On the Design of Food Labeling Policies[a]

Nano Barahona[†] Cristóbal Otero[⋆] Sebastián Otero[§] Joshua Kim[‡]

April 18, 2022

Abstract: We study a regulation in Chile that mandates front-of-package warning labels on products whose sugar or caloric concentration exceeds certain thresholds. We document an overall decrease in sugar and caloric intake of 7-9%. To unpack the underlying mechanisms, we provide descriptive evidence of the impact of the policy on consumer choice, both across and within categories and firms' behavior. We find no noticeable substitution of products across food categories and show that most of the demand effect of the regulation comes from within-category substitution. We also find that a substantive portion of the overall effect comes from product reformulation. We discuss how these findings can inform the design of effective labeling policies.

Keywords: Food labels, nutrition, obesity, product reformulation, sugar taxes.
JEL Codes: D12, D22, I12, I18, L11, L81

[a]First version: April 2022. We want to thank Hunt Allcott, Liran Einav, and Matthew Gentzkow for their comments and suggestions. We also thank Camila Corvalán and Marcela Reyes for benefitial conversations on institutional details; Christine Von Dessauer, Roberto Cases, and Romina Quagliotti 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) 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.
[†]University of California, Berkeley. Email: nanobk@berkeley.edu. [⋆]University of California, Berkeley. Email: cotero@berkeley.edu. [§]Stanford University. Email: sotero@stanford.edu. [‡]Facebook. Email: joshkim@fb.com.

1. Introduction

The average American adult weighs nearly 23 pounds more today 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 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.

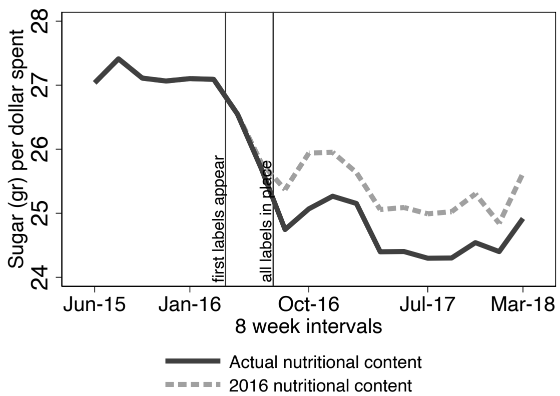
An increasingly popular policy is to provide simplified information about products' healthiness to consumers through front-of-package (FOP) warning labels. These labels are simple symbols that clearly signal to consumers when a product is considered unhealthy based on whether their sugar or caloric concentration exceeds certain thresholds. Chile was a pioneer country in implementing government-mandated FOP warning labels in 2016. Since then, more than 25 countries have approved or are considering the introduction of similar regulations. In this article, we review the Chilean experience and provide guidelines for the effective design of FOP warning-label policies.

To study how the regulation affects consumer choice, we use scanner data on all purchases made in Walmart, the largest food retailer in Chile, from 2015 to 2018. The data contain information on prices and quantities, alongwith consumer demographics. We combine these records with the nutrition facts tables for products before and after the policy to study its effects on the overall demand and supply of nutrients.

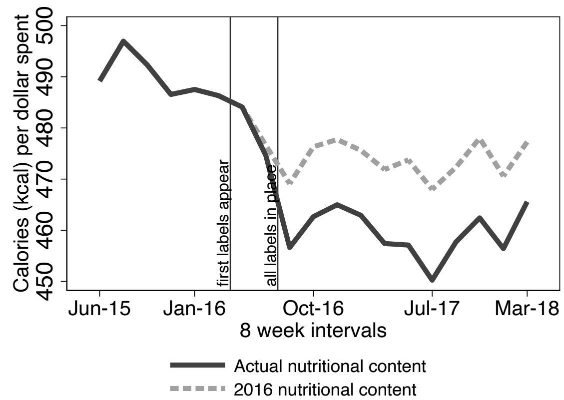
We start by documenting a sharp overall decrease in sugar and caloric intake of 8.8% and 6.5% per dollar spent, respectively, immediately after the policy was phased in. We show these findings in Figure 1, which we produced by using a panel of Walmart consumers and computing their sugar and caloric intake during every 8-week period. This reduction, which persists for the 2-year post-policy window in our data, is explained by a combination of demand- and supply-side responses. On the one hand, consumers reacted to the regulation by making healthier choices—even when the nutritional content of products is kept constant over time (dashed curves). On the other hand, 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).[1] To the best of our knowledge, this is the first evaluation of the effects of the Chilean food labeling policy on nutritional intake across all food and beverage categories.

In the rest of this article, we study how policymakers should design food labels to harness demand- and supply-side responses effectively. Recent warning-label policies use

[1] 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: After the labeling policy was introduced, the total sugar intake decreased from 27.3 to 24.9 grams of sugar per dollar, and the total caloric intake decreased from 488 to 457 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.

the same thresholds for all food and beverage groups, with the only differences being for liquids and solids. In the Chilean case, regulatory cutoffs were set to distinguish healthy from unhealthy products based on the natural foods and liquids that are considered to be the gold standards of a healthy diet (Corvalán et al., 2019). Next, we discuss other considerations that policymakers should take into account when setting regulatory thresholds.

Food labels work largely by providing information to consumers and correcting mistaken beliefs about the healthiness of products (Barahona et al., 2022). As a result, labels can induce consumers to properly internalize all the costs they impose on themselves, shift their purchasing behavior, and improve consumer welfare. Accordingly, labels should be set to maximize the relevant information for consumers, which is defined by two conditions. First, information should be relevant in the sense that consumers are willing to substitute labeled for unlabeled products as a result of changes in perceived product healthiness. Second, information should be informative in the sense that it can correct mistaken beliefs about product healthiness.

We perform several empirical exercises to understand the extent to which both conditions hold across different food products. We start by studying whether the food labeling policy has the potential to shift consumer demand between categories (e.g., from ready-to-eat cereal to fruits). To do so, 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.

Next, we document important within-category substitution across labeled and unlabeled products in several product categories. We use an event-study design to compare the quantities sold of labeled and unlabeled products. 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 treatment effects across different categories. To understand the source of this heterogeneity, we implement a survey in which we elicit consumers' beliefs about the nutritional content of soft drinks and cereal. These categories present the smallest and largest treatment effects, respectively. We find that beliefs are very accurate for soft drinks but not for cereal. These results are consistent with the idea that food labels are more effective in shifting demand categories in which labels are more informative (e.g., cereal).

On the supply side, firms might respond to labeling policies by reformulating their products and avoiding labels. In addition to taking demand responses into account, policymakers may want to consider firm responses to enhance the overall effects of labels. Policymakers interested in promoting product reformulation should have two criteria. First, they should identify food and beverage categories in which labels induce important changes in consumer demand toward unlabeled products. Second, they should design labeling requirements such that firms in those categories have incentives to reformulate, given their cost structure.

To empirically assess supply-side 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 5% of ex ante labeled cookies changed their nutritional composition. We argue that 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. Our results are consistent with this pattern, whereby categories with liquid products present a larger share of products that bunch.

Our results are important for policy design. First, the lack of between-category substitution implies that policymakers need to set regulatory thresholds to maximize the effects within specific categories. Second, it is important to target categories that represent a large share of consumers' bundles and overall intake of critical nutrients. Third, thresholds must be set to maximize substitutability within targeted categories, which means that: (a) the categories need to have both healthy and unhealthy products that are close substitutes and (b) consumers must be misinformed about the healthiness status of those products. Fourth, thresholds need to be set tighter on targeted categories in which reformulation is feasible at a low cost.

It is important to note that, in addition to the misinformation internality highlighted in this article, there is growing evidence that consumers may not always choose the food products that are best for them as a result of lack of self-control or time inconsistency (Sadoff et al., 2020; Samek, 2019). In those cases, a better policy tool may be to implement sugar taxes (Barahona et al., 2022). Consequentially, because labels and sugar taxes counteract different internalities, they should be seen as complementary policies rather than substitutes.

This paper contributes to several strands of the literature. First, it contributes to the literature that examines the effect of food labeling regulations on the demand for food (Kiesel and Villas-Boas, 2013; Zhu et al., 2015; Allais et al., 2015) and on firms' strategic responses by reformulating products (Moorman et al., 2012; Lim et al., 2020; Reyes et al., 2020; Barahona et al., 2022). Our paper contributes to these studies by providing evidence of and quantifying the effects of national food labeling regulations on multiple categories.

This is not the first paper to study the Chilean Food Act. Using a before-after analysis, Taillie et al. (2020) document a significant decline in purchases of labeled beverages following the policy's implementation. Araya et al. (2022) take advantage of the staggered introduction of labeled products in store inventories and find that labels decrease demand in the breakfast cereal category, but not for chocolates or cookies. Three other papers focus on supply-side responses to the policy in the cereal market. Pachali et al. (2022) study price adjustments and conclude that the prices of labeled products increased due to increased product differentiation. Alé-Chilet and Moshary (2022) provide evidence of bunching just below regulatory thresholds and conclude that reformulation reinforces the policy's effects by lowering the caloric content of cereal. Barahona et al. (2022) develop and estimate an equilibrium model that allows both price adjustments and product reformulation. In this paper, we go beyond cereal and study the effects of the policy on several product categories. We document important heterogeneity across categories in both demand- and supply-side responses and discuss how these differential effects matter for policy design.

