unhealthy nutrients in that category (Barahona et al., 2023). This article builds on that study in two ways. First, we evaluate the policy's impact on overall nutritional intake. Second, we decompose the overall impact of the policy into supply- and demand-side responses and document substantial heterogeneity in the substitution effects of labels and in product reformulation across categories, which are key for effective policymaking.

To investigate the impact of the regulation across all product categories, we leverage access to Walmart's scanner data in Chile. The data contain the universe of food purchases made in Walmart-Chile between 2015 and 2018. Walmart is the country's largest

[1]There are ongoing discussions to implement similar policies in the US (The Washington Post, 2024).

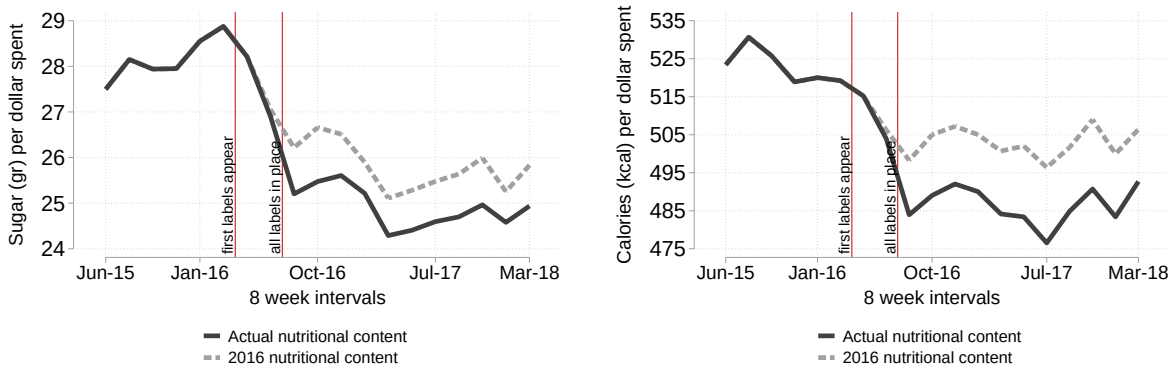
food retailer and accounts for more than 40% of supermarket sales. We combine these data with the products' nutrition facts tables from before the policy (2016), after its initial implementation (2018), and after the final stage, under which regulatory thresholds became stricter (2021). Finally, we use Walmart's loyalty program to follow consumers over time and produce individual-level measures of nutritional intake.

Our first contribution is to document a sharp aggregate decrease in sugar and caloric intake of 9% and 6% per dollar spent, respectively, immediately after the policy was phased in. We present these findings in Figure 1. The reduction in sugar and caloric intake—which persists for the 2-year post-policy window in our data—is explained by a combination of demand- and supply-side responses. Consumers reacted to the regulation by making healthier choices, even when the nutritional content of products is kept constant over time (dashed curves). Firms responded by reducing the concentration of critical nutrients in their products, thus offering a healthier bundle of products (the difference between solid and dashed curves).[2]

Separating this overall effect into its demand- and supply-side components matters for food-labeling policy. Unlike a tax, whose efficiency cost can often be summarized by the equilibrium quantity response (Chetty, 2009), the welfare implications of a label also depend on who responds because labels can affect consumers heterogeneously (Allcott et al., 2025). Moreover, a given reduction in sugar intake may involve different economic costs depending on whether consumers switch products or firms reformulate them.

We begin with the demand side, decomposing its effects into *between-* and *within-*category substitution. First, we study whether the food labeling policy has the potential to shift consumer demand between categories. To test for the presence of cross-category substitution effects, we compare categories with different shares of labeled products and examine whether categories with a low share of labeled products increased their revenue relative to categories with a high share of labeled products. We find that the extent to which consumers substituted between categories due to the presence of labels is negligible and cannot explain the patterns we document in Figure 1. The lack of between-category

[2]In Appendix A, Figure A.1, we show figures that divide nutritional intake by the volume of food purchased instead of by dollars spent. Overall, the findings and key takeaways are similar to the analysis above.



(a) Sugar intake

(b) Calorie intake

Figure 1: Nutritional intake per dollar spent before and after the policy

Notes: We produce this figure using a panel of Walmart consumers and computing their sugar and caloric intake during every 8-week period. After the labeling policy was introduced, total sugar intake decreased from 27.5 to 24.9 grams of sugar per dollar, and total caloric intake decreased from 523 to 493 kcal per dollar. The solid curve represents the total amount of sugar or calories purchased for every dollar spent in every 8-week period. The dashed curve is constructed in the same way as the solid curve, but fixing products' nutritional content at their 2016 values. The left vertical line corresponds to when the first labels appeared, and the right vertical line corresponds to when the Food Act became mandatory. We have two snapshots of nutritional information data: one from early 2016, before the policy was introduced, and one from 2018, after the policy was introduced. We assume that all changes in nutritional content occurred around the date of policy implementation (June 2016), and thus use these two snapshots for all pre-policy and post-policy nutritional values, respectively, in our calculations.

substitution is consistent with the limited demand effects observed in cross-category price promotions (Leeflang and Parreño-Selva, 2012). This phenomenon could be attributed to several factors: the spatial arrangement of cross-category products in stores (Kan et al., 2023), the pre-planned purchasing behavior targeting certain categories, often referred to as "destination categories" (Briesch et al., 2013), and the general stickiness and persistence of dietary habits (Hut and Oster, 2022; Kluser and Pons, 2026).

Next, we document important *within*-category substitution between labeled and unlabeled products in several product categories. We compare the quantities sold of labeled and unlabeled products before and after the introduction of the policy. We focus on categories for which there is enough variation in the share of labeled products and for which labeled and unlabeled products followed similar trends before the implementation of the policy. We document substantial heterogeneity in these substitution effects across categories, including a 26% decrease in equilibrium quantities of labeled relative to unlabeled products for cereals and a 12% decrease for soft drinks.

To understand the source of this heterogeneity, we implemented a survey in Argentina in which we elicited consumers' beliefs about the nutritional content of soft drinks and cereal, two important categories with smaller and larger substitution effects, respectively. We find that beliefs are accurate for soft drinks but not for cereal. In line with Barahona et al. (2023), these results are consistent with the idea that food labels are more effective in shifting demand in categories in which labels are more informative.

On the supply side, firms might respond to labeling policies by reformulating their products and avoiding labels. To empirically assess these responses, we compare the distribution of nutritional content before and after the policy's implementation. We document a significant amount of bunching at the regulatory threshold in several product categories, with important heterogeneity across product categories. For instance, whereas virtually all products above the regulatory thresholds were reformulated in the yogurt category, only 8% of ex-ante labeled cookies changed their nutritional composition.

Product reformulation is more likely when the demand effects of labels are larger, when the threshold is close to the original nutritional content, and when reformulation costs are lower. For example, in categories such as yogurt or juice, firms can reformulate their products by substituting sugar with other low-cost sweeteners that mimic the products' taste. On the other hand, in cereal or cookies, sugar serves as a bulking agent, and replacing it with low-cost sweeteners may cause them to crumble. Consistent with this pattern, we also find that as the regulatory thresholds became tighter, the supply side responded by sequentially bunching at the stricter thresholds, but only in categories where reformulation was feasible at a low cost. An important implication of these supply-side responses is that they improve nutritional intake even for consumers who do not respond to the labels themselves.

Our findings are also important for policy design. First, the lack of between-category substitution implies that policymakers need to consider the effects of regulatory thresholds within specific categories. Second, if the policy objective is to maximize aggregate reductions in targeted nutrients, the threshold for each nutrient can be set to target categories that represent a large share of consumers' bundles in terms of the overall intake of that nutrient. Third, these categories need to have both healthy and unhealthy products that are close substitutes, and labels are likely to be more effective when they provide

consumers with new information about the healthiness of those products. Fourth, policymakers should also consider the extent to which reformulation is feasible at a low cost in the targeted categories.

Where to set the threshold depends on how demand-side substitution and supply-side reformulation combine. A stricter threshold increases overall reformulation by exposing more products to the label, but it also leaves consumers with fewer unlabeled alternatives to substitute toward. The net effect of moving the threshold is therefore an empirical question. We conclude by exploring it with a calibration that combines the estimated demand and reformulation responses of each category, holding prices and aggregate demand by category fixed. Under these assumptions, predicted sugar purchases are lowest at a threshold of about 5 grams per 100 grams for solids and about 3.50 grams per 100 milliliters for liquids, below the thresholds adopted by Chile.

Our paper contributes to four strands of the literature. First, it adds to the literature that investigates the effect of food labeling regulations on the demand for food. Most of this work has focused on specific food categories, chain restaurants, or experiments in the lab and in the field.[3] We add to this literature by providing evidence of and quantifying the effects of a national food labeling regulation on overall consumption of sugar and calories across all product categories.

Second, it speaks to the studies that examine the Chilean Food Act. The main overarching finding is that labels induce demand effects within a category. Taillie et al. (2020) document a decline in purchases of labeled beverages following the policy's implementation. Araya et al. (2022) take advantage of the staggered introduction of labeled products to store inventories and find that—in the very short run—labels decrease demand in the breakfast cereal category, but not for chocolates or cookies. Alé-Chilet and Moshary (2022), Pachali et al. (2023), and Barahona et al. (2023) study the effects of the policy on breakfast cereal and find strong substitution from labeled to unlabeled products.

[3]Examples include salad dressing (Mathios, 2000); microwave popcorn (Kiesel and Villas-Boas, 2013); sugar-sweetened beverages (Taillie et al., 2020); cheese and yogurt (Allais et al., 2015); ready-to-eat breakfast cereal (Zhu et al., 2015); chain restaurants (Wisdom et al., 2010; Bollinger et al., 2011; Finkelstein et al., 2011); laboratory experiments (Bix et al., 2015; Ono and Ono, 2015; Crosetto et al., 2016, 2020; Ravaioli, 2021); and field experiments (Shangguan et al., 2019; Dubois et al., 2021). Moreover, different FOP label designs may exhibit different levels of effectiveness; see Dubois et al. (2021) and Ravaioli (2021) for a discussion.

Avishay-Rizi and Reshef (2023) study a similar food labeling policy in Israel and document heterogeneous demand-side responses across food categories. Our paper contributes to these studies by focusing on substitution between categories and within multiple categories, and by rationalizing these heterogeneous effects.

Third, we contribute to the literature that documents product reformulation responses to nutritional information policies. In the context of the same Chilean regulation we study, Alé-Chilet and Moshary (2022) and Barahona et al. (2023) provide evidence of bunching just below regulatory thresholds in the cereal market.[4] Relative to these studies, we document important heterogeneity across categories in supply-side responses and discuss the drivers of these differences, and how they matter for policy design. We also add to these papers by looking at supply-side responses in the longer run after the strictest regulatory thresholds were in place.

Fourth, we contribute to a broader literature that studies how governments can help consumers make better nutritional choices (Cawley, 2015), whether by improving access to healthy food in poor neighborhoods (Allcott et al., 2019), restricting junk-food advertising (Dubois et al., 2017), or taxing sugar-sweetened beverages and calorie-dense products (Falbe et al., 2015, 2016; Taylor et al., 2019; Silver et al., 2017; Dubois et al., 2020; Allcott et al., 2019a; Aguilar et al., 2021; Lee et al., 2019). In line with Paraje et al. (2026), who find that cohorts with greater exposure to the Chilean Food Act were less likely to be overweight or obese, we show that food labels can effectively improve diet quality when carefully designed.

The remainder of the paper is organized as follows. Section 2 describes the setting and the data. In Section 3, we present empirical evidence of how labels impact overall nutritional intake through demand- and supply-side responses. We discuss policy implications in Section 4 and conclude in Section 5.

[4]For evidence of reformulation responses to nutritional information policies in other settings, see Unnevehr and Jagmanaite (2008), Moorman et al. (2012), Griffith et al. (2017), and Lim et al. (2020). Reyes et al. (2020) and Quintiliano Scarpelli et al. (2020) also show a reduction in critical nutrient concentration of multiple products after the policy's implementation.

2. Setting and data

In recent years, many countries have introduced FOP labels to help consumers make healthier food choices. Unlike nutrition facts tables, FOP labels simplify nutritional information, which makes it easier to use and interpret in a context in which shoppers make quick purchasing decisions (Temple, 2020). A recent development is warning labels that indicate whether a food product has a relatively high content of a critical nutrient, such as sugar, sodium, fat, or calories. Relative to other FOP labels, warning labels are simple binary symbols that clearly signal to consumers when a product has a high concentration of a given nutrient. Perhaps due to their simplicity, warning labels have become popular in the last few years. Following Chile's implementation in 2016, more than 30 countries, including Argentina, Brazil, Canada, Israel, and Mexico, have either implemented or are discussing the implementation of country-wide mandatory food labeling policies.[5]

Mandatory warning labels can be justified from both a demand- and supply-side perspective. In terms of consumer behavior, labels could help mitigate biases that drive consumers to over-purchase unhealthy products beyond their true preferences, such as inattention to potentially harmful health effects, and poorly calibrated beliefs over products' nutritional content (Bernheim and Taubinsky, 2018; Allcott et al., 2019b; Barahona et al., 2023).

Food labels can also incentivize firms to produce healthier products. In the absence of mandatory labels, firms do not have incentives to invest in reducing the concentration of critical nutrients in their products.[6] If consumers increase the demand for healthier products in response to labels, firms can benefit from reducing the concentration of regulated nutrients to avoid getting a label. In equilibrium, regulated markets with labels can induce consumers to make better nutritional decisions and firms to offer a healthier bundle of products (Barahona et al., 2023).

[5] Table A.2 presents the list of countries, including the critical nutrients targeted in the regulation.

[6] Although suppliers of healthy products have incentives to solve these inattention and informational problems in a well-functioning market, several frictions make self-regulation unlikely, providing grounds for government intervention. Dranove and Jin (2010) discuss several conditions that are required for markets to unravel.

2.1. *The Chilean Food Act*

Chile was the first country to introduce a nationwide mandatory FOP warning-label policy.[7] In response to high rates of obesity, the most prevalent chronic disease in the country, Chile implemented Law 20.606 in 2016 (hereafter, the Food Act), which introduced FOP warning labels to inform consumers about products' healthiness and help guide purchasing decisions. [8] The rationale was that nutritional information available in the form of a fact table on the back of the product was too complex and "did not allow [consumers] to make an informed decision" (Historia de la Ley 20.606, 2011, p. 170).

The Chilean Food Act mandated that products with calories, added sugar, saturated fats, and sodium higher than a given threshold must include a FOP warning label for each nutrient threshold surpassed. Figure 2 shows what Chilean FOP warning labels look like and how they are displayed on actual products. The thresholds were established uniformly for all food products, depending on whether the product is a solid or a liquid. To define the thresholds, the legislators chose the 90th percentile of the distribution of the concentration of critical nutrients from non-processed food products according to the USDA. The introduction of the thresholds was gradual and implemented in three stages. Stages 1, 2, and 3 took place in June 2016, 2018, and 2019, respectively. All stage thresholds were announced at the outset of the policy, ensuring they were publicly known. Threshold values are presented in Appendix A, Table A.1.[9]

[7]Chile was the first country to approve mandatory national FOP warning labels for food high in calories, added sugars, saturated fats, and sodium and to implement the warning labels for all processed food products. Before 2016, several countries had implemented voluntary FOP labels. For example, Sweden, Denmark, Norway, Lithuania, and Iceland used the Keyhole logo; The Netherlands, Belgium, and Poland used the Choices logo; Korea and the United Kingdom used traffic-light labels; and Singapore used the Healthier Choice Symbol. Finland implemented a mandatory warning label in 1993 but only for some products high in salt. Thailand introduced a mandatory GDA label in 2007 but only for five categories of snacks. Also, Ecuador and Iran implemented mandatory traffic-light labeling for all processed products in 2014 and 2015, respectively.

[8]The Food Act also included bans on selling labeled products in schools and on advertising labeled products aimed at children younger than 14 years of age.

[9]The Food Act has a relevant exception that limits its scope to regulating only processed and packaged foods. This means that even if a product exceeds a certain threshold, it will not require a label if it does not contain added sugars, sodium, saturated fats, honey, or syrup.



Figure 2: FOP warning labels on selected products

Notes: The figure presents both the FOP warning labels implemented in Chile and how these are displayed on various food packages. The labels say, from left to right, "High in sugar," "High in saturated fat," "High in sodium," and "High in calories." Products can have from zero to four labels. Table A.1 presents the threshold values that determine the assignment of each label.