Finally, we contribute to a broader literature that studies how governments can help consumers make better nutritional choices. Allcott et al. (2019) examine whether improving access to healthy food in poor neighborhoods can decrease nutritional inequality; Dubois et al. (2017) analyze the effect of advertising on junk food consumption; and several other papers study the effects and design of taxes for sugar-sweetened beverages and calorie-dense food 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). Our paper focuses on food labels and shows that they can be an effective tool to improve diet quality and combat obesity when carefully designed.

The remainder of the paper is organized as follows. Section 2 describes the setting and the data. In Section 3, we provide a conceptual framework to address the key demand and supply forces through which food labels can reduce the intake of critical nutrients. In Section 4, we present empirical evidence of how labels work in practice. We discuss policy implications and conclude in Section 5.

2. Setting and data

2.1. *FOP warning labels and the Chilean Food Act*

In recent years, many countries have introduced FOP labels to help consumers make healthier food choices. Unlike back-of-package labels, FOP labels simplify nutritional information, which makes it easier to use and interpret in a context in which shoppers make quick purchasing decisions. Several types of FOP labels are used around the world, which has prompted extensive discussion of the effectiveness of different visual designs and the simplicity or complexity of the information provided (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 particular product is high in a given nutrient.[2] Perhaps due to their simplicity, warning labels have become popular in the last few years. Since 2016, more than 25 countries have either implemented or are in the process of implementing country-

[2]Other common FOP labels are Multiple Traffic Lights (MTL), and Guidelines Daily Amounts (GDA). MTL uses traffic light colors and text descriptors to indicate whether a product is low, medium, or high in a given critical nutrient. They are similar to warning labels, in the sense that they provide a qualitative interpretation of the nutritional content of products. In contrast, GDA labels summarize nutritional information in percentage per portion relative to the guideline-based daily intake and do not provide any explicit recommendation (Jáuregui et al., 2020).

wide mandatory food labeling policies. Table 1 presents the list of countries, including the critical nutrients targeted in the regulation.

Table 1: 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	✓	✓	✓	
Brazil	Approved	2020	✓	✓	✓	
Argentina	Approved	2021	✓	✓	✓	✓
Colombia	Approved	2021	✓	✓	✓	
Venezuela	Approved	2021	✓	✓	✓	
Antigua and Barbuda	In discussion	-	✓	✓	✓	
Bahamas	In discussion	-	✓	✓	✓	
Barbados	In discussion	-	✓	✓	✓	
Canada	In discussion	-	✓	✓	✓	
Costa Rica	In discussion	-	✓	✓	✓	✓
Dominica	In discussion	-	✓	✓	✓	
El Salvador	In discussion	-	✓	✓	✓	✓
Guatemala	In discussion	-	✓	✓	✓	
Guyana	In discussion	-	✓	✓	✓	
Haiti	In discussion	-	✓	✓	✓	
India	In discussion	-	✓	✓	✓	✓
Jamaica	In discussion	-	✓	✓	✓	
Panama	In discussion	-	✓	✓	✓	✓
Paraguay	In discussion	-	✓	✓	✓	✓
Saint Kitts and Nevis	In discussion	-	✓	✓	✓	
Saint Lucia	In discussion	-	✓	✓	✓	
Suriname	In discussion	-	✓	✓	✓	
Trinidad and Tobago	In discussion	-	✓	✓	✓	

Notes: The table shows the introduction and discussion of mandatory FOP warning-label policies around the world. The table includes countries with laws or government resolutions implemented, approved but not implemented, or under discussion. Countries with discussions on the topic that do not have a specific government plan or law proposal to establish mandatory FOP warning labels are not included. For countries with approved or implemented policies, "Year" indicates the date the policy was approved or the first implementation stage began. The sources for this table can be found in Table A.1. Some of these policies distinguish between total fat, saturated fat, and trans fat; we group all of them together for expositional purposes.

Chile was the first country to introduce a nationwide mandatory FOP warning-label policy.[3] The policy responds to obesity being the most prevalent chronic disease in the

[3]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,

country: Before the policy, 45% of children and 74% of adults were overweight or obese (OECD, 2019). In 2016, Chile passed Law 20.606 (hereafter, the Food Act), which introduced FOP warning labels to inform consumers about products that can harm health and help guide purchasing decisions.[4] 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).[5]

The Chilean Food Act mandated that products with calories, added sugars, 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 USDA. Introduction of the thresholds was, however, gradual and implemented in three stages. Stages 1, 2, and 3 took place in June of 2016, 2018, and 2019 respectively. Threshold values are presented in Table 2.[6]

Table 2: Chilean Food Act thresholds

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

Sweden, Denmark, Norway, Lithuania, and Iceland used the Keyhole logo; Netherlands, Belgium, and Poland used the Choices logo; Korea and the United Kingdom used traffic-light labels; 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.

[4] The Food Act also included regulations to ban selling labeled products in schools and a ban to advertise labeled products aimed at children younger than 14 years old.

[5] Recent experimental evidence informs the trade-off between precision and simplicity of the information displayed on food labels and finds that coarse-categorical labels are more effective than detailed numerical ones (Ravaioli, 2021).

[6] A relevant exception in the Food Act is that it only regulates processed and packaged foods. Therefore, products with no added sugars, sodium, saturated fats, honey, or syrup do not receive a label even if they are above a given threshold.

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 2 presents the threshold values that determine the assignment of each label.

2.2. *Data*

2.2.1. *Walmart data*: To capture prices and quantities, we use scanner-level data provided by Walmart-Chile. Walmart is the largest food retailer in Chile and accounts for more than 40% of supermarket sales. Our data contain all transactions in all Walmart stores in Chile between May 2015 and March 2018. Every transaction identifies products at UPC (Universal Product Code) level and contains information on the price, revenue, product name, brand name, and discounts.[7] We can also track buyers enrolled in Walmart's loyalty program and link them to individual characteristics, such as gender, age, and household income. Finally, we supplement these data with additional information on product and store characteristics, also provided by Walmart.

Since our data only cover purchases at Walmart and most consumers may also purchase a large share of their groceries from other retailers, we restrict our analysis to regular Walmart customers only. Our final sample consists of 524,000 consumers who visited a Walmart store at least once every 8 weeks during the study period. The average age of customers in our panel is 48 years, and 69% are women. In the first year of data, from May 2015 to May 2016, the median customer shops at Walmart 24 times, at three Walmart locations, and travels 3 kilometers to get to the closest store.[8]

[7]The data comprise over 9 billion transactions by over 5 million consumers for over 20,000 different food products.

[8]We count as a visit anytime a customer spends at least $20 on food products.

2.2.2. *Nutritional Information*: The nutritional data for packaged products comes from two sources: (a) pre-policy data collected by the Institute of Nutrition and Food Technology (INTA) at the University of Chile and (b) post-policy data we collected and digitized ourselves. For non-packaged products, we use publicly available data provided by the United States Department of Agriculture (USDA).

Pre-policy: Anticipating implementation of the Food Act, the INTA collected nutritional information for a sample of products in January 2016 at UPC level. This included the nutrition facts table, whether the product is a liquid or a solid, and the size of the package.

Post-policy: To supplement the INTA's nutritional data, we hand-collected nutritional information as follows: We developed a camera phone app to take pictures of nutrition facts tables and linked them to the Walmart scanner-level data. A team of enumerators visited the three largest Walmart stores in Chile and used the app to digitize the nutritional content of all available products. Our dataset covers 90% of Walmart's revenue from packaged food products. We collected this information in March 2018, two years after the first stage of the law was implemented in June 2016.

USDA: For products sold without nutrition facts tables, such as fresh produce or meat, we rely on FoodData Central data that is publicly available from the USDA. We use this data to fill in any missing data for all food categories.

2.2.3. *Consumer beliefs*: We conducted a survey to elicit consumers' beliefs about the nutritional characteristics of a set of cereal and soft drink products in the absence of warning labels. We administered the survey in Argentina using Qualtrics in August 2019 and surveyed 1,500 individuals. Ideally, we would have elicited consumer beliefs in Chile before the policy's implementation. However, since this was not possible, we mimicked this exercise by conducting the survey in Argentina, whose population and food market are similar to Chile's but which has not been exposed to any food labeling policy. We asked consumers to provide their best estimate of the sugar and caloric concentration of a set of products and to state how confident they were about their answers. We elicit the first and second moments of consumer beliefs about each product's nutritional content using this information.

3. Conceptual Framework

To study the design of effective food labeling policies, it is critical to understand what market failures they can solve. In this section, we show that mandatory warning labels can help mitigate consumer biases linked to the demand for food products. We also discuss why some of these biases are suitable for government intervention. In addition, we argue that sophisticated policymakers should design policies that consider the equilibrium outcomes that may arise from demand and supply responses to the labels.

3.1. *Food labels and consumer biases in the demand for food*

Obesity is a worldwide pandemic driven by highly processed and unhealthy food consumption. It is well known that many consumers choose to buy highly palatable foods despite their health consequences. This is a perfectly rational decision if the experience and tastiness benefits outweigh the associated health risks. However, a growing body of research suggests that consumers may overpurchase unhealthy products beyond their true preferences because they fail to internalize the harm these choices represent (Bernheim and Taubinsky, 2018; Allcott et al., 2019b).

Several behavioral reasons and cognitive biases can explain food consumption's unaccounted internalities. The first source arises from inattention to the potential harm some nutrients can cause (Allcott et al., 2019). A second source is misinformation about the healthiness of products, in the sense that consumer beliefs over product's nutritional content are poorly calibrated (Barahona et al., 2022). Finally, a third source is food addiction in the form of a lack of self-control and time inconsistency (Sadoff et al., 2020; Samek, 2019). Each source has different implications for the design of policy interventions that are appropriately targeted. Naturally, these internalities are not mutually exclusive and may occur simultaneously. As such, policymakers should use a combination of different tools to combat food overconsumption efficiently.

Food labels in the form of FOP warning labels can mitigate the first two sources of internalities: inattention and misinformation. First, labels can be used as a "nudge" to manipulate the salience of specific product attributes, similar to what graphic imagery does on cigarette packs (Bernheim and Taubinsky, 2018). Warning labels have the potential to induce consumers to become more aware of the health consequences of food, thus shifting their demand away from unhealthy products. Second, labels can provide simplified information to consumers about non-observed attributes of products. Even though the nutritional content of products is available in the nutrition facts table on the back of the package, FOP labels can affect product choices by easing the process of acquiring information and translating the meaning of nutritional content into simplified health recommendations.

3.2. *The contribution of mandatory labels in promoting a healthier bundle of products*

It is natural to ask whether profit-maximizing firms could solve these inattention and informational problems in a well-functioning market. But unfortunately, even though suppliers that produce healthy products have incentives to develop healthiness certification marks to increase demand for their products, we do not observe firms or third-party organizations engaging in this behavior.

Dranove and Jin (2010) discuss several conditions that are required for markets to unravel, such as vertical differentiation in a well-defined measure of quality; competitive markets with no strategic interactions among sellers; homogeneous consumers; rational expectations of consumers regarding the quality of non-disclosed products; and the distribution of available quality being public information in the absence of disclosure. None of these conditions are met in the food industry. From a policy standpoint, these frictions imply that it is unlikely that the market itself will self-regulate, which provides ground for government intervention. This is perhaps the most important argument for government-mandated labels.

In the absence of mandatory labels, firms do not have incentives to invest in reducing the concentration of critical nutrients in their products, which results in unhealthier products. On the other hand, if consumers increase the demand for healthier products in response to labels, firms can benefit from reducing the concentration of labeled nutrients such as sugar or fat. That is, firms may have incentives to change the composition of some of their products to avoid labels and increase their demand (Reyes et al., 2020). In equilibrium, regulated markets with labels can yield consumers making better nutritional decisions and firms offering a healthier bundle of products (Barahona et al., 2022).

3.3. *How should policymakers design their food labels?*

Policymakers implementing labels need to decide where to set label thresholds. Recent warning-label policies use the same thresholds for all food and beverage groups, with the only differences being for liquids and solids. In the Chilean case, regulatory cutoffs were set to distinguish healthy from unhealthy foods based on the natural foods and liquids that are considered to be gold standards of a healthy diet (Corvalán et al., 2019). Should labels be set differently based on the above considerations? The answer depends on how much labels affect consumer demand and how the supply side responds in equilibrium.

If labels work via information, then they should be design to maximize relevant information for consumers. Information is more relevant for consumers when (a) it corrects mistaken beliefs about product healthiness, and (b) consumers are willing to substitute labeled and unlabeled products. The first condition states that labels will be more or less informative depending on how accurately consumers can differentiate healthy from unhealthy products in the absence of labels. For instance, in the soft-drinks category, consumers can clearly distinguish between Coke and Coke Zero, and thus the labels do not provide new information. The second condition implies that the effect of labels will also depend on whether labeled and unlabeled products are close substitutes. For example, if consumers have high substitution rates between cereal and fruit, regulatory cutoffs should be strict to induce consumers to switch to a healthier category. In contrast, if there is no substitution across these categories, then labels should be set to maximize substitution only among cereal products.

In addition to affecting consumer behavior, labels have the potential to encourage firms to reformulate their products. Policymakers who seek to promote product reformulation should have two criteria. First, they should identify food and beverage categories in which labels induce important changes in consumer demand toward unlabeled products. Second, they should set labels such that firms in those categories have incentives to reformulate, given their cost structure. For instance, reformulation in yogurt and juice is relatively easy to implement, and thus policymakers can set stricter regulatory thresholds. In contrast, modifying cereal products is more expensive, and therefore more lenient thresholds are more likely to promote reformulation.

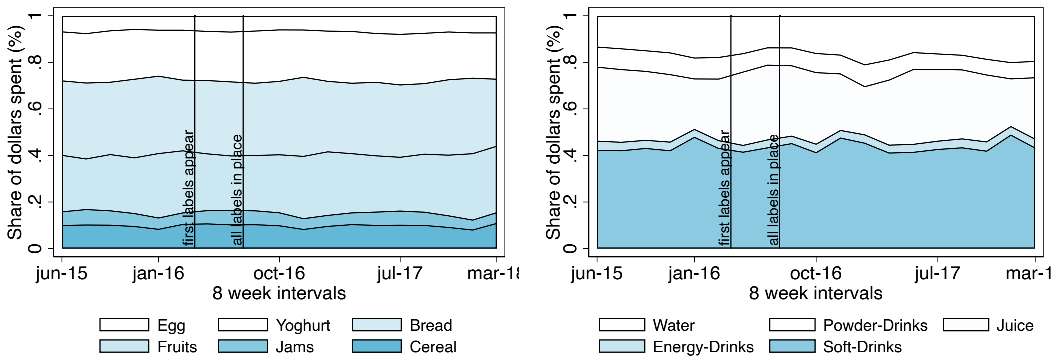
In the next section, we study whether labels induce between- or within-category substitution, the relevance of the accuracy of beliefs for the effectiveness of food labels, and the extent to which products were reformulated to avoid receiving a label. We then use these results to discuss how policymakers should think about their choice of regulatory thresholds.

4. Empirical evidence

4.1. *Between-category substitution*

We start by examining whether consumers shift their consumption of food products across food categories. 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 between categories for product categories that are likely to be eaten at breakfast: eggs, yogurt, bread, fruits, jams, and breakfast cereal. Then, within each broad group, we compare revenues before and after the policy for categories with a high and a low share of labeled products.

In Figure 3, we plot changes in the share of revenue over time of the food categories that fall into the breakfast and the 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—62% of the products are labeled.



(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 (e.g., snacks) are categories in which all products received at least one label. We show no differential changes in dollars spent between low-in-labels and high-in-labels categories.

Figure 3 suggests there is little to no evidence that consumers are shifting consumption from highly labeled categories, such as breakfast cereals or soft drinks, to low labeled categories, such as eggs or juices. For instance, in Panel (a), the share of breakfast spending going to cereals averages 9.9% both in the pre-and post-policy period. [9]

To formalize these results, we pool all food categories together and run the following regression:

$$\log(r_{cst}) \quad = \quad \beta_t \cdot L_c + d_{cs} + \delta_t + \varepsilon_{cst}, \tag{1}$$

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. Finally, δ_t denotes period fixed effects and d_{cs} refers to category-store fixed effects. We normalize the β_t coefficient corresponding to the first period post-adoption to zero. Observations are weighted by category-store pre-policy revenues and standard errors clustered the category level.

Figure 4 displays the results from estimating Equation (1). In both the pre-and post-policy period, the difference in coefficients is small, and none of them is significantly different from zero.[10] Regression results are consistent with the results in Figure 3. This

[9]In Online Appendix A Figure A.2, we show that this finding extends to several other food groups, such as carbs, meats, desserts, and snacks.

[10]Note that the variance of the estimates increases in January and February. Since we are pooling many

evidence suggests 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.

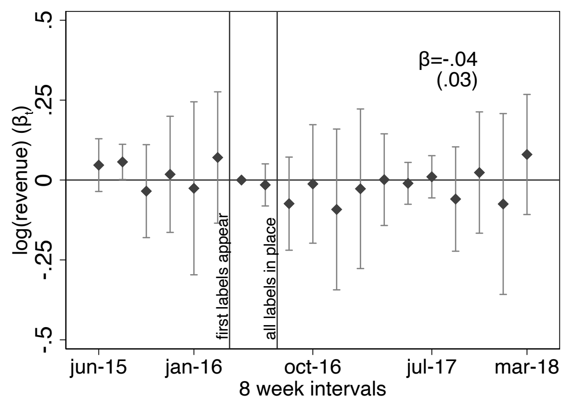


Figure 4: Total spending per capita on different categories

Notes: The figure presents the β_t coefficients from Equation (1). Vertical lines delimit the 95% confidence intervals. These regressions are run on a sample of 69 categories. The average share of labeled products in each category is 0.3, with a minimum of 0 and a maximum of 1.