2.2. *Data*

The main outcome measures used in the analysis come from scanner-level data from Walmart, which we expand using nutritional information and a survey that captures consumers' beliefs about product healthiness.

2.2.1. *Walmart data*: We use data from Walmart-Chile, the largest food retailer in Chile, responsible for over 40% of supermarket sales. The data cover all Walmart store transactions between May 2015 and March 2018, and identify products by a Universal Product Code (UPC). For each transaction, we have access to information such as a product's price, revenue, product name, brand name, and discounts.[10]

We use Walmart's loyalty program to connect transactions with individual shoppers over time. We focus on regular Walmart customers who visited a store at least once every 8 weeks during the study period, leading to a total of roughly 524,000 individuals. We have access to information on these customers, including their gender, age, and household income. The average customer in our study is 48 years old, and 69% are women.

[10]The data comprise over 9 billion transactions by over 5 million consumers for 26,297 different food products.

Before the policy was introduced, the median customer shopped at Walmart 24 times, at three different Walmart locations, and traveled about 3 kilometers to get to the nearest store. The sample is broadly representative of the Chilean urban population, though high-income consumers are somewhat overrepresented. One-third of consumers fall within the bottom 50% of the national income distribution, another third between the 50th and 85th percentiles, and the final third in the top 15%.

2.2.2. *Nutritional Information*: The nutritional data for packaged products come from three sources: (a) pre-policy data collected by the Institute of Nutrition and Food Technology (INTA) at the University of Chile, (b) post-policy data we collected and digitized in 2018, and (c) longer-run post-policy data, current through 2021, collected by a private firm.

INTA collected nutritional information for a sample of products in January 2016 at the UPC level. This included the nutrition facts table, whether the product is a liquid or a solid, and the package size. We developed a phone app that linked images of nutrition facts tables to UPCs to collect nutritional information for the post-policy period. We then digitized the nutritional content of all available products in the three largest Walmart stores in Chile. We collected this information in March 2018, two years after the first stage of the labeling law was implemented in June 2016. For nutritional information data after the third stage, we partnered with "OK to Shop," a startup founded in 2018 that collects detailed information on food products. We accessed 2021 data for all their available products and matched it at the UPC level with nutritional information from previous years.

To include information on non-packaged products, such as fresh produce or meat—which do not have nutrition facts tables—we consulted publicly available data from the USDA's FoodData Central. We used these data to complete any missing information on critical nutrients across all food categories.

2.2.3. *Consumer beliefs*: We surveyed 1,664 consumers to elicit beliefs about the nutritional characteristics of packaged food products. We asked participants to provide an estimate of the sugar and caloric content of certain cereal and soft drink products. We

conducted the survey in Argentina in August 2019, when there was no food labeling policy in place.[11]

3. Empirical evidence

This section provides evidence of demand- and supply-side responses to food labels. We investigate the degree to which labels prompt consumers to substitute products within and between categories, the importance of the accuracy of consumers' beliefs for the effectiveness of food labels, and the extent to which products have been reformulated to avoid being labeled.

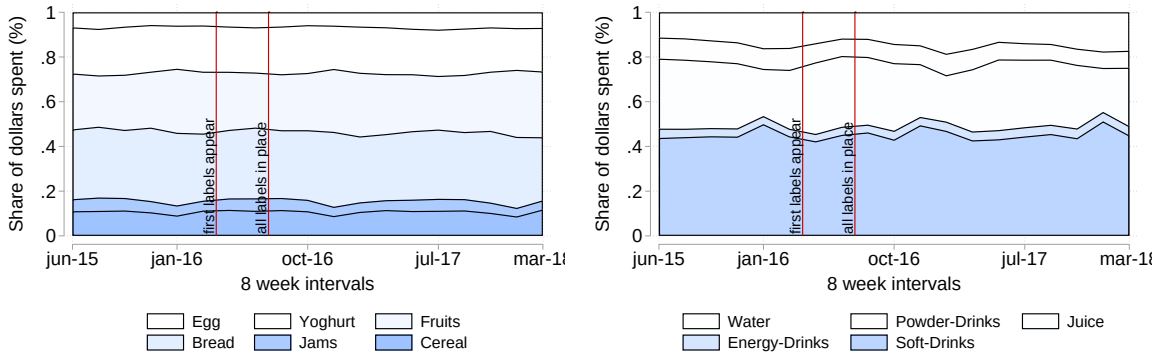
3.1. *Demand-side responses*

3.1.1. *Between-category substitution*: We start by examining whether consumers shift consumption across food categories as a response to food labels. To do so, we define broader groups of products that contain multiple categories in which we could expect substitution to occur. For instance, we check whether there was substitution among product categories that are likely to be eaten at breakfast: eggs, yogurt, bread, fruits, jams, and breakfast cereal. Then, within each broader group, we compare revenues before and after the policy for categories with a high and a low share of labeled products, where we define labeled products as all products with at least one warning label.

In Figure 3, we plot changes in the share of revenue over time of the food categories that fall into the breakfast and drinks food groups. In each group, categories are ordered from top to bottom according to the share of labeled products they contain (weighted by pre-policy revenue). The darker the area's color, the larger the share of labeled products. For instance, in breakfast products, 0% of egg products are labeled, while in cereals—the category with the highest share of labeled products in this group—61% of the products are labeled.

Figure 3 suggests there is little to no evidence that consumers are shifting their spending from highly labeled categories, such as breakfast cereals or soft drinks, to low-labeled

[11] Although we would have preferred to conduct the survey in Chile before the policy's implementation, we used Argentina as a proxy because it has a population and food market similar to Chile's and, at the time of the survey, had not been exposed to a labeling policy.



(a) Breakfast

(b) Drinks

Figure 3: Share of dollars spent across categories

Notes: The figure shows the evolution of the share of dollars spent in each category within broader groups of products. Colors represent the share of labeled products within each category. White areas are categories in which no product received a label, and dark-blue areas are categories with a high share of labeled products. We show no differential changes in dollars spent between low-in-labels and high-in-labels categories.

categories, such as eggs or juices. In Panel (a), cereals account for an average of 10.8% of breakfast spending both before and after the policy, and Panel (b) reveals a similar stability for water, soft drinks, and related beverages. In Appendix A, Figure A.2, we show that this finding extends to several other food groups, such as carbs, meats, desserts, and snacks.

To formalize these results, we pool all food categories together (not only breakfast and drinks) and run the following regression:

$$\log(r_{cst}) \quad = \quad \sum_{k} \beta_k \cdot L_c \cdot \mathbb{1}\{k = t\} + d_{cs} + \delta_t + \varepsilon_{cst}, \tag{1}$$

where r_{cst} denotes the total revenue from products in category c sold in store s in period t, and L_c is the (weighted) share of products in category c that have at least one label.[12] Finally, δ_t denotes period fixed effects, and d_{cs} refers to category-store fixed effects. We normalize the β_k coefficients so that their average value over the pre-policy period is equal to zero. Observations are weighted by category-store pre-policy revenues, and standard errors are clustered at the category level.

Figure 4 displays the results from estimating Equation (1). We find that β_k esti-

[12] L_c is calculated by weighting observations by their pre-policy revenue.

mates are small in magnitude, not statistically significant, and revolve around zero after the policy is in place. Reassuringly, we do not observe that categories with different shares of labeled products are differentially trending before the policy after accounting for category-store and period fixed effects. Regression results are consistent with the results for breakfast products and drinks presented in Figure 3. Overall, these results suggest that the extent to which consumers substituted toward other categories due to the presence of labels is negligible and cannot explain the patterns we document in Figure 1.

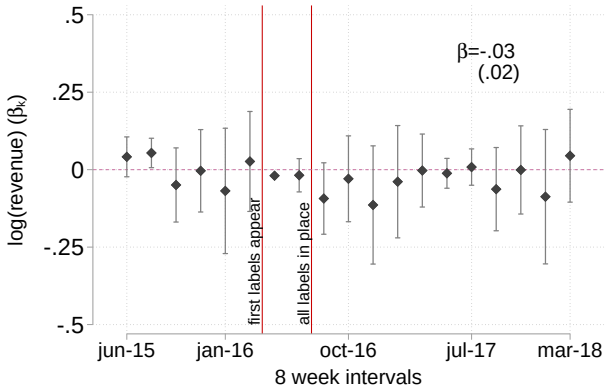


Figure 4: Changes in total spending per capita across categories

Notes: The figure presents the β_k coefficients from Equation (1). These regressions are run on a sample of 69 categories, 19 periods, and 167 stores. There is a total of 209,290 observations. The average share of labeled products in each category is 0.3, with a minimum of 0 and a maximum of 1. Vertical lines delimit the 95% confidence intervals of a test against the normalized coefficient (May 2016). Since different categories are affected by seasonality differently, coefficients closer to the winter season in Chile (May-July) are more precisely estimated when we compare them to the normalized coefficient.

A caveat of our analysis is that the no-interference assumption—standard in these types of research designs—does not hold. Naturally, a decrease in demand for labeled products will likely increase purchases of unlabeled ones. Since all markets are treated with the policy, no category or product can be used as a clean control group. This complicates the interpretation of the estimated coefficients. Nevertheless, the lack of significant coefficients suggests there is no detectable systematic shift in consumption from high- to low-labeled categories.

3.1.2. *Within-category substitution*: Next, we examine the effects of food labels within each food category, where products are more likely to be close substitutes. For our analysis, we limit our attention to categories in which labeled products represent more than 5% and less than 95% of the total category revenue, and in which there are more than ten labeled and ten unlabeled products. We end up with 17 categories that represent 34.2% of the pre-policy revenue of all packaged food products and 40.7% of the pre-policy revenue of all labeled products. Appendix B details the sample selection procedure and provides descriptive statistics on the share and type of labels across products and categories.

We define a product as the union of UPCs that share the same product name and brand. For example, we assign all *Diet Coke* UPCs the same product ID regardless of can or bottle size. We assign labels to a product based on its 2018 post-policy nutritional content. We collapse our original data into product-store-period data bins (in which a period is defined as 8 consecutive calendar weeks) and estimate the following regression for each category:

$$\log(q_{jst}) \quad = \quad \beta \cdot L_{jt} + \gamma \cdot \log(p_{jst}) + \delta_{js} + \delta_t + \varepsilon_{jst}, \tag{2}$$

where q_{jst} denotes the grams (ml) of product j sold in store s in period t, p_{jst} refers to the product's price per 100 grams (ml), and L_{jt} is an indicator variable that takes the value of one in the post-policy period if the product has at least one label. Finally, δ_{js} refers to product-store fixed effects and δ_t to period fixed effects. Observations are weighted by product-store pre-policy revenues. Products that do not appear in the pre-period have zero weight and are thus excluded from the estimation sample. Standard errors are clustered at the product level.[13]

In Figure 5, we plot the estimated changes in demand for each of the categories. We find a significant impact of the policy for most categories. These results confirm that consumers substituted from labeled to unlabeled products within most of these product categories and that the effect holds in the medium run. Overall, these findings suggest

[13]We assume that products with zero sales in a given supermarket-period were not available in that market and do not include them in the regression.

that the policy is effective in shifting consumption in many product categories and thus complements the policy impact on breakfast cereal documented by Barahona et al. (2023). Interestingly, the effects vary significantly across product categories. In categories like bread toast and cereal, the proportion of labeled products sold compared to unlabeled ones dropped by an average of 25-30% following the implementation of the regulation. In contrast, categories such as soft drinks and ice creams saw a smaller change, around 10%. For categories like burger patties and candy, there was no notable difference in sales before and after the regulation.

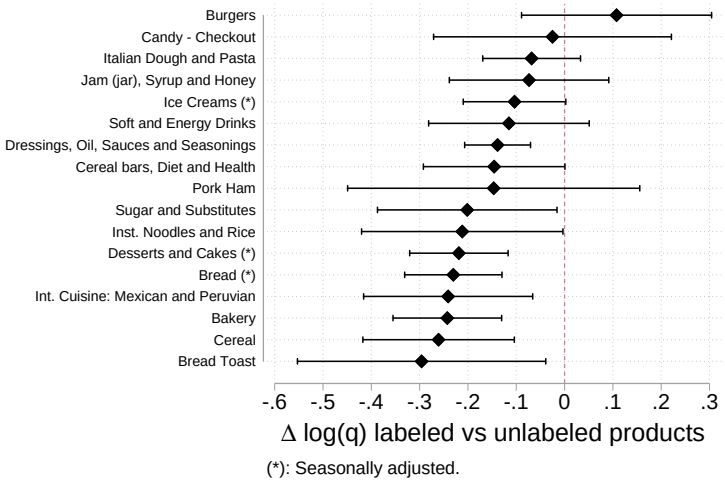


Figure 5: Changes in demand for selected categories

Notes: This figure shows the changes in equilibrium quantities between labeled and unlabeled products from estimating Equation (2) for different product categories. The set of products in this analysis represents 34.16% of the pre-policy revenue of all products in the sample. We provide more details on sample selection in Appendix B.

As in the previous analysis, consumers substitute from one product to another, and the no-interference assumption does not hold. In the extreme case of one-to-one substitution, a β of -10% would reflect a 5% decrease in labeled products and a 5% increase in unlabeled products. As a result, our coefficients should be interpreted as the before-after impact of the policy on the relative differences in equilibrium quantities between labeled and unlabeled products and not as the direct effect of labels on purchases of labeled products.

A second caveat of this analysis is that prices are endogenous, and including them as controls can complicate the interpretation of the coefficients even further. In the current specification of Equation (2), β accounts for the change in the relative differences

in equilibrium quantities between labeled and unlabeled products before and after the policy, *conditional* on the observed prices. If prices were omitted from Equation (2), β would capture changes in the relative differences in equilibrium quantities induced both by the labels and the endogenous changes in prices. In Appendix A, Figure A.4, we show that results omitting prices from Equation (2) look qualitatively similar for almost all categories.[14]

Next, we explore a mechanism to make sense of the dispersion in observed effects.

3.1.3. *The importance of prior beliefs*: In related work, Barahona et al. (2023) show that the effect on cereals is mostly explained by substitution away from products that consumers believed to be healthy but which ended up with a label. In other words, at the product level, labels are more effective if they provide new information to consumers. This mechanism echoes Bollinger et al. (2011), who show that calorie posting in chain restaurants operates by correcting consumers' underestimates of caloric content, and who also measure beliefs using surveys in jurisdictions without the policy. To investigate how information and beliefs affect the extent of *within*-category substitution, we use the beliefs survey described in Subsection 2.2.3 and compare the effects of the policy on cereal and soft drinks, the two categories covered by the survey.

The pre-policy concentration of sugar in soft drinks follows a bimodal distribution driven by diet and non-diet drinks, which is highly correlated with consumers' beliefs about sugar concentration in these products. The correlation between the average value of respondents' beliefs about the sugar concentration of each product and the product's observed pre-policy sugar concentration in the soft drinks category is 0.95. In cereal, however, consumers have mistaken beliefs about the caloric content of products. The correlation between the average value of respondents' beliefs about the caloric concentration of each product and the product's observed pre-policy caloric concentration is 0.28. We present the relationship between consumer beliefs and pre-policy nutritional content in

[14]We measure price as the unit value at the UPC-store-period level (revenue divided by quantity sold) and aggregate to the product level using contemporaneous sales-volume weights. Because a product bundles UPCs of different package sizes, and price per unit of volume varies with size, this index can move mechanically if the labels shift consumers across sizes within a product. In Appendix A, Figure A.5, we estimate Equation (2) with a price index that holds the UPC weights fixed at their pre-policy values. The estimated demand effects are quantitatively similar. In Appendix A, Figure A.6, we investigate the effects of the policy on prices. Overall, the effects across categories are very limited.

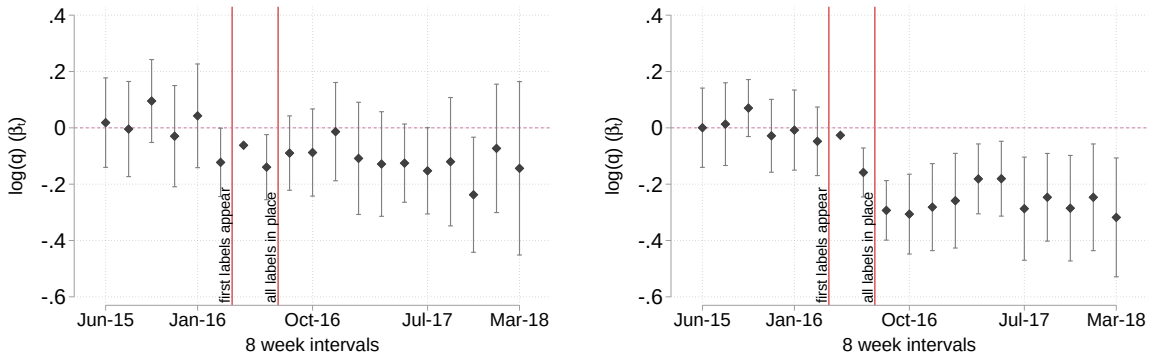
Appendix A, Figure A.7.