4.2. *Within-category substitution*

In this subsection, we study 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 eight different categories and subcategories that meet the following criteria: (a) products are sufficiently similar such that consumers would consider substituting from one to another as a result of the regulation, (b) there is sufficient variation in terms of the share of products that received a label, and (c) unlabeled and labeled products within the category follow similar pre-trends in the absence of the policy. We select eight categories that represent 6.4% of the pre-policy revenue of all food products and 11.7% of the pre-policy revenue of all labeled products. In Appendix B we explain in more detail the selection process for these categories and sample coverage.

We define a product as the union of UPCs that share the same product name and brand. For example, we assign all *Diet Coke* the same product ID regardless of their can or bottle size. We assign labels to a product based on its 2018 nutritional content. We

different products together, seasonality affects them differently. However, as long as the seasonality of different categories is not correlated with the shares of products labeled, this should not bias our estimates. To control for seasonality, we need to compare coefficients belonging to the same period of the year as the normalized coefficient. When we proceed this way, coefficients are more precise and still not statistically different from zero.

collapse our original data into product-store-period data bins (in which a period is defined as eight consecutive calendar weeks) and estimate the following regression:

$$\log(q_{jst}) \;=\; \beta_t \cdot L_j + \gamma \cdot \log(p_{jst}) + \delta_{js} + \delta_t + \varepsilon_{jst}, \tag{2}$$

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_j is an indicator variable that takes the value of one if the product has one or more labels. Finally, δ_{js} refers to product-store fixed effects and δ_t to period fixed effects. We normalize the β_t coefficients so that their average value over the pre-policy period is zero. 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 product level.[11]

In Figure 5, we plot the estimated changes in demand for each of the categories. We find an effect for all eight categories, ranging from 10% in the case of soft drinks to 26% in the case of cereals. These results confirm that consumers substituted from labeled to unlabeled products within several categories and that the effect holds in the medium run, which complements the results for breakfast cereals of Barahona et al. (2022).[12]

4.3. *The importance of prior beliefs*

In related work, Barahona et al. (2022) 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 product level, labels are more effective if they provide new information to consumers. 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.

Figure 6 shows a scatter plot of the first moments of beliefs about each product's nutritional content versus the true nutritional content in the pre-policy period. The size of each circle represents the total pre-policy revenue of the given product. Figure 6(a) focuses on soft drinks and Figure 6(b) on cereal. The observed pre-policy concentration

[11] Given our context, in which consumers substitute from one product to another, it is natural that the no-interference assumption—standard in the impact evaluation literature—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 impact on the relative change in the equilibrium quantities of labeled versus unlabeled products sold.

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

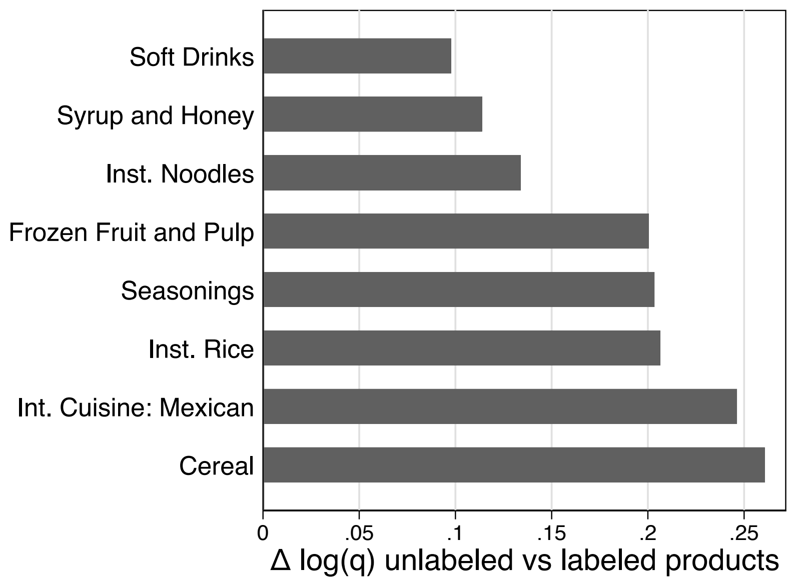


Figure 5: Changes in demand for selected categories

Notes: This figure summarizes the findings on changes in equilibrium quantities from Subsection 4.2. It 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 12% of the pre-policy revenue of labeled products in the sample. We provide more details on sample selection in Appendix B.

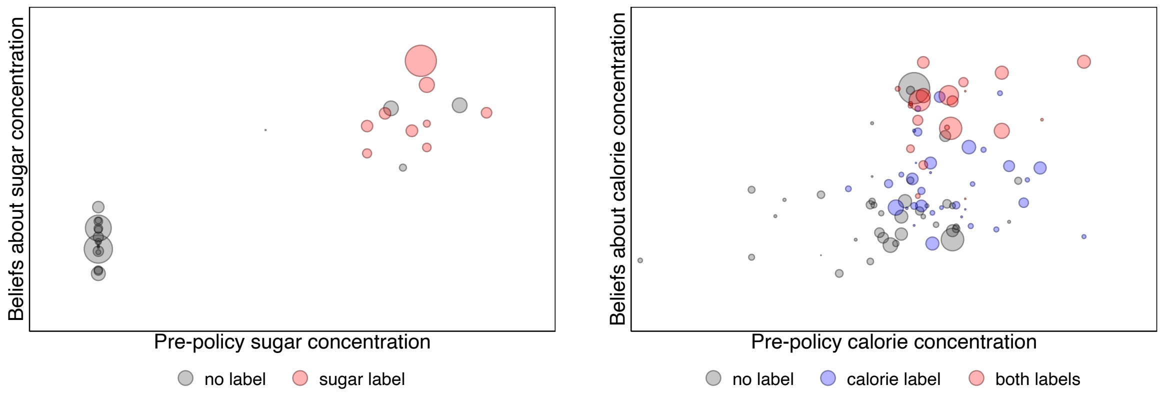
of sugar in soft drinks follows a bimodal distribution driven by diet and non-diet drinks, which shows that respondents do a good job of understanding that diet drinks are lower in sugar. In cereal, however, consumers have mistaken beliefs about the caloric content of products. The correlation between the average value of respondents' beliefs about the caloric concentration of each product and the product's observed pre-policy caloric concentration is 0.23.

The accuracy of beliefs over sugar content implies that labels came as no surprise in this category, which means that the effect of the policy should be smaller in soft drinks than in cereal. To test for this, 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).

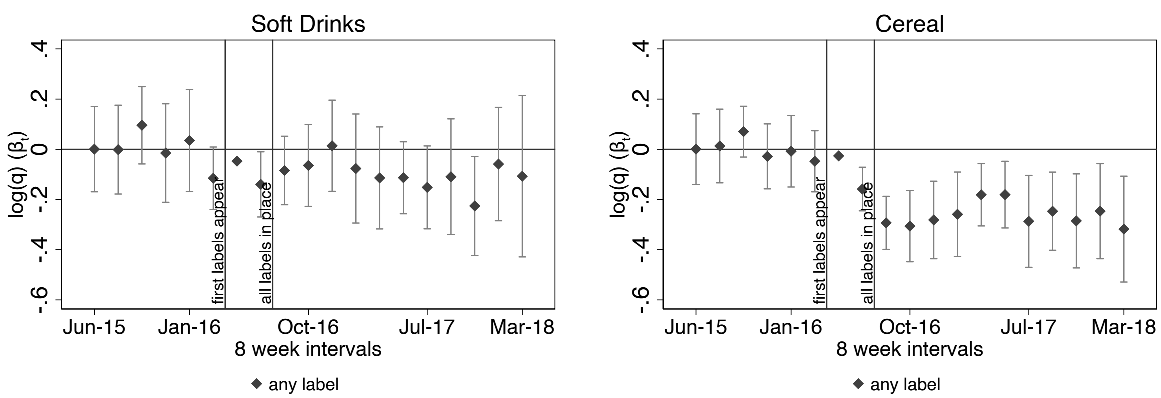
Figure 7 displays the results of estimating Equation (3) for soft drinks and cereal. Pre-period coefficients are small and not significantly different from zero in both categories. After introduction of the labels, we observe a noticeable drop in the demand for labeled products relative to that for unlabeled products in the cereal market but not in the soft drinks market.



(a) Soft Drinks (b) Cereals

Figure 6: 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 cereal, and its size represents the total revenue from that product in our sample period. Panel (a) focuses on 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) Soft Drinks (b) Cereals

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

Notes: This figure presents the estimates of our event study regressions from Equation 3. Panel (a) presents the β_k coefficients for soft drinks. Panel (b) displays the β_k coefficients of the regression on the sample of 68 ready-to-eat cereals that show up in the pre- and post-policy periods. The sample consists of 27 unlabeled and 41 labeled products for a total of 194,510 observations.