The high correlation between consumer beliefs and actual sugar content suggests that labels provided less new information in this category, consistent with a smaller policy effect in soft drinks relative to cereal. We estimate the following regression for each category:

$$\log(q_{jst}) = \sum_k \beta_k \cdot L_j \cdot \mathbb{1}\{k = t\} + \gamma \cdot \log(p_{jst}) + \delta_{js} + \delta_t + \varepsilon_{jst}, \tag{3}$$

where all variables and specification details are defined as in Equation (2). We normalize the β_k coefficients so that their average value over the pre-policy period is equal to zero.[15]

Figure 6 displays the results of estimating Equation (3) for soft drinks and cereal. Pre-period coefficients are small and not significantly different from zero in both categories. After the introduction of the labels, we observe a noticeable drop in the demand for labeled products relative to that for unlabeled products in the cereal market but not in the soft drinks market.



(a) Soft Drinks (b) Cereals

Figure 6: Dynamic changes in demand for soft drinks and cereals

Notes: This figure presents regression coefficient estimates from Equation (3). Panel (a) presents the β_k coefficients of the regression on the sample of 72 soft drinks that appear in both the pre- and post-policy periods. The sample consists of 42 unlabeled and 30 labeled products for a total of 152,397 observations. Coca-Cola Company reformulated two products—Regular Fanta and Regular Sprite—to remove their high-in-sugar label in September of 2017. We define those products as unlabeled in Equation (3) (i.e., $L_j = 0$). Panel (b) displays the β_k coefficients of the regression on the sample of 75 ready-to-eat cereals that appear in both the pre- and post-policy periods. The sample consists of 27 unlabeled and 48 labeled products for a total of 186,612 observations.

[15] In Appendix A, Figure A.3, we present estimates of Equation (3) for all categories. Pre-period coefficients are small and not significantly different from zero in all categories. For most categories, we observe a noticeable drop in demand for labeled products relative to that for unlabeled products after the introduction of the labels. This effect persists throughout our observational time window.

Several alternative mechanisms could help explain our findings. The contrast between cereals and soft drinks, for instance, may partly reflect differences in sugar distributions (continuous vs. bimodal), which could make substitution within cereals more likely. Heterogeneous preferences across categories may also play a role, as consumers might regard sugar as more acceptable in some categories than in others—for example, due to mental accounting. Promotions and product placement may have shifted in response to the policy, and the salience of labels could also contribute to reduced demand.

While we cannot fully rule out these mechanisms, the evidence suggests they are unlikely to be the main drivers. Our estimates remain robust when we omit price controls, indicating that price promotions are not driving the results. A mechanism based solely on product placement or label salience would predict broadly similar responses across all labeled products.[16] Instead, we observe stronger demand effects in categories where labels plausibly corrected misperceptions (cereals) and weaker effects where prior beliefs were already accurate (soft drinks). Our beliefs interpretation is consistent with complementary evidence in the literature showing limited or no impacts in categories where labels conveyed little new information, such as chocolates and cookies (Araya et al., 2022).[17] Since we do not observe promotions, shelf placement, or end-of-aisle displays in our data, we cannot test these channels directly; to the extent that such supply-side responses occurred, they are subsumed in our estimated equilibrium effects.

3.2. *Supply-side responses: Product reformulation*

We next examine supply-side responses to the labeling regulation. Firms could respond either by reformulating existing products or by adjusting their product portfolio through entry and exit. As documented in Appendix C, we find little evidence of adjustment along the entry/exit margin, which motivates our focus on reformulation. Specifically, we analyze whether firms reduced sugar content in categories where reformulation was both feasible and economically meaningful—i.e., categories in which the distribution of sugar

[16] In Barahona et al. (2023), Online Appendix C.2, we test directly for salience effects in the cereal category: if labels operated mainly by making unhealthiness salient, demand responses should be stronger for labeled products with higher concentrations of critical nutrients. We find the opposite pattern.

[17] Araya et al. (2022) exploit the staggered rollout of labels across different stores before the law went into effect, which allows them to study the effect of labels in categories in which all products would eventually receive a label after full deployment of the law.

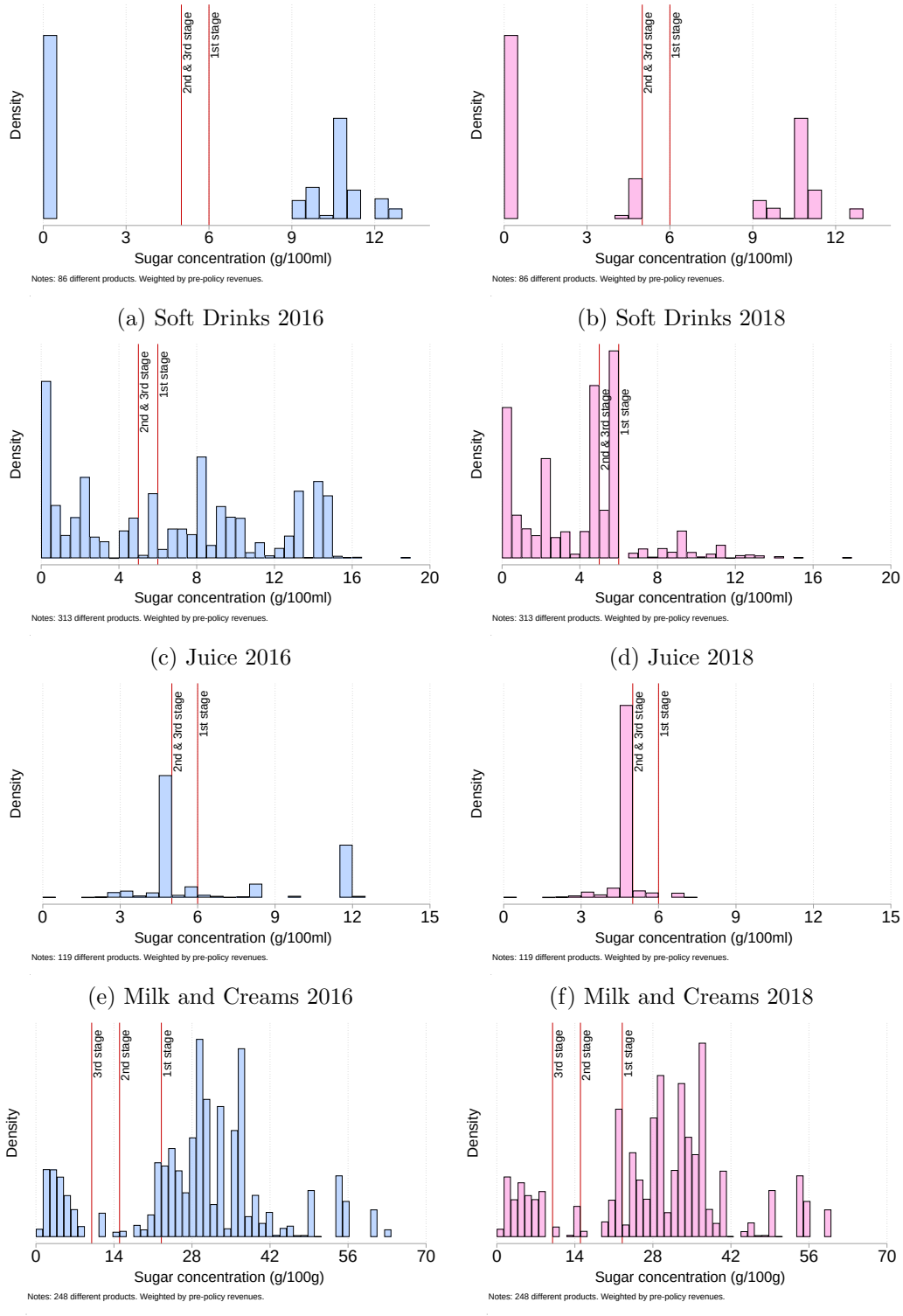
concentration was neither entirely below the labeling threshold nor so far above it that compliance was implausible. This restriction yields 13 categories that comprise 24.5% of total pre-policy revenue, and 52.6% of pre-policy revenue among products with pre-policy nutritional content above the regulatory threshold (see Appendix D for details).

To assess reformulation behavior, we compare the distribution of sugar concentration before and after the policy, with particular attention to bunching around regulatory thresholds. To discipline the choice of categories shown in the main text, we present the four largest categories in our reformulation sample by pre-policy revenue—soft drinks, juice, milk and creams, and cookies (Appendix Table D.1, column 1). Figure 7 depicts sugar concentration distributions in 2016 (pre-policy) and 2018 (post-policy), where bar height reflects pre-policy product revenue. Vertical lines indicate the policy thresholds in each stage of implementation.

Before the labeling policy, the distributions of sugar content in these categories showed no clear concentration around the regulatory thresholds. By 2018, the categories display starkly different responses. In the juice category, while about half of the products in 2016 had a sugar concentration per 100 ml exceeding the first-stage threshold, the distribution shifted leftward, with most products avoiding the first-stage labels. In contrast, reformulation is more limited in soft drinks, where the distribution is bimodal and most sugary products have sugar-free counterparts, and in cookies, where most products remained above the threshold. In milk and creams, products bunch at the most stringent threshold, 5 grams of sugar per 100 ml. The reason is that this threshold was set to coincide with the natural sugar content in whole milk.

In Figure 8, we summarize the findings for all categories by plotting the (weighted by pre-policy revenue) share of the products in each category that surpassed the sugar threshold in the pre-policy period and were reformulated to fall below the threshold in the post-policy period. We show histograms for each category in Appendix A, Figures A.8 and A.9.

We find substantial heterogeneity across categories: in some, such as yogurt, nearly all products were reformulated to fall below the regulatory threshold, while in others, like cookies, fewer than 10% were reformulated. One potential concern with reformulation is that firms reformulating to avoid a warning label might improve one nutrient



(a) Soft Drinks 2016 (b) Soft Drinks 2018

(c) Juice 2016 (d) Juice 2018

(e) Milk and Creams 2016 (f) Milk and Creams 2018

(g) Cookies 2016 (h) Cookies 2018

Figure 7: Distribution of sugar content pre- and post-policy in selected categories

Notes: This figure plots the distribution of sugar concentration before and after policy implementation for the four largest categories in the reformulation sample by pre-policy revenue (Appendix Table D.1, column 1). Each observation corresponds to a product and is weighted by its pre-policy revenue. Appendix Figures A.8 and A.9 present the remaining categories.

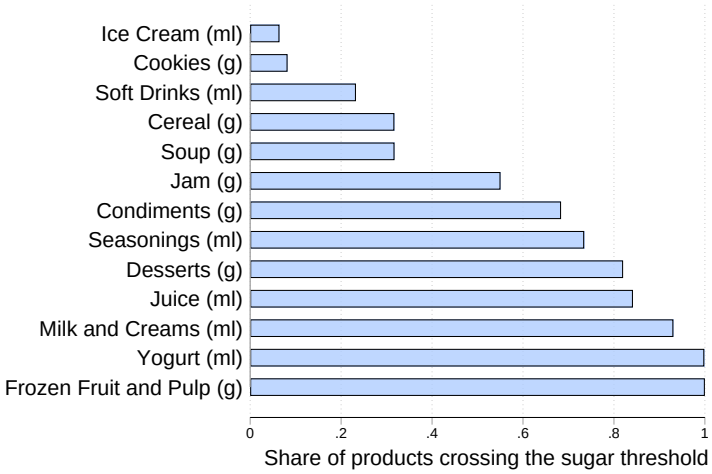


Figure 8: Share of products crossing the sugar threshold due to reformulation

Notes: This figure summarizes the findings from Appendix A, Figures A.8 and A.9. It shows the pre-policy revenue-weighted share of products that were above the sugar threshold before the policy and reduced their sugar concentration below the regulatory threshold after the policy. The products used represent 53% of the pre-policy revenue of all products to the right of the policy threshold in the pre-policy period. We provide more details on sample selection in Appendix D.

while worsening others. Appendix A, Panel B of Table A.3 shows this was not the case: reformulation typically improved several regulated nutrients simultaneously, with little evidence of trade-offs.[18]

3.2.1. *Long-run supply-side responses*: To study product reformulation following the implementation of the most stringent policy threshold, we analyze nutritional information from 2021. These data are merged at the UPC level with data from 2016 (pre-policy) and 2018 (post-first stage). We focus on products (UPCs) for which nutritional information is available for all three periods, covering 65% of the pre-policy revenue of solid products and 73% of the pre-policy revenue of liquid products.

Panels (a), (c), and (e) from Figure 9 present the results for solids. Panel (a) presents the pre-policy distribution of sugar content for solid products. Panel (c) shows that, consistent with our findings across individual categories, some products that were just to the right of the threshold before the policy tend to bunch at the first-stage threshold afterward. Panel (e) reveals significant bunching at the third-stage threshold in 2021.

[18]We focus our analysis on regulated nutrients. Firms may replace regulated nutrients with substitutes that we do not observe in our data.

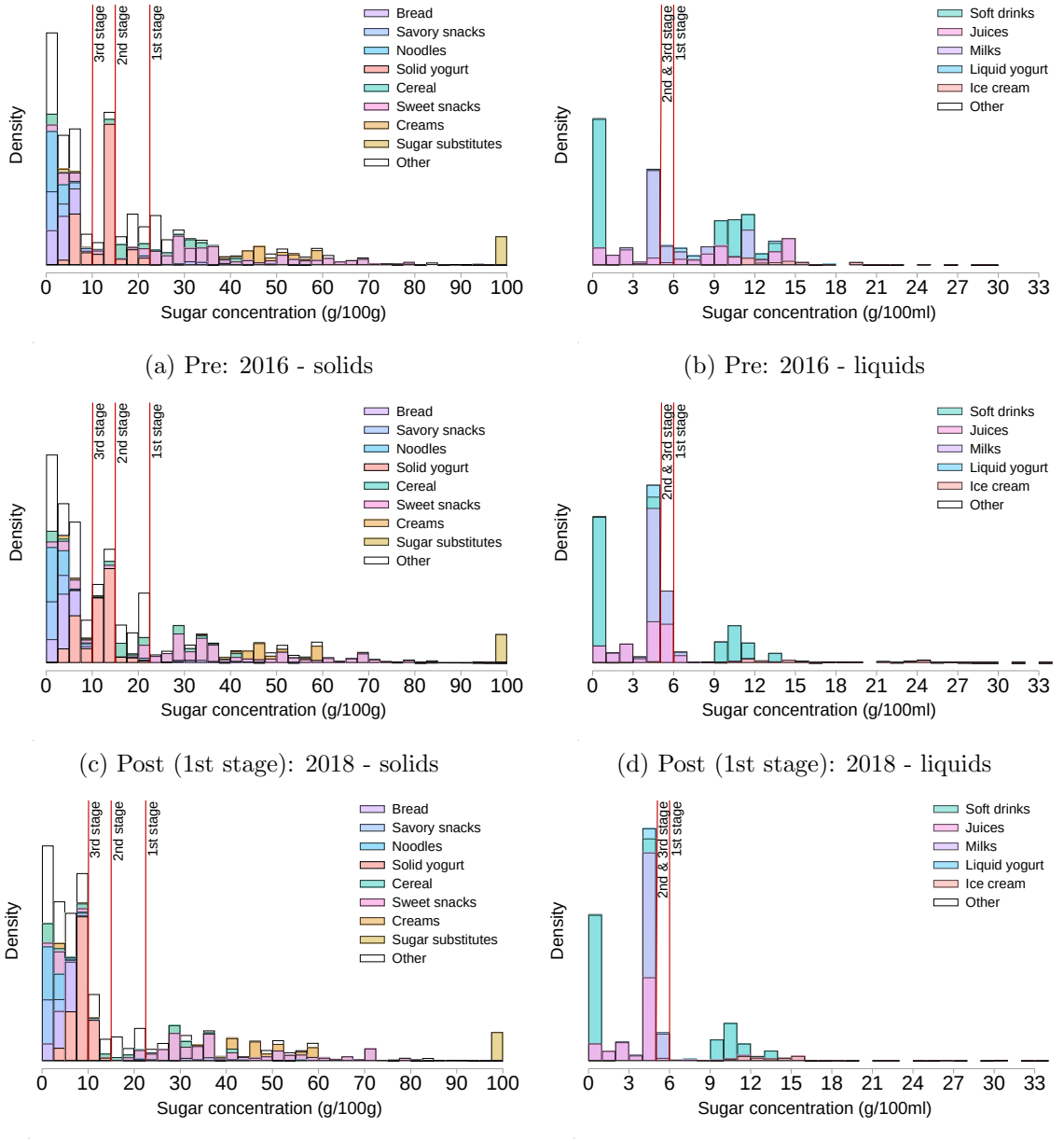
Notably, a substantial portion of this bunching is driven by yogurts, which are sometimes classified as solids rather than liquids, likely making them easier to reformulate compared to other solid products.

Panels (b), (d), and (f) from Figure 9 present the results for liquids. Once the first stage is enacted, the majority of the product distribution bunches to the left of either the first or second stage threshold. Panel (f) shows that after the third stage is implemented, most products bunch at that threshold. Notably, some soft drinks remain unchanged in formulation. This is likely because many of these products have sugar-free alternatives (e.g., *Coke* and *Diet Coke*).

Overall, Figure 9 indicates substantial reformulation among liquid products, whereas responses among solids are more modest and largely concentrated in yogurt.

Three important features of a product category can affect the extent to which products are reformulated. First, firm responses may depend on the expected impact of labels on product demand. In categories with close substitutes and in which labels can provide more information, the returns to reformulation are higher. Second, reformulation is a function of the distance between the products' current nutritional content and the regulatory threshold.[19] Third, firms are more likely to reformulate products when they are able to do so without substantially affecting their quality (e.g., taste). For example, in categories such as yogurt or juice, firms can reformulate their products by substituting sugar with other low-cost sweeteners that mimic the products' taste. On the other hand, sugar serves as a bulking agent in cereal or cookies, and replacing it with low-cost sweeteners may cause them to crumble. Our results are consistent with this pattern, whereby categories with liquid products have a larger share of products that bunch.

[19]In Appendix A, Figure A.10, we document a negative correlation between the share of products reformulated from above the sugar threshold before the policy to below it afterward and the average pre-policy distance from the threshold among products initially above it.



(a) Pre: 2016 - solids

(b) Pre: 2016 - liquids

(c) Post (1st stage): 2018 - solids

(d) Post (1st stage): 2018 - liquids

(e) Post (3rd stage): 2021 - solids

(f) Post (3rd stage): 2021 - liquids

Figure 9: Distribution of sugar content pre- and post-policy

Notes: The figure shows the pre- and post-policy distribution of sugar concentration for all solid and liquid products (full bar) and selected categories (in colors). Panels (a) and (b) represent the pre-policy distributions, Panels (c) and (d) the post-policy distributions after the first stage of implementation, and Panels (e) and (f) the distributions after all stages were implemented. Product categories that are not covered by the law (e.g., fresh meat, fish, eggs, unpacked fruits and vegetables, beans, flour), in which most products do not contain any added sugar (e.g., butter, rice, frozen meals), or in which labels are mostly not required (e.g., cheese, sausages) are excluded. Only products with nutritional information for the three periods are included. The products included are 64.51% of the pre-policy revenue of solid products and 72.95% of the pre-policy revenue of liquid products, among the categories included in the exercise. Observations are weighted by the pre-policy mass of each product purchased by the consumer panel. Stages 1, 2, and 3 in Chile took place in June 2016, June 2018, and June 2019.

4. Policy implications

Our findings have direct implications for food-labeling policy design. Because labels induce little substitution across categories, their effects operate mainly within categories, through consumer substitution from labeled to unlabeled products and firm reformulation. The relevant demand margin is therefore substitution from, say, a labeled cereal to an unlabeled cereal, rather than from cereal to fruit.

A category's contribution to the aggregate effect of a labeling policy depends on three features. First, it depends on how much of the targeted nutrient consumers purchase in that category. Second, the demand response depends on whether labeled and unlabeled products are close substitutes and whether labels provide consumers with new information about product healthiness. Third, the supply response depends on products' distance from the threshold and whether firms can reformulate them without substantially compromising taste or increasing production costs.

These features vary substantially across categories. Most food labeling policies, including Chile's, apply nutrient thresholds that are common across product categories, often distinguishing only between solids and liquids. Under this design, moving a threshold changes which products are labeled and alters demand and reformulation responses across many heterogeneous categories simultaneously. Consequently, its effect on aggregate purchases of the targeted nutrient cannot be inferred from any single category or from demand and reformulation responses considered separately.

4.1. *Calibration under alternative policy thresholds*

We use an accounting calibration to bring these considerations together. We focus on sugar because it is a central nutrient targeted by the Chilean regulation and because our data allow us to combine category-level demand estimates with product-level reformulation around explicit thresholds for solids and liquids. For each counterfactual threshold, the calibration aggregates predicted demand and reformulation responses across categories, accounting for each category's contribution to total sugar purchases, and calculates predicted aggregate sugar purchases separately for solids and liquids.

The calibration starts from the pre-policy distribution of product quantities and sugar

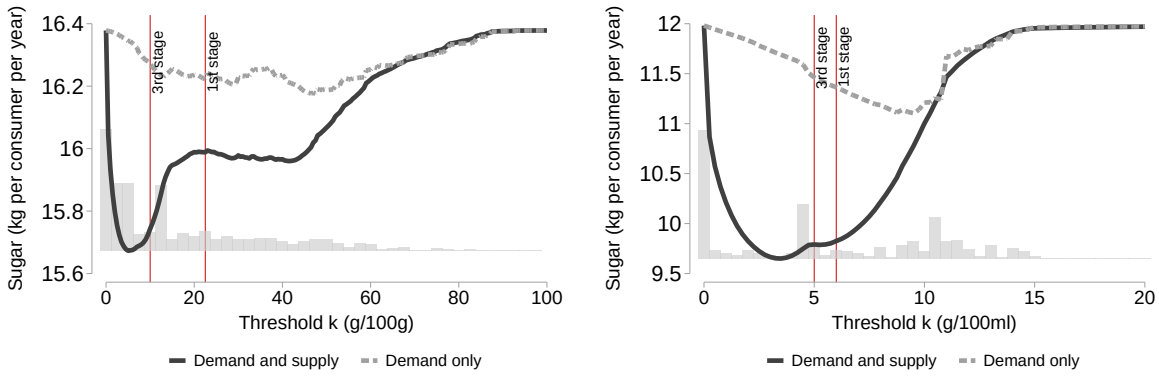
content. Because the regulation sets different thresholds for solids and liquids, we treat solid and liquid products within the same category as separate category-by-form cells. For each counterfactual threshold, we first predict each product's probability of reformulating to the cutoff. Second, we use these reformulation probabilities to calculate the post-reformulation labeled share within each cell. Third, we apply the estimated or imputed category-level demand response, reallocating a fixed quantity from labeled to unlabeled products within the cell.[20] Finally, we aggregate the resulting sugar purchases across cells. Throughout the exercise, we hold prices, total quantities within each cell, and all product characteristics other than sugar content fixed. Appendix E describes the procedure and its assumptions in detail.

Figure 10 traces predicted aggregate sugar purchases over a grid of counterfactual thresholds, separately for solids and liquids. The solid line allows for both demand substitution and product reformulation, while the dashed line shuts down reformulation and includes only the demand response. The vertical lines mark the Chilean Stage 1 and Stage 3 thresholds, and the histogram shows the pre-policy distribution of sugar content. When reporting percentage reductions, we use total sugar purchases under the pre-policy nutritional content distribution as the baseline.

Three patterns stand out. First, predicted sugar purchases are non-monotonic in the threshold. Moving the threshold changes not only the number of labeled products and the availability of unlabeled substitutes within a given category, but also which categories are most affected by the labels. Second, incorporating reformulation produces substantially larger reductions than demand substitution alone, particularly for liquids. Third, both the magnitude of the predicted reductions and the thresholds at which they are largest differ between solids and liquids, reflecting differences in their sugar distributions and in the feasibility of reducing sugar without substantially changing product quality.

For solids, the lowest predicted level of sugar purchases occurs at a threshold of about 5 grams per 100 grams, below Chile's Stage 3 threshold of 10. At this threshold, predicted sugar purchases are 4.3% below the baseline, with most of the reduction coming

[20] The demand estimates are based on whether a product received any warning label. In the calibration, we apply these responses to changes in sugar-label status while holding the thresholds for other nutrients fixed.



(a) Solids

(b) Liquids

Figure 10: Predicted sugar purchases under alternative sugar thresholds

Notes: The figure plots predicted total sugar purchases under alternative sugar thresholds for solids and liquids. The solid line combines supply-side reformulation and demand-side substitution; the dashed line uses only demand-side responses. The vertical red lines mark the Chilean sugar thresholds: for solids, the Stage 1 cutoff of 22.5 and the Stage 3 cutoff of 10 grams per 100 grams; for liquids, the Stage 1 cutoff of 6.0 and the Stage 3 cutoff of 5.0 grams per 100 milliliters. The gray histogram in the background shows the mass-weighted distribution of pre-policy sugar content across packaged products. Because the labeling policy applies only to packaged goods, the exercise is restricted to packaged products. Within the category-by-form cells that enter the calibration, the products for which we observe the nutritional information required for the exercise represent 57.58% of pre-policy packaged solid revenue and 82.01% of pre-policy packaged liquid revenue.

from reformulation. Yogurt accounts for the largest share of the decline, followed by ready-to-eat desserts, breakfast cereals, cookies, ice cream, and condiments. The solid-products curve also has a second, shallower local minimum at approximately 42 grams per 100 grams, where predicted purchases are 2.6% below the baseline. This second reduction is driven by a different set of categories, principally condensed milk, cream substitutes, jams, and chocolate. The two minima illustrate how a common threshold can generate distinct aggregate responses as it moves through the sugar distributions of heterogeneous categories.

For liquids, the lowest predicted level of sugar purchases occurs at a threshold of approximately 3.50 grams per 100 milliliters, below Chile's Stage 3 threshold of 5. At this threshold, predicted purchases are 19.5% below the baseline, a substantially larger reduction than for solids. Most of the reduction is attributable to reformulation. Juice makes the largest contribution, followed by liquid dairy products, cream substitutes, and sugar-sweetened soft drinks. For soft drinks, both reformulation and substitution away from labeled products contribute to the decline. The larger predicted response for liquids

is consistent with sugar being easier to replace in many beverages without substantially changing their taste.

Taken together, these results show that the effect of a common threshold depends on how it interacts with the distribution of sugar purchases and heterogeneous demand and reformulation responses across categories. The thresholds identified above minimize predicted sugar purchases within the calibration and under its maintained assumptions; they should not be interpreted as welfare-maximizing policy thresholds.

4.2. *Other dimensions of label design*

The calibration varies the nutrient threshold while holding other features of the Chilean labeling system fixed. We now discuss three additional design choices: whether cutoffs vary across categories, how many nutrient levels they distinguish, and whether they report individual nutrients or summarize overall nutritional quality.

First, cutoffs may be common across categories or tailored to specific sets of categories. The Chilean policy, for example, applies one threshold to solids and another to liquids. Category-specific thresholds could better accommodate the heterogeneous demand and reformulation responses we document across product categories. Chilean legislators initially considered this approach but rejected it because it could complicate consumers' interpretation of the labels, require regulators to define category boundaries, and obscure differences in absolute nutrient content across categories. Whether the potential gains justify these costs remains an open question.

Second, labels can use a binary warning—as in the Chilean system—or distinguish multiple levels of nutrient content. Under a single cutoff, categories representing 57% of spending are almost entirely labeled or entirely unlabeled, leaving little scope for substitution within those categories. Moreover, products far above the cutoff have weaker incentives to reformulate, as reformulation declines with distance from the threshold (Appendix A, Figure A.10). Additional levels could preserve distinctions among products in these categories and provide firms with more attainable reformulation targets. Ecuador's traffic-light labels and Singapore's Nutri-Grade illustrate this graded approach. Its potential disadvantage is that additional levels may be harder to interpret and may generate

weaker responses than a binary warning (Ravaioli, 2021).

Third, labels may report whether a product is high in particular nutrients or provide a summary measure of its overall nutritional quality. Chile's nutrient-specific warnings do not rank products by their overall healthiness. A common summary score could facilitate comparisons across categories and potentially encourage substitution between them. Constructing such a score, however, requires regulators to decide how different nutrients and other product attributes are weighted. Because the Chilean policy induced little substitution across categories, our estimates do not reveal how consumers would respond to such a design.

4.3. *Complementary policies*

Changes in the scope, granularity, or content of labels may make them more informative, but they do not address every reason consumers purchase products high in critical nutrients. Labels are likely to have limited effects when they convey little new information—for example, if all chocolates and candy get labeled. They may also be ineffective if consumers understand the nutritional consequences of their choices but, because of self-control problems or time inconsistency, make decisions that conflict with their long-run objectives (Sadoff et al., 2020; Samek, 2019). In these cases, other instruments that operate through different channels can be more effective. Sugar taxes, for example, directly change relative prices and can reduce purchases even when consumers are already informed (Barahona et al., 2023). Taxes can therefore complement labels in settings or categories where providing information alone does not substantially change behavior.

5. Conclusions

We study how the Chilean Food Act affected purchases across the full range of food and beverage categories. Following the introduction of front-of-package warning labels, consumers purchased substantially less sugar and fewer calories per dollar spent. These aggregate reductions reflect changes in both consumer choices and the nutritional content of the products offered by firms.

On the demand side, we find little evidence that consumers shifted purchases across

product categories. Instead, they substituted between labeled and unlabeled products within categories. The magnitude of this response varies considerably and is larger when labels provide consumers with new information about products' nutritional content. Labels therefore appear to be most effective when consumers have suitable alternatives within a category and do not already know which products are high in critical nutrients.

On the supply side, firms reformulated many products to avoid warning labels, improving nutritional content even for consumers who did not change their choices. Reformulation was also highly heterogeneous: it was more common when products were close to the threshold and when firms could reduce nutrient content without substantially changing their taste. The tighter thresholds introduced in later stages generated additional reformulation, primarily in categories where such changes were feasible.

Together, these findings show that the effects of food labels depend on the markets to which they are applied. Labels are most effective in categories where they provide new information, labeled and unlabeled products are close substitutes, and reformulation is feasible. Because these conditions differ substantially across categories, evaluating a labeling policy requires looking beyond its average effect and understanding the demand- and supply-side responses that produce it.

References

Aguilar, A., E. Gutierrez, and E. Seira (2021). The effectiveness of sin food taxes: evidence from Mexico. *Journal of Health Economics 77*, 102455.

Alé-Chilet, J. and S. Moshary (2022). Beyond consumer switching: Supply responses to food packaging and advertising regulations. *Marketing Science 41*(2), 243–270.

Allais, O., F. Etilé, and S. Lecocq (2015). Mandatory labels, taxes and market forces: An empirical evaluation of fat policies. *Journal of Health Economics 43*, 27–44.

Allcott, H., D. Cohen, W. Morrison, and D. Taubinsky (2025). When do "nudges" increase welfare? *American Economic Review 115*(5), 1555–1596.

Allcott, H., R. Diamond, J.-P. Dubé, J. Handbury, I. Rahkovsky, and M. Schnell (2019).

Food deserts and the causes of nutritional inequality. *The Quarterly Journal of Economics 134*(4), 1793–1844.

Allcott, H., B. B. Lockwood, and D. Taubinsky (2019a). Regressive sin taxes, with an application to the optimal soda tax. *The Quarterly Journal of Economics 134*(3), 1557–1626.

Allcott, H., B. B. Lockwood, and D. Taubinsky (2019b). Should we tax sugar-sweetened beverages? An Overview of theory and evidence. *Journal of Economic Perspectives 33*(3), 202–27.

Araya, S., A. Elberg, C. Noton, and D. Schwartz (2022). Identifying food labeling effects on consumer behavior. *Marketing Science 41*(5), 982–1003.

Avishay-Rizi, O. and O. Reshef (2023). Firms' Anticipation and Adaptation to Changing Environments: Evidence from Nutritional Regulation. Mimeo.

Barahona, N., C. Otero, and S. Otero (2023). Equilibrium effects of food labeling policies. *Econometrica 91*(3), 839–868.

Bernheim, B. D. and D. Taubinsky (2018). Behavioral public economics. *In: Handbook of Behavioral Economics—Foundations and Applications 1*(5), 381–516.