The main takeaway is that labels have a larger within-category effect when they provide useful information to consumers. This is consistent with the evidence of Araya et al.

(2022), who study the same regulation for breakfast cereals, chocolates, and cookies. They exploit the staggered rollout of labels across different stores before the law went into effect. Therefore, their research design allows them to study the effect of labels even if all products within the category received a label. They find a null effect of food labels in chocolates and cookies, in which labels are likely to offer no new information to consumers.

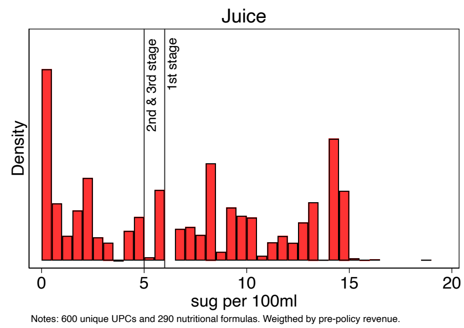
4.4. *Heterogeneity in the cost of reformulation*

Finally, we examine supply-side responses to the regulation. We look at the extent to which firms reformulated their products in different categories due to the introduction of food labels. We focus on reformulation in sugar and restrict our attention to categories in which the distribution of sugar concentration is neither entirely to the left of the regulation threshold nor too far to the right, such that it would be not feasible to modify the nutritional content up to threshold levels. This gives us a total of 13 categories representing 14.7% of the pre-policy revenue of all food products and 53.8% of the pre-policy revenue of products to the right of the threshold for which we collected nutritional content data. We discuss further details of the selection of categories and sample coverage in Appendix C.

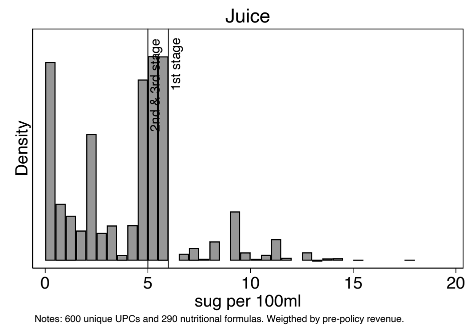
To assess behavioral responses on the supply side, we compare the distribution of nutritional content before and after the policy was implemented and explore bunching at the regulatory thresholds. Figure 8 plots the distribution of sugar concentration in 2016 and 2018 for products in the juice and cereal categories. The size of the bars represents the pre-policy revenue of the products included in that bar. Each subfigure also includes three vertical lines that indicate the thresholds of the policy in each of the three stages.

Before the introduction of food labels, we do not observe any noticeable pattern of bunching at any of the thresholds, in either the juice or cereal categories. However, in 2018 we find some evidence of bunching in both categories. In juice, we find that whereas in 2016 more than half of the products had a sugar concentration per 100 ml above the first-stage threshold, the distribution shifted to the left, and most products avoided the first-stage labels. In cereal, we also find that some products were reformulated in order to be on the left side of the threshold. Nevertheless, reformulation occurred to a much lesser extent than in juice.

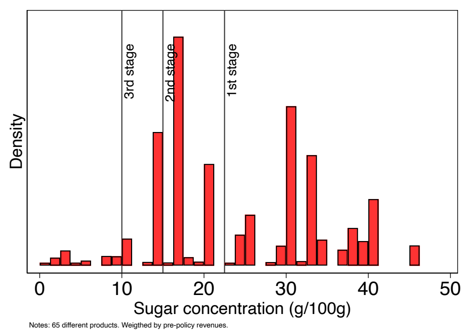
In Figure 9, we summarize the findings for all categories by plotting the (weighted by pre-policy revenue) share of the products in each category that surpassed the sugar threshold in the pre-policy period and were reformulated to be to the left of the threshold in the post-policy period. We show the histograms for each category in Online Appendix A, Figures A.4 and A.5.



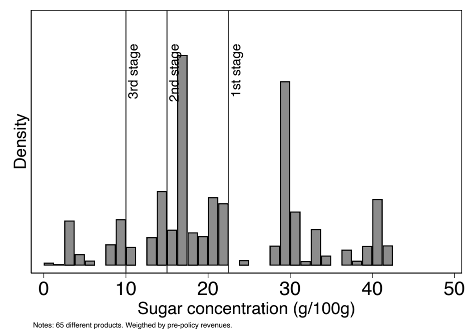
(a) Juice 2016



(a) Juice 2018



(b) Cereal 2016



(b) Cereal 2018

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

Notes: This figure plots the distribution of sugar concentration for juice and cereal before and after policy implementation. Observations are weighted by pre-policy revenue.

We find that while in some categories 100% of the products were reformulated to cross the regulatory threshold, in other categories less than 10% of the products were reformulated. Three important features of a product category can affect the extent to which products were reformulated. The first is the response firms expect on the demand side to the presence of labels. In categories with close substitutes and in which labels can provide more information, the returns to reformulation are higher. The second feature is the distance between products' current nutritional content and the regulatory threshold. The third is the cost at which firms can reformulate products without substantially affecting their quality. 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 present a larger share of products that bunch.

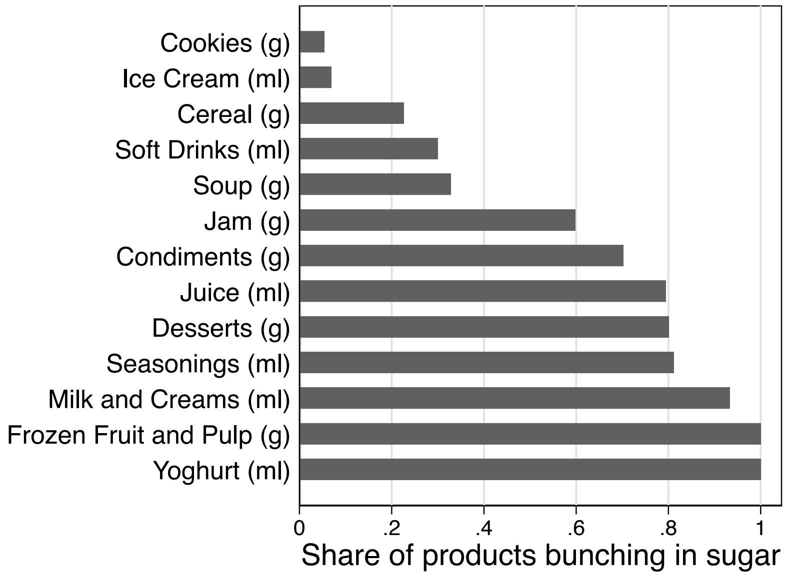


Figure 9: Share of products bunching in sugar

Notes: This figure summarizes the findings for bunching from Online Appendix A, Figures A.4 and A.5. It shows the pre-policy revenue weighted share of products to the right of the threshold in sugar in the pre-policy period that reduced the concentration of sugar to be to the left of the regulatory threshold in the post-policy period. The products used represent 54% 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 C.

5. Discussion

The Chilean Food Act suggests that FOP warning labels have the potential to reduce the overall intake of calories and sugar. We use access to rich micro-data—the universe of Walmart transactions in Chile between 2015 and 2018—and perform several empirical exercises to unpack the mechanisms through which labels affect consumer and firm behavior to inform policy design.

We find that labels are ineffective in shifting consumption across product categories. In other words, we do not find evidence of substitution from unhealthy to healthy categories. Instead, most of the policy effects arise from substitution within a product category. The lack of between-category substitution implies that food labels should be designed to focus on effects within specific categories. For instance, labels should not be designed to maximize substitution from cereal to fruit but instead from unhealthy to healthy cereal. We also find that labels are substantially more effective in product categories in which beliefs about product healthiness are poorly calibrated. Finally, we find that labels have the potential to promote product reformulation across several product categories and that these responses are highly heterogeneous.

Given these results, policymakers choosing regulatory thresholds for labeling policies should take into account several dimensions. First, it is important to target categories that represent a large share of consumers' intake of critical nutrients. Second, thresholds must be set to maximize substitutability within targeted categories, which means that (a) the categories need to have both healthy and unhealthy products that are close substitutes, and (b) consumers need to be misinformed about the health status of those products. Third, thresholds must be tighter in targeted categories in which reformulation is feasible at a low cost. Finally, future research should explore the effectiveness of implementing category-specific thresholds to maximize demand- and supply-side responses and study how these effects compare with a labeling system that may be more difficult for consumers to understand.

Our results also shed light on the importance of combining different policies to tackle obesity. In categories such as chocolates and candy, in which all products receive labels and are known to be high in critical nutrients, food labels are less effective for improving diet quality. Other market imperfections, such as lack of self-control or time inconsistency, can be important drivers of consumer bias in these categories. As Barahona et al. (2022) show, sugar taxes can be a better tool to fight obesity in those cases. Consequently, food labels and sugar taxes should be seen as complementary policies rather than substitutes.

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*, Forthcoming.

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., 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*, forthcoming.