Bix, L., R. P. Sundar, N. M. Bello, C. Peltier, L. J. Weatherspoon, and M. W. Becker (2015). To see or not to see: do front of pack nutrition labels affect attention to overall nutrition information? *PLOS ONE 10*(10), e0139732.

Bollinger, B., P. Leslie, and A. Sorensen (2011). Calorie Posting in Chain Restaurants. *American Economic Journal: Economic Policy 3*, 91–128.

Briesch, R. A., W. R. Dillon, and E. J. Fox (2013). Category Positioning and Store Choice: The Role of Destination Categories. *Marketing Science 32*(3), 488–509.

Cawley, J. (2015). An economy of scales: A selective review of obesity's economic causes, consequences, and solutions. *Journal of Health Economics 43*, 244–268.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics 1*, 451–488.

Crosetto, P., A. Lacroix, L. Muller, and B. Ruffieux (2020). Nutritional and economic impact of five alternative front-of-pack nutritional labels: experimental evidence. *European Review of Agricultural Economics 47*(2), 785–818.

Crosetto, P., L. Muller, and B. Ruffieux (2016). Helping consumers with a front-of-pack label: Numbers or colors?: Experimental comparison between guideline daily amount and traffic light in a diet-building exercise. *Journal of Economic Psychology 55*, 30–50.

Dranove, D. and G. Z. Jin (2010, December). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature 48*(4), 935–63.

Dubois, P., P. Albuquerque, O. Allais, C. Bonnet, P. Bertail, P. Combris, S. Lahlou, N. Rigal, B. Ruffieux, and P. Chandon (2021). Effects of front-of-pack labels on the nutritional quality of supermarket food purchases: evidence from a large-scale randomized controlled trial. *Journal of the Academy of Marketing Science 49*(1), 119–138.

Dubois, P., R. Griffith, and M. O'Connell (2017, 04). The Effects of Banning Advertising in Junk Food Markets. *The Review of Economic Studies 85*(1), 396–436.

Dubois, P., R. Griffith, and M. O'Connell (2020, November). How well targeted are soda taxes? *American Economic Review 110*(11), 3661–3704.

Falbe, J., N. Rojas, A. H. Grummon, and K. A. Madsen (2015). Higher retail prices of sugar-sweetened beverages 3 months after implementation of an excise tax in Berkeley, California. *American Journal of Public Health 105*(11), 2194–2201.

Falbe, J., H. R. Thompson, C. M. Becker, N. Rojas, C. E. McCulloch, and K. A. Madsen (2016). Impact of the Berkeley excise tax on sugar-sweetened beverage consumption. *American Journal of Public Health 106*(10), 1865–1871.

Finkelstein, E. A., K. L. Strombotne, N. L. Chan, and J. Krieger (2011). Mandatory Menu Labeling in One Fast-Food Chain in King County, Washington. *American Journal of Preventive Medicine 40*(2), 122–127.

Griffith, R., M. O'Connell, and K. Smith (2017). The importance of product reformulation versus consumer choice in improving diet quality. *Economica 84*(333), 34–53.

Historia de la Ley 20.606 (2011). Sobre composición nutricional de los alimentos y su publicidad. http://www.bcn.cl/obtienearchivo?id=recursoslegales/10221.3/37370/1/HL20606.pdf.

Hut, S. and E. Oster (2022). Changes in Household Diet: Determinants and Predictability. *Journal of Public Economics 208*, 104620.

Kan, C., Y. L. Liu, D. R. Lichtenstein, and C. Janiszewski (2023). The Negative and Positive Consequences of Placing Nonpromoted Products Next to Promoted Products. *Journal of Marketing 87*(6), 928–948.

Kiesel, K. and S. B. Villas-Boas (2013). Can information costs affect consumer choice? Nutritional labels in a supermarket experiment. *International Journal of Industrial Organization 31*(2), 153–163.

Kluser, F. and M. Pons (2026). The Apple Does Not Fall Far From the Tree: Intergenerational Persistence of Dietary Habits. *Review of Economics and Statistics*. Forthcoming.

Lee, M. M., J. Falbe, D. Schillinger, S. Basu, C. E. McCulloch, and K. A. Madsen (2019). Sugar-sweetened beverage consumption 3 years after the Berkeley, California, sugar-sweetened beverage tax. *American Journal of Public Health 109*(4), 637–639.

Leeflang, P. S. H. and J. Parreño-Selva (2012). Cross-category Demand Effects of Price Promotions. *Journal of the Academy of Marketing Science 40*(4), 572–586.

Lim, J. H., R. Rishika, R. Janakiraman, and P. Kannan (2020). Competitive effects of front-of-package nutrition labeling adoption on nutritional quality: Evidence from Facts Up Front–style labels. *Journal of Marketing 84*(6), 3–21.

Mathios, A. D. (2000). The impact of mandatory disclosure laws on product choices: An analysis of the salad dressing market. *The Journal of Law and Economics 43*(2), 651–678.

Moorman, C., R. Ferraro, and J. Huber (2012). Unintended Nutrition Consequences: Firm Responses to the Nutrition Labeling and Education Act. *Marketing Science 31*(5), 717–737.

NCHS (2018). National health and nutrition examination survey data, national center for health statistics. *U.S. Department of Health and Human Services, Centers for Disease Control and Prevention*.

Ono, M. and A. Ono (2015). Impacts of the FoSHU (Food for Specified Health Uses) system on food evaluations in Japan. *Journal of Consumer Marketing 32*(7), 542–550.

Pachali, M. J., M. J. Kotschedoff, A. van Lin, B. J. Bronnenberg, and E. van Herpen (2023). How do nutritional warning labels affect prices? *Journal of Marketing Research 60*(1), 92–109.

Paraje, G., N. Valdés, A. V. Macaya, C. Corvalán, and B. Popkin (2026). The Impact of Chile's Multipronged Food Labelling and Advertising Law on Early Childhood Excess Weight: A Cohort Difference-in-Differences Study. *The Lancet 407*(10548), 2630–2640.

Quintiliano Scarpelli, D., A. C. Pinheiro Fernandes, L. Rodriguez Osiac, and T. Pizarro Quevedo (2020). Changes in nutrient declaration after the food labeling and advertising law in Chile: a longitudinal approach. *Nutrients 12*(8), 2371.

Ravaioli, S. (2021). Coarse and Precise Information in Food Labeling. Mimeo.

Reyes, M., L. Smith Taillie, B. Popkin, R. Kanter, S. Vandevijvere, and C. Corvalán (2020). Changes in the amount of nutrient of packaged foods and beverages after the initial implementation of the Chilean Law of Food Labelling and Advertising: A nonexperimental prospective study. *PLOS Medicine 17*(7), e1003220.

Sadoff, S., A. Samek, and C. Sprenger (2020). Dynamic inconsistency in food choice: Experimental evidence from two food deserts. *The Review of Economic Studies 87*(4), 1954–1988.

Samek, A. (2019). Gifts and goals: Behavioral nudges to improve child food choice at school. *Journal of Economic Behavior & Organization 164*, 1–12.

Shangguan, S., A. Afshin, M. Shulkin, W. Ma, D. Marsden, J. Smith, M. Saheb-Kashaf, P. Shi, R. Micha, F. Imamura, et al. (2019). A meta-analysis of food labeling effects on consumer diet behaviors and industry practices. *American Journal of Preventive Medicine 56*(2), 300–314.

Silver, L. D., S. W. Ng, S. Ryan-Ibarra, L. S. Taillie, M. Induni, D. R. Miles, J. M. Poti, and B. M. Popkin (2017). Changes in prices, sales, consumer spending, and beverage consumption one year after a tax on sugar-sweetened beverages in Berkeley, California, US: A before-and-after study. *PLOS Medicine 14*(4), e1002283.

Taillie, L. S., M. Reyes, M. A. Colchero, B. Popkin, and C. Corvalán (2020). An evaluation of Chile's law of food labeling and advertising on sugar-sweetened beverage purchases from 2015 to 2017: A before-and-after study. *PLOS Medicine 17*(2), e1003015.

Taylor, R., S. Kaplan, S. B. Villas-Boas, and K. Jung (2019). Soda wars: Effect of a soda tax election on soda purchases. *Economic Inquiry 57*(3), 1480–1496.

Temple, N. J. (2020). Front-of-package food labels: A narrative review. *Appetite 144*, 104485.

The Washington Post (2024). The New Food Warning Labels: How the FDA is Helping Americans Make Healthier Choices. Accessed: 2024-09-15.

Unnevehr, L. J. and E. Jagmanaite (2008). Getting rid of trans fats in the US diet: Policies, incentives and progress. *Food Policy 33*(6), 497–503.

WHO (2018). Factsheet No. 311. *World Health Organization*.

Wisdom, J., J. S. Downs, and G. Loewenstein (2010). Promoting Healthy Choices: Information versus Convenience. *American Economic Journal: Applied Economics 2*(2), 164–178.

Zhu, C., R. A. Lopez, and X. Liu (2015). Information Cost and Consumer Choices of Healthy Foods. *American Journal of Agricultural Economics 98*(1), 41–53.

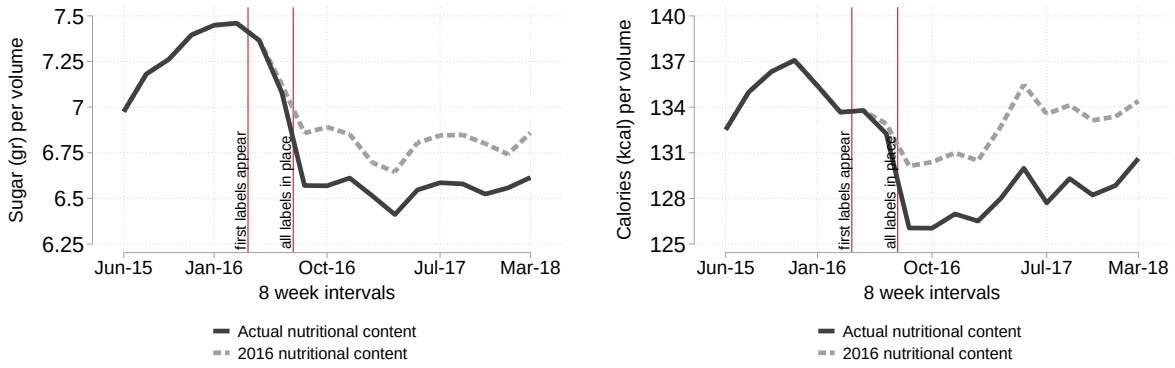
Online Appendix for:

Food Labeling Policies: Aggregate Impacts and Heterogeneity Across Categories

Nano Barahona Cristóbal Otero Sebastián Otero Joshua Kim

September 1, 2026

Appendix A: Additional Figures and Tables

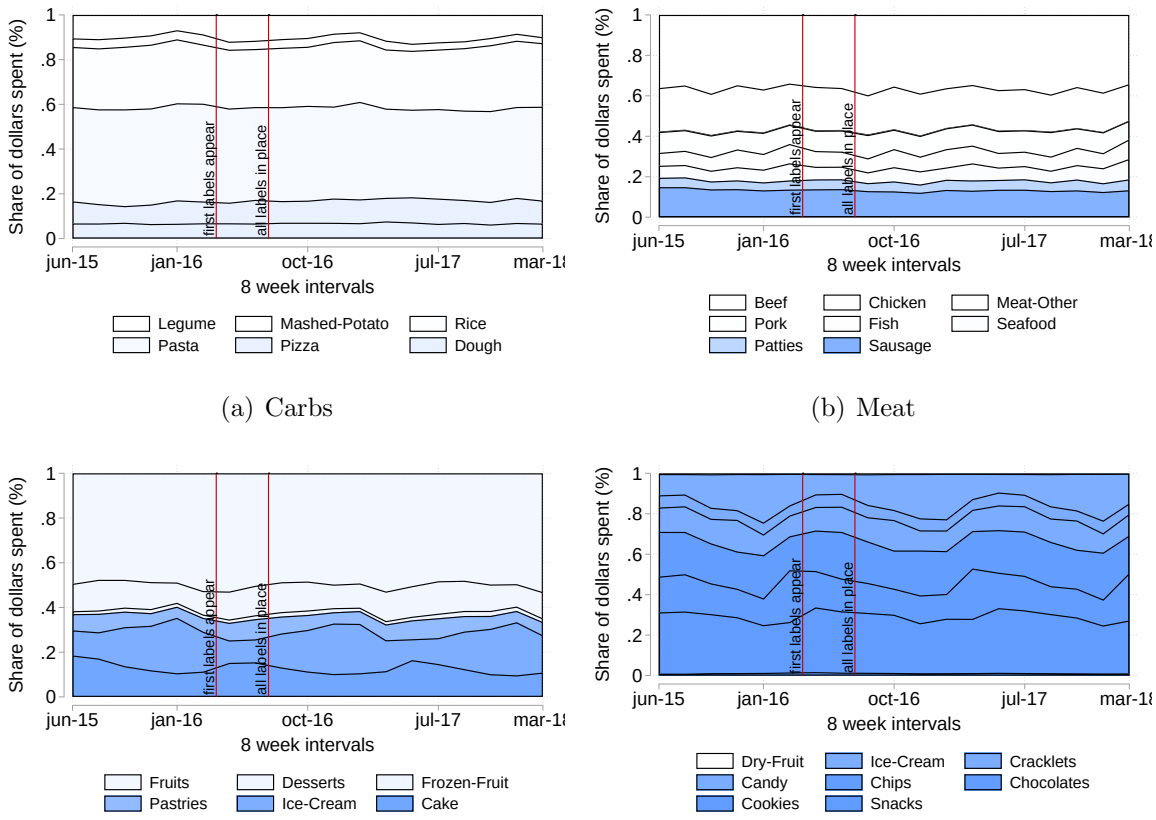


(a) Sugar intake

(b) Calorie intake

Figure A.1: Nutritional intake per gram or milliliter of product purchased before and after the policy

Notes: This figure shows changes in nutritional intake per unit of mass or volume of food products purchased at Walmart. For mass and volume, we calculate the total kilograms and liters of products purchased at Walmart. We then divide total sugar intake by the total mass or volume of products. Measures of product mass and volume are subject to measurement error from potential coding errors in package sizes.



(a) Carbs

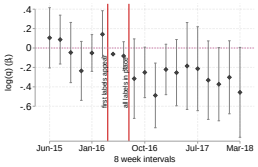
(b) Meat

(c) Desserts

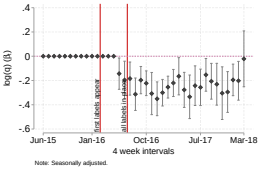
(d) Snacks

Figure A.2: Share of dollars spent across different categories

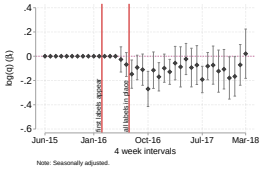
Notes: This figure shows the evolution of the share of dollars spent in each category within broader groups of products. Colors represent the share of labeled products within each category. White areas are categories in which no product received a label, and dark-blue areas (e.g., snacks) are categories in which all products received at least one label. We show that there are no differential changes in dollars spent between low-in-labels and high-in-labels categories.