Barahona, N., C. Otero, and S. Otero (2022). Equilibrium effects of food labeling policies. Mimeo.

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

Corvalán, C., M. Reyes, M. L. Garmendia, and R. Uauy (2019). Structural responses to the obesity and non-communicable diseases epidemic: Update on the chilean law of food labelling and advertising. *Obesity Reviews 20*(3), 367–374.

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., 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.

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.

Jáuregui, A., J. Vargas-Meza, C. Nieto, A. Contreras-Manzano, N. Z. Alejandro, L. Tolentino-Mayo, M. G. Hall, and S. Barquera (2020). Impact of front-of-pack nutrition labels on consumer purchasing intentions: a randomized experiment in low- and middle-income mexican adults. *BMC Public Health 463*, 1471–2458.

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.

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.

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.

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*.

OECD (2019). *OECD Reviews of Public Health: Chile*.

Pachali, M. J., M. J. Kotschedoff, A. van Lin, B. J. Bronnenberg, and E. van Herpen (2022). How do nutritional warning labels affect prices? *Working Paper*.

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

Reyes, M., L. S. 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*(2), e1003015.

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.

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.

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

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.

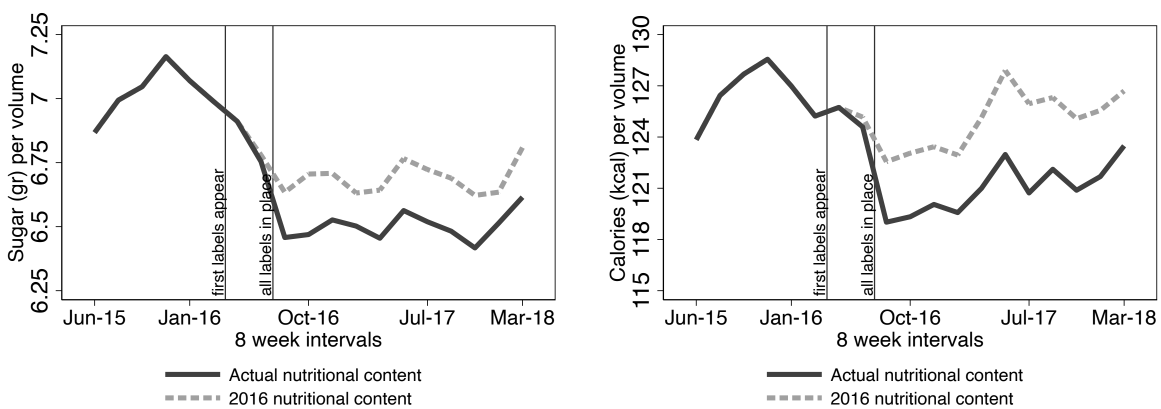
Online Appendix for:

On the Design of Food Labeling Policies

Nano Barahona Cristóbal Otero Sebastián Otero Joshua Kim

April 18, 2022

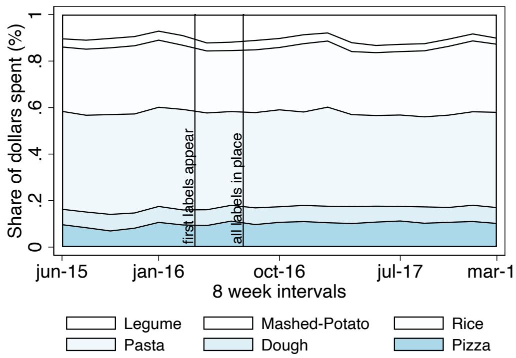
Appendix A: Additional Figures



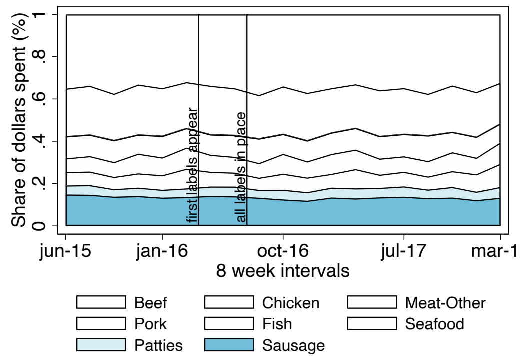
(a) Sugar intake (b) Calorie intake

Figure A.1: Sugar intake per grams and milliliters consumed before and after the policy

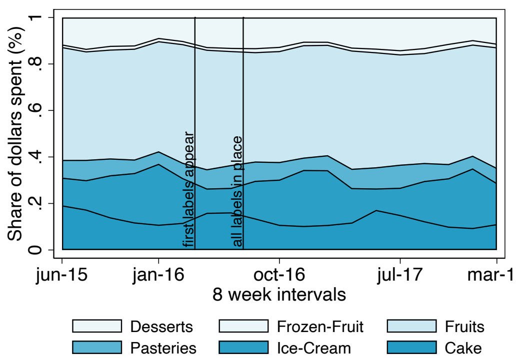
Notes: This figure shows the changes in nutritional intake per volume/mass of food products purchased at Walmart. For volume, we calculate the total amount of kilograms and total liters of products purchased at Walmart. We then divide the total intake of sugar by the total volume/mass of products. Measures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



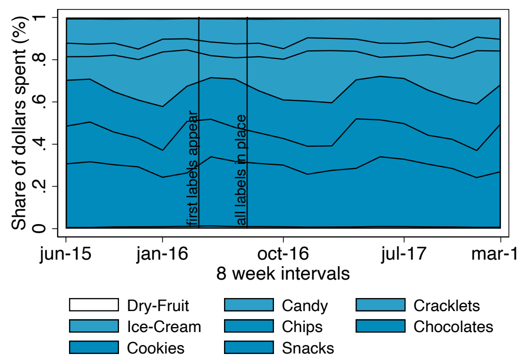
(a) Carbs



(b) Meat



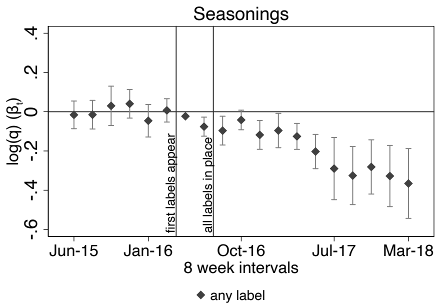
(c) Dessert



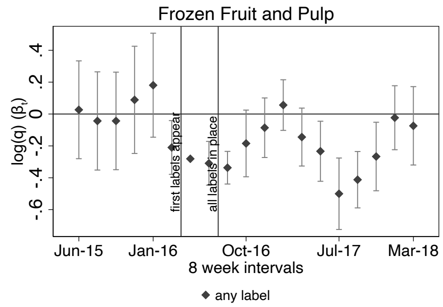
(d) Snacks

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

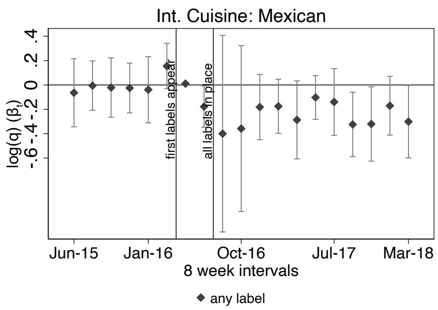
Notes: This figure shows the evolution of the share of dollars spent in each category within broader groups of products. Colors represent the share of labeled products within each category. White areas are categories in which no product received a label, dark-blue areas (e.g. snacks) are categories in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