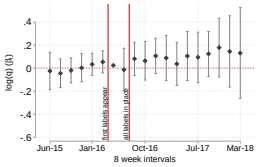
(a) Bread Toast



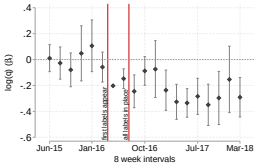
(b) Bread



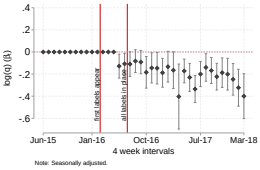
(c) Ice Creams



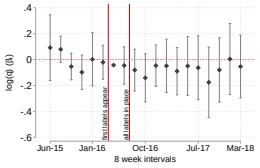
(d) Burgers



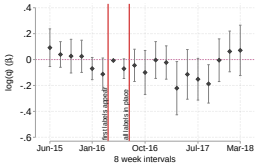
(e) Bakery



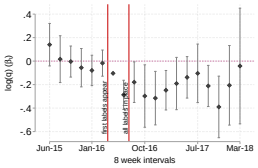
(f) Desserts and Cakes



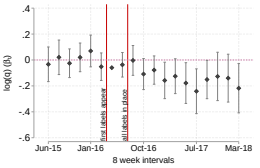
(g) Jam (jar), Syrup and Honey



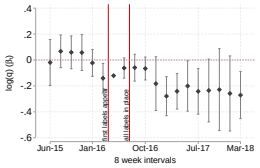
(h) Italian Dough and Pasta



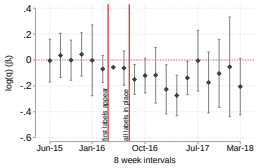
(i) Inst. Noodles and Rice



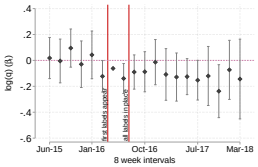
(j) Dressings, Oil, Sauces and Seasonings



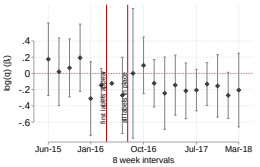
(k) Sugar and Substitutes



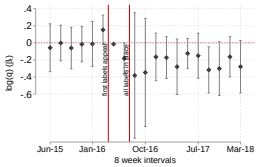
(l) Cereal bars, Diet and Health



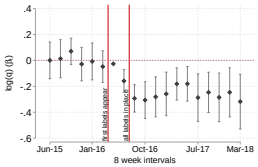
(m) Soft and Energy Drinks



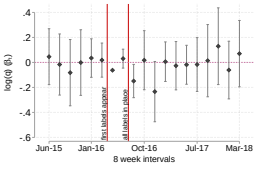
(n) Pork Ham



(o) Int. Cuisine: Mexican and Peruvian



(p) Cereal



(q) Candy - Checkout

Figure A.3: Estimates by category

Notes: This figure presents regression coefficient estimates. Each panel presents the β_k coefficients from Equation (3) for selected categories.

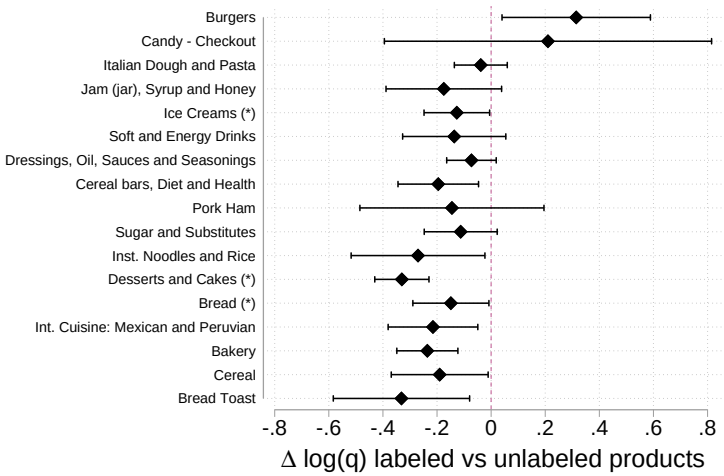


Figure A.4: Changes in demand for selected categories (without controlling for prices)

Notes: This figure shows the changes in equilibrium quantities between labeled and unlabeled products from estimating Equation (2) for different product categories after fixing $\gamma = 0$. The set of products in this analysis represents 34.16% of the pre-policy revenue of all products in the sample. We provide more details on sample selection in Appendix B.

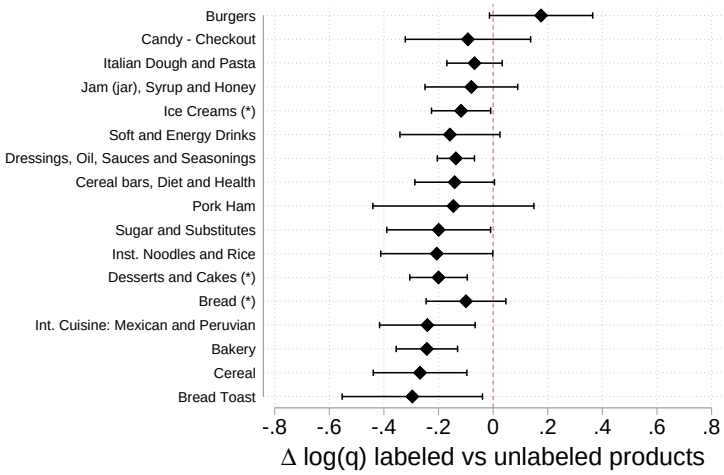


Figure A.5: Changes in demand for selected categories (price index built from fixed pre-policy sales-volume weights)

Notes: This figure re-estimates Equation (2) replacing the baseline price index—which aggregates UPC-store-period unit values using contemporaneous sales-volume weights—with one that holds the UPC weights fixed at their pre-policy sales volumes. UPCs with no pre-policy sales receive zero weight. The set of products in this analysis represents 34.16% of the pre-policy revenue of all products in the sample. We provide more details on sample selection in Appendix B.

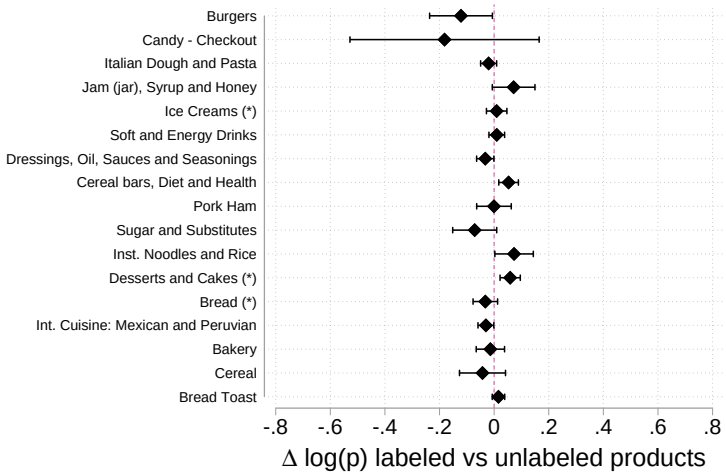
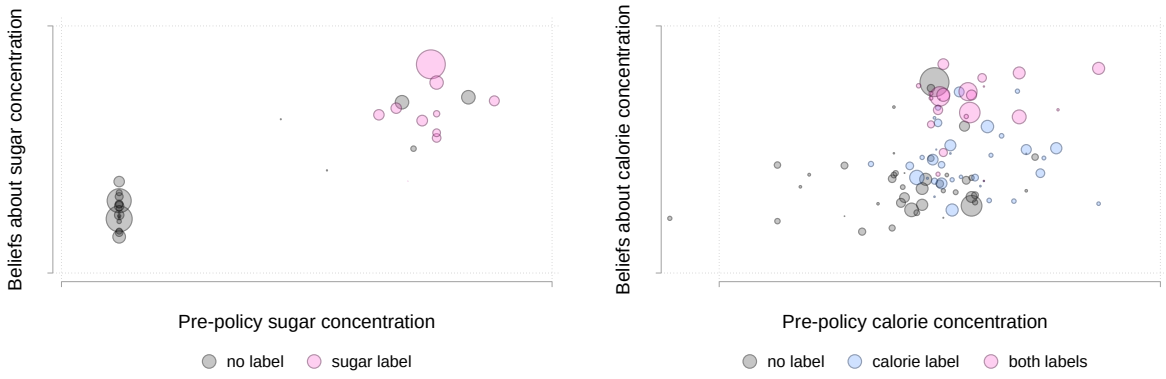


Figure A.6: Changes in prices for selected categories

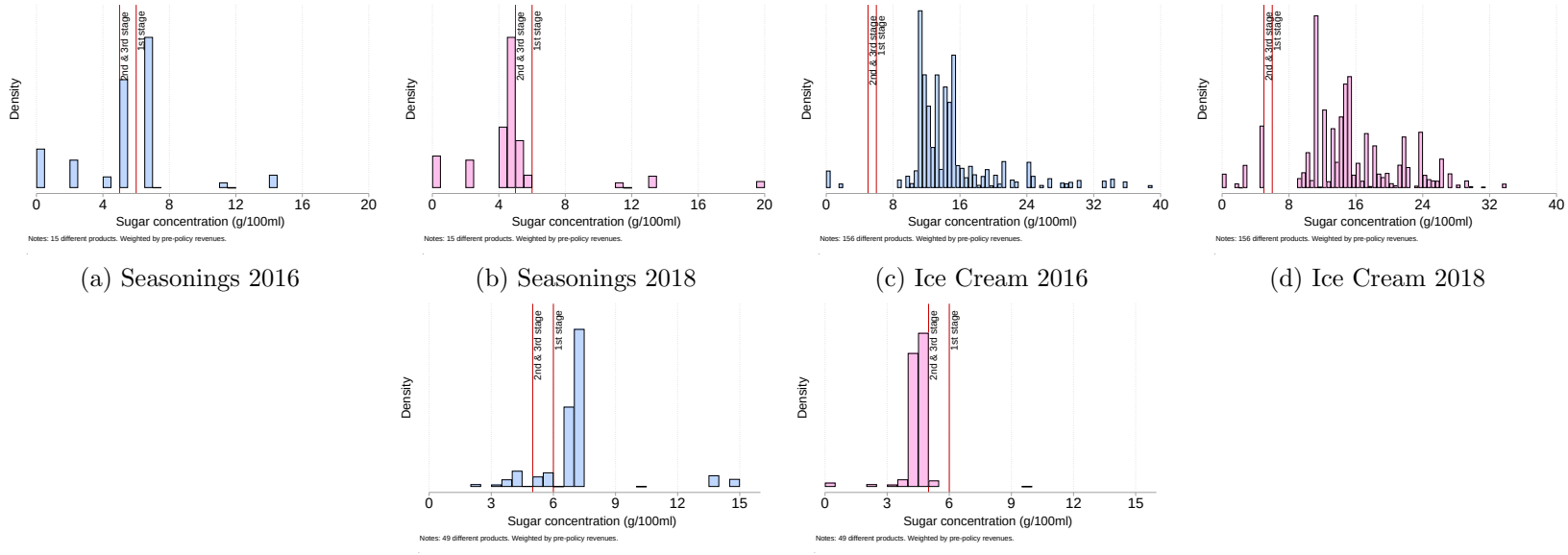
Notes: This figure shows the changes in equilibrium prices between labeled and unlabeled products by estimating Equation (2) for different product categories, replacing the left-hand side variable with log prices (and omitting the right-hand side price control). The set of products in this analysis represents 34.16% of the pre-policy revenue of all products in the sample. We provide more details on sample selection in Appendix B.



(a) Soft Drinks (b) Cereals

Figure A.7: Correlation between beliefs about nutritional content and true nutritional content

Notes: This figure shows the first moments of beliefs about each product's nutritional content vs. its real nutritional content. Each circle corresponds to a different product, and its size represents the total revenue from that product in our sample period. Panel (a) focuses on the sugar concentration of soft drinks as measured by g sugar/g product and panel (b) on the caloric concentration of cereals, as measured by kcal/g product. Since we focus on the relative distance between survey responses for different products, we do not provide numerical labels for the x- and y-axes.

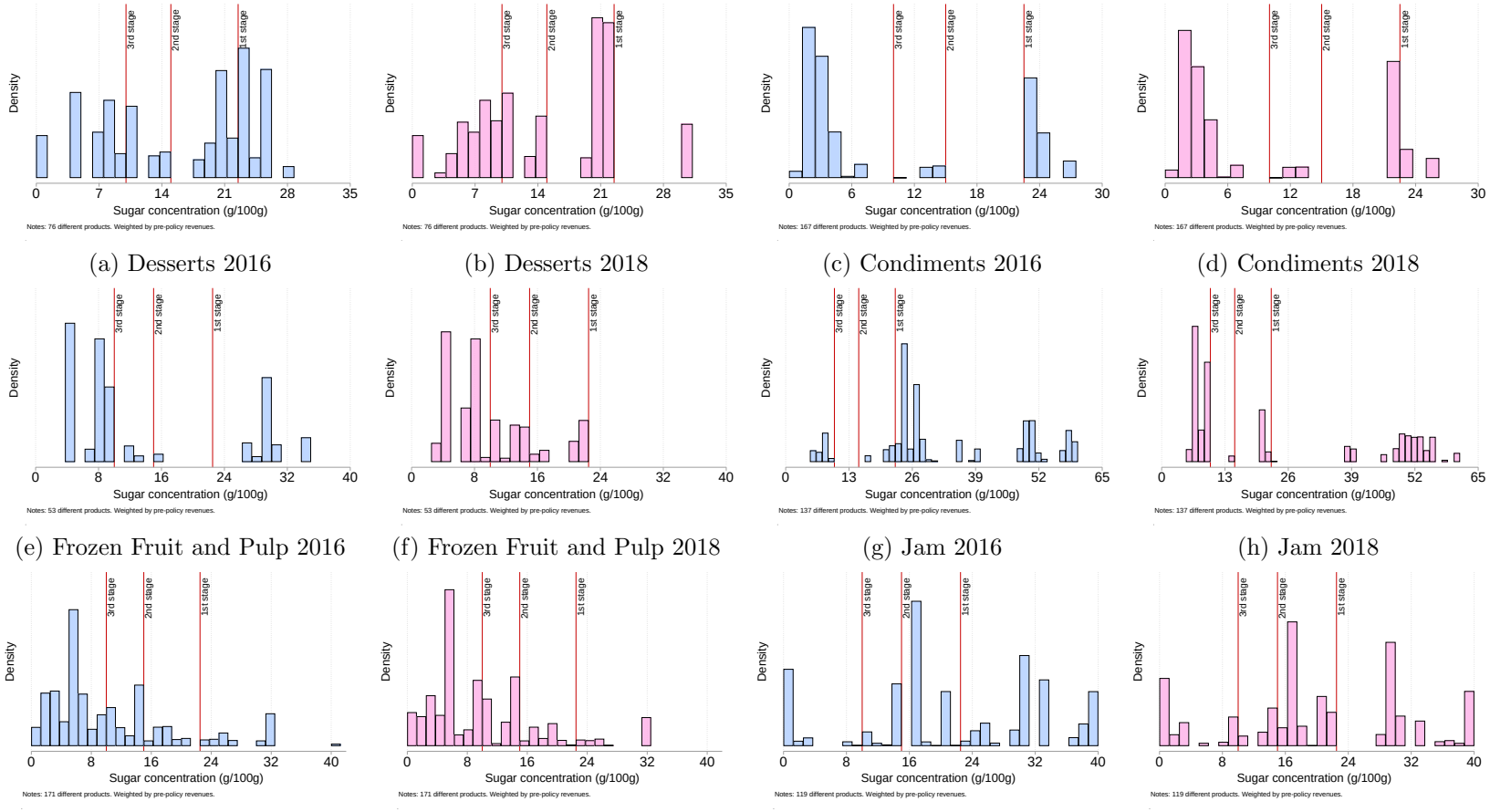


(a) Seasonings 2016 (b) Seasonings 2018 (c) Ice Cream 2016 (d) Ice Cream 2018

(e) Yogurt 2016 (f) Yogurt 2018

Figure A.8: Distribution of sugar content pre- and post-policy for liquids in selected categories

Notes: This figure plots the distribution of sugar concentration for liquid products in different categories before and after the policy implementation. Observations are weighted by pre-policy revenue.



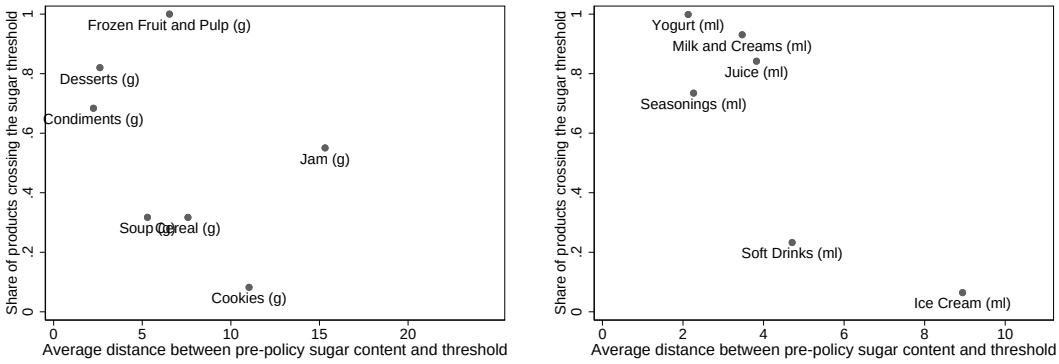
(a) Desserts 2016 (b) Desserts 2018 (c) Condiments 2016 (d) Condiments 2018

(e) Frozen Fruit and Pulp 2016 (f) Frozen Fruit and Pulp 2018 (g) Jam 2016 (h) Jam 2018

(i) Soup 2016 (j) Soup 2018 (k) Cereal 2016 (l) Cereal 2018

Figure A.9: Distribution of sugar content pre- and post-policy for solids in selected categories

Notes: This figure plots the distribution of sugar concentration for solid products in different categories before and after the policy implementation. Observations are weighted by pre-policy revenue.