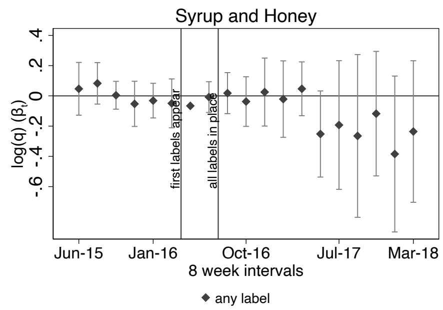
(a) Seasonings



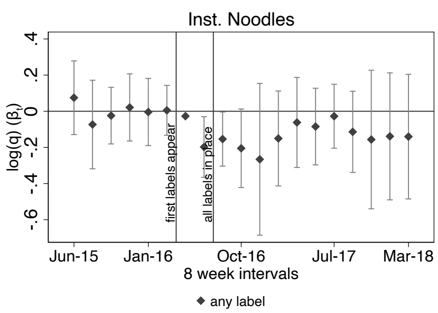
(b) Frozen Fruit and Pulp



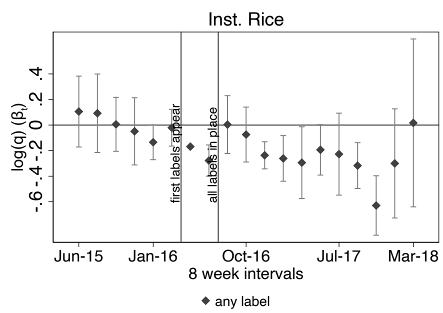
(c) Int. Cuisine



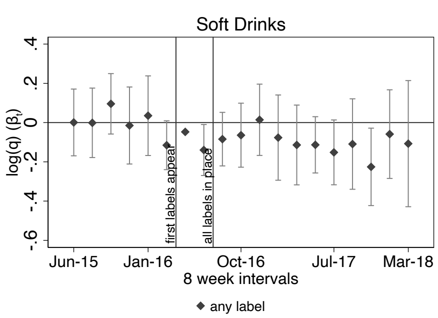
(d) Syrup and Honey



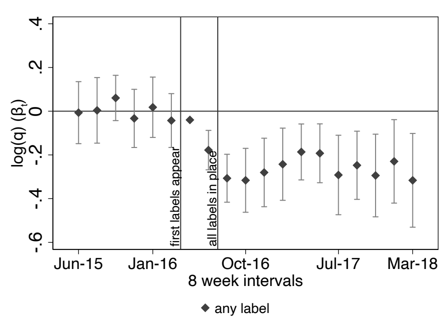
(e) Inst. Noodles



(f) Inst. Rice



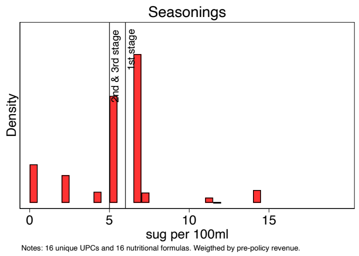
(g) Soft Drinks



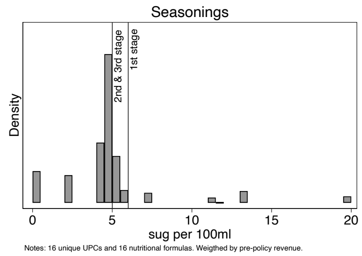
(h) Cereal

Figure A.3: Event study by category

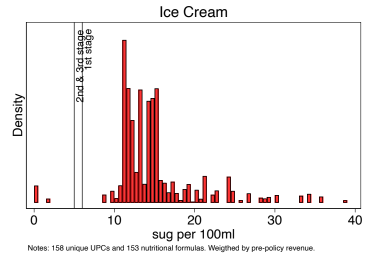
Notes: This figure presents the coefficients of our event study regressions. Each panel presents the β_t coefficients from Equation (2) for selected categories.



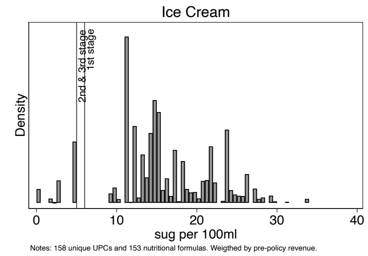
(a) Seasonings 2016



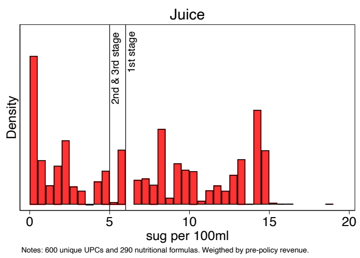
(a) Seasonings 2018



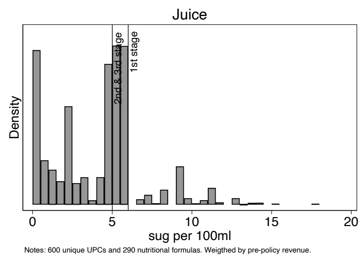
(b) Ice Cream 2016



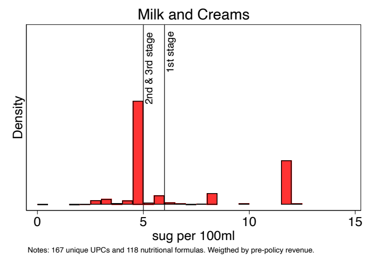
(b) Ice Cream 2018



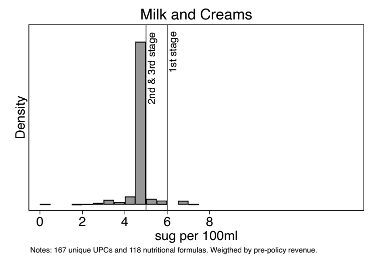
(c) Juice 2016



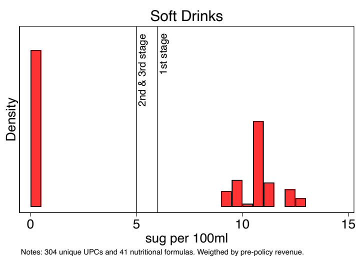
(c) Juice 2018



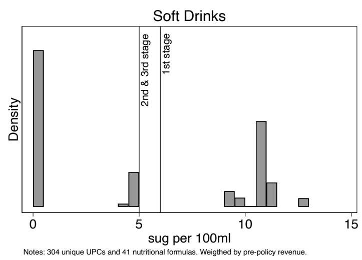
(d) Milk and Creams 2016



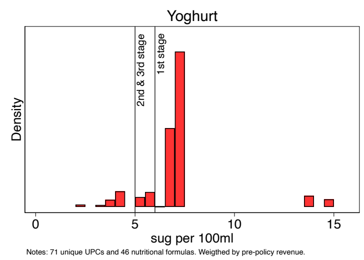
(d) Milk and Creams 2018



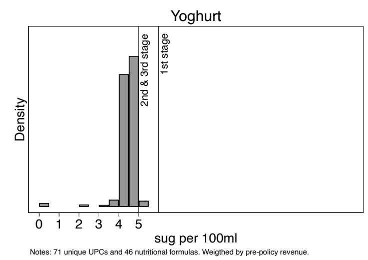
(e) Soft Drinks 2016



(e) Soft Drinks 2018



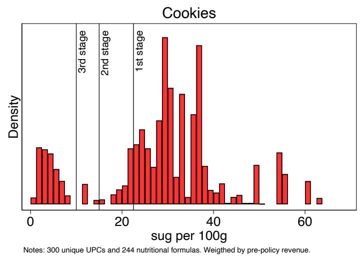
(f) Yoghurt 2016



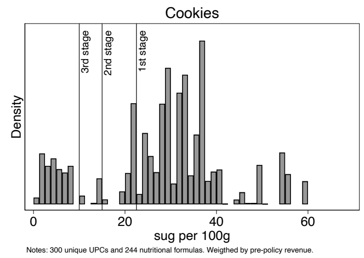
(f) Yoghurt 2018

Figure A.4: Distribution of sugar content pre- and post-policy for liquids in selected categories

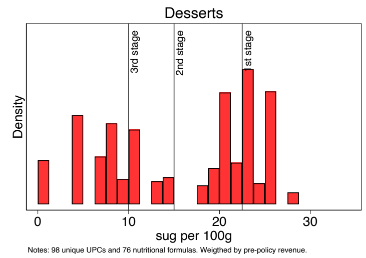
Notes: This figure plots the distribution of sugar concentration for liquid products in different categories before and after the policy implementation. Observations are weighted by pre-policy revenue.