(a) Solids (b) Liquids

Figure A.10: Share of products crossing the sugar threshold due to reformulation as a function of the average distance to the threshold

Notes: This figure shows the relationship between the pre-policy revenue-weighted share of products to the right of the sugar threshold that reduced their sugar concentration to fall below the regulatory threshold after the policy and the pre-policy revenue-weighted average distance between products to the right of the sugar threshold and the regulatory threshold. In Panel (a) we show solid categories and in Panel (b) we show liquid categories.

Table A.1: Chilean Food Act thresholds

	Solids (per 100 g)			Liquids (per 100 ml)		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
Energy (kcal)	350	300	275	100	80	70
Sodium (mg)	800	500	400	100	100	100
Sugar (g)	22.5	15	10	6	5	5
Saturated fat (g)	6	5	4	3	3	3

Notes: The table shows the level of calories, sodium, sugar, and fat at which a product would need to be labeled. The policy was implemented in three stages, with each stage setting stricter thresholds. Stages 1, 2, and 3 took place in June 2016, June 2018, and June 2019.

Table A.2: FOP warning-label policies

Country	Status	Year	Critical nutrients			
			Sugar	Sodium	Fat	Calories
Chile	Implemented	2016	✓	✓	✓	✓
Peru	Implemented	2019	✓	✓	✓	
Israel	Implemented	2020	✓	✓	✓	
Mexico	Implemented	2020	✓	✓	✓	✓
Uruguay	Implemented	2021	✓	✓	✓	
Argentina	Implemented	2022	✓	✓	✓	✓
Brazil	Implemented	2022	✓	✓	✓	
Colombia	Implemented	2023	✓	✓	✓	
Venezuela	Implemented	2024	✓	✓	✓	
Canada	Implemented	2026	✓	✓	✓	
Paraguay	Approved	2023	✓	✓	✓	
Bangladesh	In discussion	-	✓	✓	✓	
Barbados	In discussion	-	✓	✓	✓	
Costa Rica	In discussion	-	✓	✓	✓	✓
Dominican Republic	In discussion	-	✓	✓	✓	
El Salvador	In discussion	-				
Ethiopia	In discussion	-	✓	✓	✓	
Ghana	In discussion	-				
Guatemala	In discussion	-	✓	✓	✓	
Honduras	In discussion	-	✓	✓	✓	✓
Jamaica	In discussion	-	✓	✓	✓	
Kenya	In discussion	-	✓	✓	✓	
Nigeria	In discussion	-	✓	✓	✓	
Pakistan	In discussion	-				
Panama	In discussion	-	✓	✓	✓	
Philippines	In discussion	-	✓	✓	✓	✓
South Africa	In discussion	-	✓	✓	✓	
Tanzania	In discussion	-				
Trinidad and Tobago	In discussion	-	✓	✓	✓	
Uganda	In discussion	-	✓	✓	✓	
United States	In discussion	-	✓	✓	✓	

Notes: The table shows the adoption and discussion of mandatory FOP warning-label policies around the world. It includes countries where the policy has been implemented, approved but not yet implemented, or is under discussion. For countries with approved or implemented policies, "Year" indicates the date the policy was approved or the first implementation stage began. For policies under discussion, checkmarks reflect the nutrient-specific scope stated in a current proposal or government plan; blank cells indicate that the current proposal or plan does not specify a nutrient-specific warning scope. Information is current as of August 25, 2026.

Table A.3: Label probabilities

	Solids				Liquids			
	Calorie	Sugar	Sodium	Fat	Calorie	Sugar	Sodium	Fat
Panel A: Label probability								
Cond. on calorie label	1.00	0.48	0.05	0.51	1.00	0.63	0.27	0.83
Cond. on sugar label	0.75	1.00	0.02	0.50	0.31	1.00	0.08	0.23
Cond. on sodium label	0.28	0.07	1.00	0.29	0.87	0.52	1.00	0.80
Cond. on fat label	0.73	0.44	0.07	1.00	1.00	0.57	0.30	1.00
Panel B: Reformulation								
Cond. on calorie reformulation	-115.8	-2.5	-96.9	-2.7	-23.8	-2.7	-30.7	0.3
Cond. on sugar reformulation	-29.0	-9.4	-63.5	-1.3	-20.2	-5.5	-5.2	0.1
Cond. on sodium reformulation	-27.9	-0.5	-456.7	-1.4	-9.2	-1.7	-178.4	0.0
Cond. on fat reformulation	-43.3	-1.7	-83.0	-4.7	-29.8	-7.4	-25.5	-1.6

Notes: Panel A reports conditional probabilities of displaying multiple labels. For example, conditional on having a calorie label, 48 percent of solid products and 63 percent of liquid products also had a sugar label. All proportions are weighted by pre-policy revenue. Panel B reports changes in other nutrients for products that reformulated at least one nutrient. Reformulation is defined as a product's content per 100g/ml exceeding the regulatory threshold in the pre-policy period and falling below it in the post-policy period. Reformulation in one nutrient often coincided with improvements in others: among solid products that reformulated sugar, average sugar content declined by 9.4 g per 100 g without offsetting increases in other nutrients.

Appendix B: Sample Selection in our Within-Category Substitution Analysis

In this appendix, we discuss the sample selection process for the categories in our analysis of within-category substitution. We also provide basic descriptive statistics regarding sample coverage.

B.1. *Selection of Categories*

For the purpose of estimating the impact of labels on consumer demand, we first need to define the set of products contained in a given category. The ideal definition of a category for this exercise is a set of products that are sufficiently similar, such that consumers would consider substituting from one to another as a result of the regulation. We combine Walmart's category definitions with prior knowledge on substitution patterns to define adequate categories. We limit our attention to categories in which labeled products represent more than 5% and less than 95% of the total category revenue, and in which there are more than ten labeled and ten unlabeled products.

B.2. *Sample Coverage*

Table B.1 reports the share of revenue covered by the categories used in our demand analysis. Column (6) shows the share of total Walmart food revenue represented by each category in our consumer panel, which sums to 100 percent. We classify categories where nearly all products fall above or below the policy thresholds as "Mostly unlabeled" and "Mostly labeled." Together, these account for 57 percent of sales and include packaged fruit, meat, salads, candy, and chocolate. Another 6.9 percent corresponds to categories with fewer than 10 labeled and unlabeled products, while 1.9 percent reflects categories where treatment status could not be determined (e.g., peanuts, some cheeses). Our selected categories represent the remaining 34.2 percent of revenue. Products not subject to regulation, such as fresh fruit, vegetables, and alcohol, are excluded.

Column (1) reports the revenue-weighted share of products within each category that received a warning label. Column (6) reports each category's share of total market revenue; the categories in our working sample together account for 34.16 percent. Column

Table B.1: Selected categories used to study the impact of food labels on consumer demand

	Share of labeled products					Market share	
	Any label	Calories	Sugar	Sodium	Fat	Overall	Within
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Included	42.46	20.89	27.22	8.09	12.91	34.16	40.73
Bakery	41.94	40.69	38.64	1.86	14.49	0.39	0.46
Bread	21.89	16.41	18.12	0.00	7.53	3.70	2.28
Bread Toast	26.69	26.69	0.00	7.41	3.32	0.12	0.09
Burgers	60.02	0.00	0.00	3.41	56.61	1.08	1.83
Candy - Checkout	68.91	66.39	68.91	0.00	51.58	0.73	1.41
Cereal	62.69	62.69	36.47	0.00	0.00	2.35	4.14
Cereal bars, Diet and Health	63.47	63.18	17.66	0.14	11.51	0.83	1.48
Desserts and Cakes	47.78	17.63	45.07	0.00	16.95	3.23	4.34
Dressings, Oil, Sauces and Seasonings	32.25	16.39	3.39	21.57	12.36	8.52	7.72
Ice Creams	82.16	58.48	82.06	10.33	43.44	2.45	5.66
Inst. Noodles and Rice	38.73	25.06	0.00	34.71	19.77	0.24	0.27
Int. Cuisine: Mexican and Peruvian	27.55	0.84	1.10	27.55	0.00	0.12	0.10
Italian Dough and Pasta	20.82	19.57	0.00	1.24	0.00	0.59	0.34
Jam (jar), Syrup and Honey	61.80	26.15	55.31	0.00	25.98	0.53	0.91
Pork Ham	59.09	2.87	0.00	54.09	14.91	0.91	1.50
Soft and Energy Drinks	42.59	0.08	42.59	0.00	0.00	6.49	7.77
Sugar and Substitutes	8.57	5.86	8.44	0.00	0.00	1.86	0.45
Not included	32.05	25.20	14.21	2.20	17.73	65.84	59.27
Mostly labeled	98.71	84.72	51.47	5.33	50.06	15.21	42.17
Mostly unlabeled	0.75	0.13	0.55	0.15	0.12	41.88	0.88
Not enough control/treated products	69.22	43.67	12.41	8.29	57.29	6.90	13.40
Others	53.81	34.57	23.88	0.21	3.16	1.86	2.81
Total	35.61	23.73	18.65	4.21	16.08	100	100

Notes: Column (1) reports the share of labeled products within each category. Columns (2)–(5) report the share of products carrying each type of label—calories, sugar, sodium, and fat—within the category. Column (6) reports each category's share of total Walmart food revenue in our consumer panel; by construction, these shares sum to 100 percent. Column (7) reports each category's share of total labeled-product revenue, which sums to 100 percent across categories.

(7) restricts the sample to labeled products and shows that our working sample represents 40.7 percent of their pre-policy revenue.

Appendix C: Entry and Exit of Products

In this appendix, we examine whether the labeling policy affected product entry and exit. In principle, labels could induce firms to withdraw affected products or introduce unlabeled or reformulated alternatives, thereby shifting the composition of available UPCs. To assess this channel, we compare the number of unique UPCs across categories with high versus low label incidence, estimating:

$$\log(n_{cst}) = \sum_k \beta_k \cdot L_c \cdot \mathbb{1}\{k = t\} + d_{cs} + \delta_t + \varepsilon_{cst}, \tag{C.1}$$

where n_{cst} is the number of UPCs in category c and store s at time t, and L_c is the pre-policy revenue-weighted share of products in category c with at least one label. The specification includes store-by-category fixed effects, d_{cs}, and period fixed effects, δ_t. Coefficients are normalized so that their pre-policy average equals zero. Observations are weighted by pre-policy category-store revenue, with standard errors clustered at the category level.

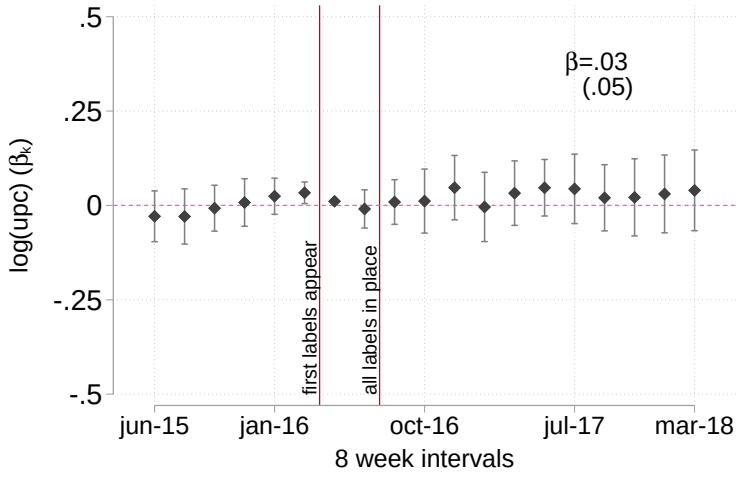


Figure C.1: Changes in the number of UPCs across categories

Notes: This figure plots the coefficients β_k from Equation (C.1), which capture differential changes in the number of UPCs between categories with higher versus lower label incidence. Coefficients are normalized so that their pre-policy average equals zero. Vertical lines mark the introduction of the first labels and the date of full coverage. Error bars denote 95 percent confidence intervals, with standard errors clustered at the category level.

Figure C.1 plots the event-study coefficients. The estimates are small, statistically

insignificant, and close to zero throughout the post-policy period. The post-policy average effect is 0.03 (s.e. = 0.05), and period-specific estimates fluctuate around zero without discernible patterns. Together with the absence of differential pre-trends, these results suggest that the labeling regulation did not systematically alter the extensive margin of product availability across categories.

To separately examine entry and exit, we divide the sample into four periods: (1) pre-policy; (2) transition (gradual rollout); (3) the first six months after full coverage; and (4) subsequent periods. For clarity, we compare only (1) and (4). Between these periods, 1.75% of UPCs (revenue-weighted) disappeared, while 9.73% of post-policy UPCs were newly introduced.

To assess whether these patterns were systematically related to labeling, we estimate:

$$\text{ShareExit}_c = \alpha + \beta \cdot L_c + \varepsilon_c,$$

where ShareExit_c is the share of UPCs in category c that disappeared post-policy, and L_c is the pre-policy revenue-weighted share with at least one label.

For entry, we analogously estimate:

$$\text{ShareEntry}_c = \alpha + \beta \cdot L_c + \varepsilon_c,$$

where ShareEntry_c is the share of UPCs in category c first appearing after the policy was fully implemented.

Table C.1 reports the results. Categories with a higher label share experienced slightly more exit and slightly more entry, but the coefficients are small and statistically insignificant. Taken together, these patterns suggest that product entry and exit are not first-order margins of adjustment to the labeling policy, and any compositional responses appear limited in scope.

Table C.1: Entry and exit probabilities

	Share Exit	Share Entry
	(1)	(2)
Share labeled (L_c)	0.018	0.020
	(0.012)	(0.025)
Constant	0.017	0.09
Observations	69	70

Notes: Each column reports results from a cross-sectional regression of exit or entry shares on the pre-policy share of labeled products at the category level. Robust standard errors in parentheses. The dependent variable in column (1) is the share of UPCs disappearing after the policy; in column (2) it is the share of new UPCs introduced after the policy. All estimates are revenue-weighted by pre-policy category size.

Appendix D: Sample Selection in our Product Reformulation Analysis

In this appendix, we discuss the sample selection process for the categories in our analysis of product reformulation. We also provide basic descriptive statistics regarding sample coverage.

D.1. *Selection of Categories*

Organizing the categories for this exercise requires different criteria from those used to study demand substitution in Appendix B. We focus on categories in which the distribution of sugar and calories is not entirely to the left of the regulatory threshold. Naturally, unlabeled products do not face any incentives to change their nutritional content. We also dropped categories with products that were too far to the right and for which it was not feasible to reduce the nutritional content to the threshold level.

D.2. *Sample Coverage*

In Table D.1, we show the share of revenue covered by the categories included in the supply-side analysis. Column (1) reports the share of total revenue for each category provided by Walmart. This column contains the universe of food products purchased by individuals from our consumer panel and adds up to 100%. We group a set of categories for which most products lie either below or above the policy threshold in the pre-policy period and label them "Mostly below" and "Mostly above." Together, these two groups represent close to 52% of revenue and include categories such as fruit, meat, salads, candy, and chocolate. Another 23.6% of total revenue is for categories with products that are exempt from the regulation (e.g., nuts) or products for which we are missing the pre-policy nutritional content. Our selected categories cover the remaining 24.5% of total revenue.

Column (2) reports the share of products, weighted by revenue, that are above the sugar threshold in the pre-policy period within a given category. Column (3) reports the share of total revenue for all products that are above the sugar threshold in the pre-policy period. Our working sample comprises around 53% of pre-policy revenue among products with the potential to bunch.

Table D.1: Selected categories used to study the impact of food labels on product reformulation

	Market share	Share above the sugar threshold before the policy	Market share within products above the sugar threshold
	(1)	(2)	(3)
Included	24.5	51.5	52.6
Cereal (g)	1.6	45.0	3.2
Cookies (g)	3.1	79.5	10.7
Desserts (g)	1.0	30.2	1.3
Condiments (g)	1.9	18.6	1.5
Seasonings (ml)	0.1	44.8	0.3
Frozen Fruit and Pulp (g)	0.2	24.9	0.2
Ice Cream (ml)	1.4	96.7	6.1
Jam (g)	0.6	87.2	2.4
Juice (ml)	3.8	47.4	7.9
Milk and Creams (ml)	4.0	28.1	4.9
Soft drinks (ml)	4.1	55.3	9.8
Soup (g)	0.8	6.2	0.2
Yogurt (ml)	1.8	51.8	4.1
Not Included	75.5	19.5	47.4
Mostly below	43.2	0.1	0.2
Mostly above	8.7	99.7	37.7
Others	23.6	56.8	9.4
Total	100	28.95	100

Notes: Column (1) reports the share of total revenue for each category provided by Walmart. This column contains the universe of food products purchased by individuals from our consumer panel and adds up to 100%. Column (2) reports the share of products, weighted by revenue, that are above the sugar threshold in the pre-policy period within a given category. Column (3) reports the share of total revenue for all products that are above the sugar threshold in the pre-policy period.