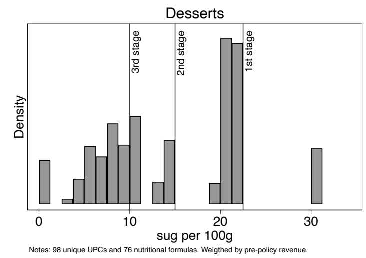
(g) Cookies 2016



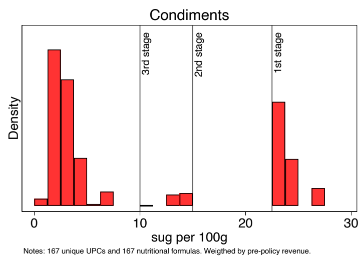
(g) Cookies 2018



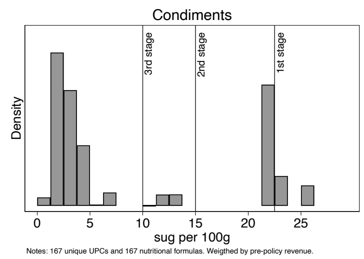
(h) Desserts 2016



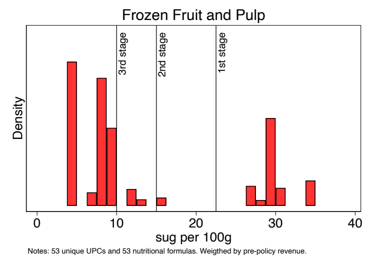
(h) Desserts 2016



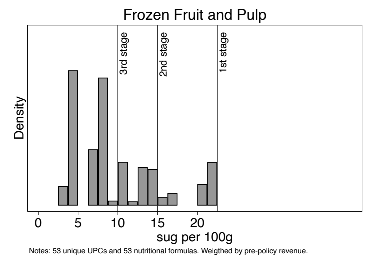
(i) Condiments 2016



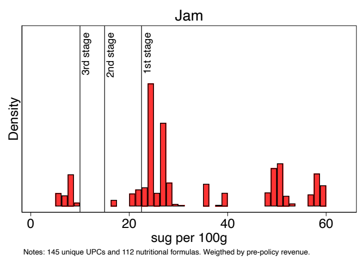
(i) Condiments 2016



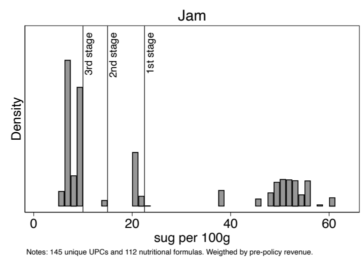
(j) Frozen Fruit and Pulp 2016



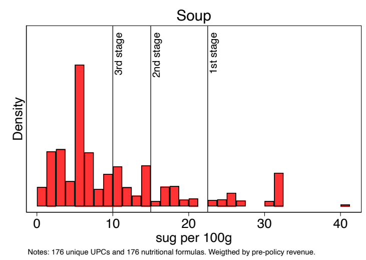
(j) Frozen Fruit and Pulp 2018



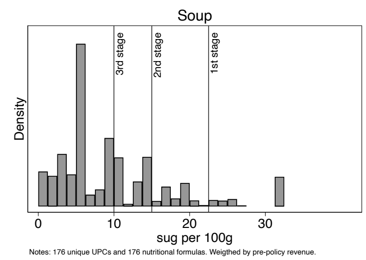
(k) Jam 2016



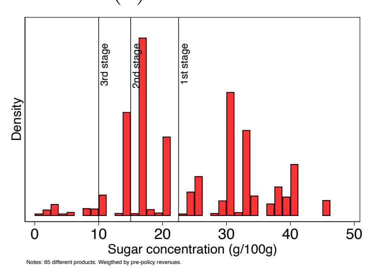
(k) Jam 2018



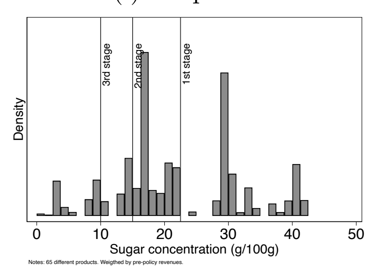
(l) Soup 2016



(l) Soup 2018



(m) Cereal 2016



(m) Cereal 2018

Figure A.5: 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.

Table A.1: FOP warning-label policies - Sources

Country	Status	Year	Sources
Chile	Implemented	2016	Legislation (Spanish)
Peru	Implemented	2019	Legislation (Spanish), Other (English), Other (Spanish)
Israel	Implemented	2020	Legislation (Hebrew), Other (English)
Mexico	Implemented	2020	Legislation, Other (Spanish), Other (English), Other (English)
Uruguay	Implemented	2021	Legislation (Spanish), Legislation (Spanish), Other (Spanish), Other (English)
Brazil	Approved	2020	Legislation (Portuguese, Legislation (Portuguese), Other (English)
Argentina	Approved	2021	Legislation (Spanish), Other (English)
Colombia	Approved	2021	Legislation (Spanish), Legislation (Spanish), Other (English), Other (English)
Venezuela	Approved	2021	Legislation (Spanish), Other (English)
Antigua and Barbuda	In discussion	-	Regional Standard (English), Other (English)
Bahamas	In discussion	-	Regional Standard (English), Other (English)
Barbados	In discussion	-	Regional Standard (English), Other (English)
Canada	In discussion	-	Government plan (English), Government consultation (English)
Costa Rica	In discussion	-	Law proposal (Spanish), Other (Spanish)
Dominica	In discussion	-	Regional Standard (English), Other (English)
El Salvador	In discussion	-	Law proposal (Spanish)
Guatemala	In discussion	-	Law proposal (Spanish)
Guyana	In discussion	-	Regional Standard (English), Other (English)
Haiti	In discussion	-	Regional Standard (English), Other (English)
India	In discussion	-	Other (English), Other (English)
Jamaica	In discussion	-	Regional Standard (English), Other (English)
Panama	In discussion	-	Law proposal (Spanish), Other (Spanish), Other (Spanish)
Paraguay	In discussion	-	Law proposal (Spanish), Other (Spanish)
Saint Kitts and Nevis	In discussion	-	Regional Standard (English), Other (English)
Saint Lucia	In discussion	-	Regional Standard (English), Other (English)
Suriname	In discussion	-	Regional Standard (English), Other (English)
Trinidad and Tobago	In discussion	-	Regional Standard (English), Other (English)

Notes: This table reports the sources for Table 2. The table includes countries that have laws or government resolutions already implemented, approved but not implemented, or under discussion. Countries with discussions of the topic that do not have a law proposal or discussion at the government level to establish mandatory FOP warning labels are not included. For countries with approved or implemented policies, "Year" indicates the date the policy was approved or the first stage of implementation began.

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 meets three criteria: (a) products are sufficiently similar, such consumers would consider substituting from one to another as a result of the regulation; (b) there is sufficient variation in terms of the share of products that receive a label; and (c) for the purpose of estimating a differences-in-difference model, we would need unlabeled and labeled products within a category to follow similar pre-trends in the absence of the policy.

Examining the categories that meet these conditions in our data is not straightforward. First, the product categories in our data are defined by Walmart for administrative and internal processes, and in many cases they include products that are not necessarily substitutes. Second, about 35% of total revenue comes from products that belong to a category with significant variation in exposure to labels (as defined by having less than 90% of labeled and unlabeled products). Third, most products within Walmart's categories do not follow parallel trends. To deal with these issues, we selected and combined certain categories, and within those we restricted our analysis to products that have the potential to behave as close substitutes. We also visually inspected and kept those in which labeled and unlabeled products followed similar pre-trends.

B.2. *Sample Coverage*

In Table B.1, we show the share of the revenue covered by the categories included in our demand-side analysis. Column (1) displays the share of total revenue represented by 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 are either below or above the policy threshold in the post-policy period and label them "Mostly unlabeled" and "Mostly labeled". Together, these two groups represent close to 63% of the revenue, and include categories such as fruit, meat, salads, candy, and chocolate. Another 30% of total revenue corresponds to other categories in which labeled and unlabaled products did not follow similar pre-trends.

Some products in these categories include pastry, bakery products, cold cuts, and biscuits. Our selected categories cover the remaining 5.7% of total revenue.

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

	Market share	Share labeled	Market share within labeled products
	(1)	(2)	(3)
Included	5.7	47.7	11.9
Cereal	1.4	62.7	3.6
Frozen Fruit and Pulp	0.1	11.2	0.1
Instant Noodles	0.1	47.3	0.1
International Cuisine: Mexican food	0.1	27.6	0.1
Instant Rice	0.1	27.1	0.1
Seasonings	0.5	57.5	1.2
Soft Drinks	3.3	42.4	6.5
Syrup and Honey	0.1	42.4	0.2
Not Included	94.3	21.2	88.1
Mostly unlabeled	57.4	1.2	1.6
Mostly labeled	6.3	99.2	25.8
Others	30.6	46.9	60.7
Total	100	22.8	100

Notes: Column (1) presents 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) presents the share of labeled products within each of the categories. Column (3) presents the share of total revenue for labeled products.

Column (2) reports the share of products, weighted by revenue, that received a warning label within a given category. Column (3) reports the share of total revenue revenue for products that received a label. When focusing on labeled products, our working sample comprises around 12% of the pre-policy revenue of labeled products in the total sample.

Appendix C: 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.

C.1. *Selection of Categories*

Organizing the categories for this exercise requires different criteria than the ones to study demand substitution in Appendix B. We focus on categories in which the distribution of sugar and calories are 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 modify the nutritional content up to the threshold level.

C.2. *Sample Coverage*

In Table C.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 67% of revenue, and include categories such as fruit, meat, salads, candy, and chocolate. Another 17.5% of total revenue is for categories with products that are exempted from the regulation (e.g., nuts) or products for which we are missing the pre-policy nutritional content. Our selected categories cover the remaining 15.5% 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 revenue for all products that are above the sugar threshold in the pre-policy period. When focusing on products with the potential to bunch, our working sample comprises around 54% of the pre-policy revenue for them.

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

	Market share (1)	Share above the sugar threshold before the policy (2)	Market share within products above the sugar threshold (3)
Included	15.5	63.2	53.6
Cereal (g)	1.4	43.5	3.1
Cookies (g)	2.1	77.3	10.7
Desserts (g)	0.6	31.2	1.3
Condiments (g)	0.4	30.9	0.9
Seasonings (ml)	0.1	49.0	0.1
Frozen Fruit and Pulp (g)	0.1	28.4	0.2
Ice Cream (ml)	0.9	98.2	6.0
Jam (g)	0.4	83.2	2.2
Juice (ml)	2.6	54.0	9.4
Milk and Creams (ml)	2.5	92.0	5.4
Soft drinks (ml)	3.3	53.7	11.5
Soup (g)	0.6	11.4	0.2
Yogurt (ml)	0.5	83.0	2.6
Not Included	84.5	10.0	46.4
Mostly below	63.8	0.3	1.0.1
Mostly above	3.2	98.0	17.2
Others	17.5	29.3	28.1
Total	100	18.28	100

Notes: Column (1) presents 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) presents the share of labeled products within each of the categories. Column (3) presents the share of total revenue for labeled products.