Appendix E: Threshold-Design Calibration

This appendix describes the threshold-design calibration discussed in Section 4.1. Let k denote the sugar threshold. For each value of k, we first predict how much reformulation occurs among products above the threshold. We then predict how much quantity consumers shift from labeled to unlabeled products within each category. Combining these two responses yields predicted total sugar purchases. We implement the calibration separately for solids and liquids, matching the thresholds in the Chilean regulation.

Throughout the appendix, "labeled" refers to a product's sugar-label status at threshold k. The demand estimates in the paper are based on whether a product received any warning label; here we apply those within-category responses to changes in sugar-label status, holding other nutrient thresholds fixed.

E.1. *Demand response*

Let c index category-by-form cells. When a category contains both solid and liquid products, we treat them as separate cells. For each counterfactual threshold k, let $x_c(k)$ denote the share of total quantity in cell c accounted for by products that remain sugar-labeled after predicted reformulation. This share varies with k because both the set of products above the threshold and their probabilities of reformulation depend on the threshold.

For a counterfactual threshold k, we parameterize the log demand response as

$$\beta_c^{\log}(k) = A_c\, k \left[1 - 4\left(x_c(k) - \frac{1}{2}\right)^2\right], \tag{E.1}$$

where $\beta_c^{\log}(k)$ is the log change in labeled quantities relative to unlabeled quantities in cell c, and A_c governs the strength of the within-cell response. The term in brackets allows the response to vary with the labeled share. It is zero when no products are labeled or all products are labeled, and it equals one when half of the cell's quantity is labeled. Under this parameterization, demand shifts only when the threshold leaves consumers facing both labeled and unlabeled products within the cell. Because we observe the demand response only at the actual threshold, we assume that its magnitude scales linearly with

k when evaluating counterfactual thresholds.[1]

We use the observed response at the actual Stage 1 threshold to recover A_c. Let $k_{0,c}$ denote that threshold: 22.5 grams per 100 grams for solids and 6.0 grams per 100 milliliters for liquids. For cells that can be matched to a direct category-level demand estimate, let $\hat{\beta}_c$ denote the estimated post-policy change in log quantities for labeled products relative to unlabeled products reported in Figure 5. A negative value of $\hat{\beta}_c$ indicates that purchases of labeled products fell relative to purchases of unlabeled products.

Let $\bar{x}_c$ denote the share of pre-policy revenue accounted for by products that would receive a sugar label under the actual Stage 1 threshold $k_{0,c}$. We choose A_c so that Equation (E.1), evaluated at $k_{0,c}$ and $\bar{x}_c$, reproduces the observed response $\hat{\beta}_c$. This gives

$$\hat{A}_c = \frac{\hat{\beta}_c}{k_{0,c}\left[1 - 4\left(\bar{x}_c - \frac{1}{2}\right)^2\right]}. \tag{E.2}$$

We can recover $\hat{A}_c$ from direct demand estimates for only a subset of category-by-form cells (12 solid and 4 liquid). To predict A_c for the remaining cells, we regress the recovered values on observable cell characteristics:

$$\hat{A}_c = \alpha + \pi_1 \bar{x}_c + \pi_2 \bar{n}_c + \pi_3 \,\text{liquid}_c + u_c, \tag{E.3}$$

where $\bar{x}_c$ is the pre-policy revenue-weighted share of products in cell c that would receive a sugar label under the actual Stage 1 threshold, $\bar{n}_c$ is the pre-policy revenue-weighted mean sugar density in the cell, and liquid_c is an indicator for liquid cells. Table E.1 reports the estimates.[2]

We use the OLS-fitted value as A_c for every cell, including those with direct demand estimates. This smooths sampling variation in the individual estimates and ensures that the supply regression and counterfactual calibration use the same demand parameters. For each counterfactual threshold k, Equation (E.1) combines A_c with the simulated post-

[1] At a higher cutoff, only the most sugar-dense products are labeled. A warning may therefore convey that a labeled product is more extreme in its sugar content, inducing consumers to substitute away from it more strongly.

[2] Because the regression contains only 16 cells, including 4 liquid cells, we use it for prediction and do not interpret the individual coefficients. LASSO and random-forest specifications produce similar predictions.

Table E.1: OLS imputation of demand parameter A_c

	Coefficient	Std. err.
Labeled share ($\bar{x}_c$)	0.015	(0.014)
Mean sugar density ($\bar{n}_c$)	−0.0002	(0.0003)
Liquid indicator (liquid$_c$)	−0.029	(0.008)
Constant	−0.009	(0.003)
Category-by-form cells with event-study A_c	16	
R^2	0.748	

Notes: This table reports estimates of Equation (E.3). The dependent variable is $\hat{A}_c$, recovered from the direct demand estimate using Equation (E.2). Each observation is a category-by-form cell with a direct demand estimate. The regression includes 12 solid cells and 4 liquid cells (16 in total). The labeled share and mean sugar density are revenue-weighted and computed within each cell. The labeled share is evaluated at the Stage 1 threshold for each form: 22.5 g/100 g for solids and 6.0 g/100 ml for liquids. The fitted values are used as A_c for all category-by-form cells in the calibration.

reformulation labeled share $x_c(k)$. Holding the threshold and labeled share fixed, a more negative A_c implies stronger substitution from labeled to unlabeled products. Panel (a) of Figure E.1 compares the recovered and fitted values of A_c.

E.2. *Supply response*

Let j index products in category-by-form cell c, let $f \in \{\text{solid}, \text{liquid}\}$ denote product form, and let n_{jc} denote product j's pre-policy sugar density. We define a product as reformulated if its sugar density was above the actual Stage 1 threshold before the policy and at or below that threshold afterward. Using the actual Stage 1 threshold for each form, we estimate a product-level logit of reformulation on the product's distance from the threshold, the within-cell dispersion of sugar content, the demand-incentive term $\log(|A_c|k_{0,c})$, and an indicator of reformulation feasibility.

For products above a counterfactual threshold k, the fitted probability of reformulating to avoid the sugar label is

$$p_{jc}(k) = \Lambda\left(b_0^f + b_{\text{dist}}^f(n_{jc} - k) + b_\sigma^f \sigma_c + b_A \log(|A_c|k) + b_\phi \phi_c\right), \tag{E.4}$$

where $\Lambda(z) = (1 + \exp(-z))^{-1}$, σ_c is the revenue-weighted within-cell standard deviation of log sugar density, and ϕ_c equals one for reformulable cells—that is, cells in which sugar

can be reduced without substantially changing product taste.[3] We constrain b_A and b_ϕ to be common across forms. The restriction on b_A assumes that demand incentives affect reformulation similarly for solids and liquids: in both forms, firms are more likely to reformulate when avoiding the label matters more for demand. We allow the intercept, distance coefficient, and dispersion coefficient to differ between solids and liquids, capturing differences in baseline reformulation rates and reformulation technology.

The term $\log(|A_c|k)$ increases with both the cell's underlying demand sensitivity and the counterfactual threshold. With $b_A > 0$, the predicted probability of reformulation approaches zero as $|A_c| \to 0$ or as $k \to 0$. We estimate the coefficients in Equation (E.4) using an unweighted product-level logit at the actual Stage 1 thresholds. We then use the estimated coefficients to predict reformulation probabilities at counterfactual thresholds. The estimation sample contains 450 solid products in 9 solid category-by-form cells and 609 liquid products in 13 liquid category-by-form cells. Table E.2 reports the estimated coefficients.

Figure E.1 compares predicted and observed demand and supply responses. Panel (a) plots the OLS-imputed demand parameter A_c against the value recovered from the event-study estimate, and Panel (b) plots the reformulation share predicted by Equation (E.4) against the observed share.

E.3. *Calibration algorithm*

The algorithm starts from the pre-policy product distribution. For each threshold k, we first predict each product's probability of reformulating using Equation (E.4). We then calculate expected labeled and unlabeled quantities after reformulation, assuming that reformulated products move exactly to the cutoff. Finally, we apply the demand response using the resulting labeled share.

Let m_{jc} denote the pre-policy quantity of product j, measured in kilograms for solids and liters for liquids. For a product above k, $m_{jc}p_{jc}(k)$ is the expected quantity that reformulates, while $m_{jc}[1 - p_{jc}(k)]$ is the expected quantity that remains labeled. Expected

[3] We set $\phi_c = 0$ for cells in categories such as cookies, chocolate, candy, and baked goods, in which sugar affects texture and bulk. We set $\phi_c = 1$ for reformulable cells such as jams, yogurts, dressings, and beverages, in which sugar can be replaced more easily.

Table E.2: Supply reformulation logit

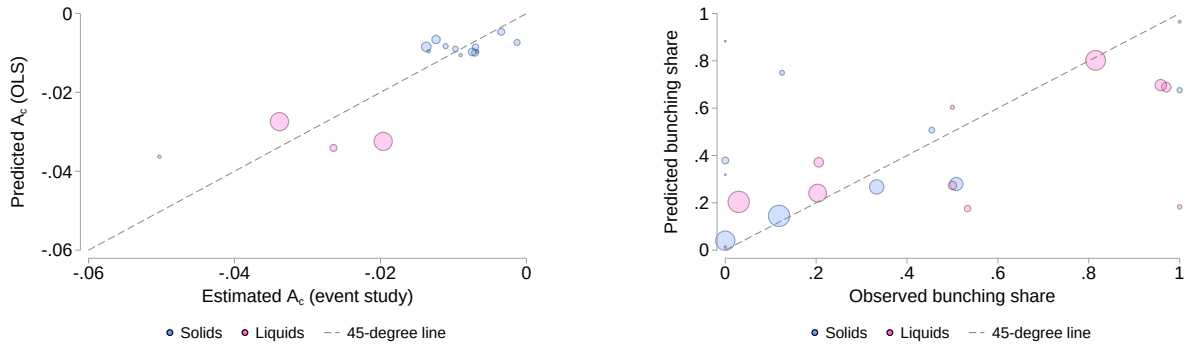
	Solids (g)	Liquids (ml)
Common across forms		
Demand incentive $\log(-A_c k_0)$, b_A	0.488	
	(1.123)	
Reformulable-cell indicator, b_ϕ	0.985	
	(0.786)	
Form-specific		
Distance to threshold $(n_{jc} - k)$, b_{dist}	−0.145	−0.338
	(0.062)	(0.099)
Log-sugar dispersion σ_c, b_σ	−1.327	4.562
	(1.862)	(0.581)
Constant, b_0	1.718	−0.240
	(2.029)	(2.204)
Products	450	609
Categories	9	13
Bunching rate	0.191	0.440
Observations	1,059	
Clusters (categories)	18	
Pseudo R^2	0.324	

Notes: This table reports estimates of Equation (E.4). The dependent variable equals one if a product was above the Stage 1 sugar threshold before the policy and at or below it afterward. The intercept, distance, and dispersion coefficients are allowed to differ between solids and liquids. The demand-incentive coefficient b_A on $\log(|A_c|k_{0,c})$ and the coefficient b_ϕ on the reformulable-cell indicator are constrained to be common across forms. The dispersion term σ_c is the within-cell standard deviation of log pre-policy sugar content, computed across purchased products using pre-policy revenue weights. The indicator ϕ_c equals one for reformulable cells, such as those containing jams, yogurts, dressings, and beverages, and zero for nonreformulable cells, such as those containing cookies, chocolate, candy, and baked goods. Standard errors, reported in parentheses, are clustered by category.

labeled quantity therefore consists of the above-threshold products that do not reformulate. Expected unlabeled quantity consists of products initially below the threshold and above-threshold products that reformulate:

$$M_c^L(k) = \sum_{j:n_{jc}>k} m_{jc}\left[1 - p_{jc}(k)\right], \tag{E.5}$$

$$M_c^U(k) = \sum_{j:n_{jc}\leq k} m_{jc} + \sum_{j:n_{jc}>k} m_{jc} p_{jc}(k), \tag{E.6}$$



(a) Demand: A_c (predicted vs. estimated)

(b) Supply: reformulation (predicted vs. observed)

Figure E.1: Model fit: predicted versus observed

Notes: Panel (a) compares the OLS-imputed demand parameter A_c from Equation (E.3) with the value recovered from the event-study estimate using Equation (E.2), for the 16 category-by-form cells with event-study estimates. Panel (b) compares the category-by-form reformulation share predicted by the reformulation logit in Equation (E.4) with the observed share, over the cells in the supply estimation sample. In both panels, marker size is proportional to the number of products in the cell (above-threshold products in Panel (b)); blue markers denote solids and pink markers denote liquids; the dashed lines are 45-degree lines.

where $M_c^L(k)$ and $M_c^U(k)$ are expected labeled and unlabeled quantities after reformulation. Because reformulated products are assigned sugar density k, the corresponding expected sugar totals are

$$S_c^L(k) = \frac{1}{100} \sum_{j:n_{jc}>k} m_{jc} \left[1 - p_{jc}(k)\right] n_{jc}, \tag{E.7}$$

$$S_c^U(k) = \frac{1}{100} \left[\sum_{j:n_{jc}\leq k} m_{jc} n_{jc} + k \sum_{j:n_{jc}>k} m_{jc} p_{jc}(k) \right]. \tag{E.8}$$

We then compute the post-reformulation labeled share,

$$x_c(k) = \frac{M_c^L(k)}{M_c^L(k) + M_c^U(k)}, \tag{E.9}$$

and apply the demand response in Equation (E.1). Let $T_c(k)$ denote the quantity shifted from labeled to unlabeled products. We hold total quantity in the cell fixed, so labeled quantity falls to $M_c^L(k) - T_c(k)$ and unlabeled quantity rises to $M_c^U(k) + T_c(k)$. When the cell contains both labeled and unlabeled products, we choose $T_c(k)$ so that the log change

in labeled quantity relative to unlabeled quantity equals $\beta_c^{\log}(k)$:

$$\beta_c^{\log}(k) = \log\left(\frac{M_c^L(k) - T_c(k)}{M_c^L(k)}\right) - \log\left(\frac{M_c^U(k) + T_c(k)}{M_c^U(k)}\right). \tag{E.10}$$

Solving for $T_c(k)$ gives

$$T_c(k) = \frac{M_c^L(k)M_c^U(k)\left[1 - \exp\left(\beta_c^{\log}(k)\right)\right]}{M_c^U(k) + \exp\left(\beta_c^{\log}(k)\right)M_c^L(k)}. \tag{E.11}$$

For cells containing both labeled and unlabeled products, the average sugar densities of the two groups are

$$d_c^L(k) = \frac{S_c^L(k)}{M_c^L(k)} \qquad \text{and} \qquad d_c^U(k) = \frac{S_c^U(k)}{M_c^U(k)}. \tag{E.12}$$

We assume that the quantity reduction is distributed proportionally across labeled products and that the corresponding increase is distributed proportionally across unlabeled products. The transferred quantity therefore removes sugar at the average labeled-product density $d_c^L(k)$ and adds sugar at the average unlabeled-product density $d_c^U(k)$. Sugar purchases after demand substitution are

$$\tilde{S}_c(k) = S_c^L(k) + S_c^U(k) - T_c(k)d_c^L(k) + T_c(k)d_c^U(k). \tag{E.13}$$

If a cell contains only labeled products or only unlabeled products, Equation (E.1) gives $\beta_c^{\log}(k) = 0$. There is then no demand substitution, and $\tilde{S}_c(k) = S_c^L(k) + S_c^U(k)$.

Finally, we aggregate sugar purchases across cells within each product form:

$$\tilde{S}_f(k) = \sum_{c \in \mathcal{C}_f} \tilde{S}_c(k), \tag{E.14}$$

where $\mathcal{C}_f$ is the set of category-by-form cells for form f. We evaluate this expression over separate grids of thresholds for solids and liquids. Figure 10 plots the resulting sugar purchases. To construct the demand-only curves, we repeat the calculation setting $p_{jc}(k) = 0$ for every product, thereby shutting down reformulation.