Received: 20 July 2021 | Revised: 19 July 2022 | Accepted: 25 July 2022

DOI: 10.1002/hec.4586

Health Economics WILEY

RESEARCH ARTICLE

The effect of soda taxes beyond beverages in Philadelphia

Felipe Lozano-Rojas[1] | **Patrick Carlin**[2]

[1]School of Public and International Affairs, University of Georgia, Athens, Georgia, USA

[2]O'Neill School of Public and Environmental Affairs, Indiana University Bloomington, Bloomington, Indiana, USA

Correspondence
Felipe Lozano-Rojas, School of Public and International Affairs, University of Georgia, Athens, Georgia, USA.
Email: flozano@uga.edu

Abstract

Soda taxes are implemented in several cities across the United States (US) with the aim of reducing sugar intake from sugar sweetened beverages (SSBs). Sugar is linked to obesity and to higher risk of diabetes and cardiovascular conditions. Sodas are targeted with these taxes since they are the main source of sugar for consumers in the US. In presence of potential substitutes, the policy can be undermined by consumers changing their sources of sugar. We examine the heterogeneous effects of the 2017 Philadelphia soda tax on purchases of other items containing sugar. We present an empirical evaluation focusing on the potential substitution toward additional sugary foods in Philadelphia and counties bordering Philadelphia. We find an increase in sugar from purchases of sweetened foods of about 4.3% following the introduction of the tax in Philadelphia and of 3.7% in the neighboring localities. The substitution to sugary foods in Philadelphia offsets 19% of the decrease of sugar from SSBs. Additionally, we find that the substitution offsets 37% of the decrease of sugar from SSBs when including counties bordering Philadelphia. These results suggest that while SSB taxes might be effective at lowering consumption of SSBs, substitution patterns may limit the effectiveness of the tax to reduce overall sugar intake.

KEYWORDS
excise taxes, health policy, local governments, public finance, soda taxes

JEL CLASSIFICATION
H21, H22, H23, H71, H75, I12, I14, I18

1 | INTRODUCTION

Taxes on sugar sweetened beverages (SSBs) have been increasingly enacted in the United States (US) because of their links to negative health outcomes such as obesity and increased risk of suffering from diabetes and cardiovascular conditions (Colchero et al., 2016; Han & Powell, 2013; Ogden et al., 2014; Roache & Gostin, 2017). By increasing the price of these goods, governments aim to generate revenue and to improve public health by lowering demand for SSBs. This policy has proven successful in the case of tobacco and alcohol. However, there are plenty of available substitutes to SSBs to as sources of sugar to maintain caloric intake. We explore several possible substitutes in this paper.

American cities began enacting SSB taxes in 2015. Berkeley first passed a 1 cent per fluid ounce (oz) tax on SSBs. Then, in 2017, Philadelphia enacted their own 1.5 cents per oz tax on sweetened and artificially sweetened beverages. Philadelphia was the first city to include artificially sweeteners. These taxes have been lauded for causing a drop in consumption of SSBs as well as increasing revenues (Cawley et al., 2019). However, these effects could be mitigated by cross-border shopping (Cawley et al., 2019a,b; Roberto et al., 2019; Seiler et al., 2021; Silver et al., 2017) and substituting toward beverages with higher concentrations of sugar due to the design of the policy, which levies the tax on the liquid and not on the sugar content (Allcott et al., 2019b; Grummon et al., 2019; Lozano-Rojas, 2018). In this paper, we analyze another potential margin of substitution-sugary foods. Specifically, we analyze items that are considered as the main sources of sugar by the American Heart Association (AHA) to see how sales of these foods change after the introduction of the tax. This margin has received considerably less attention with a few exceptions (Gibson et al., 2021; Restrepo & Cantor, 2020; Zhen et al., 2014). The AHA's list is comprehensive and includes traditional sources, such as desserts, cookies and sweeteners, as well as less traditional ones, such as condiments and dressings.

We use point-of-sale (POS) data to determine the effect of the SSB tax on sugary foods in Philadelphia. First, we analyze the changes in consumer behavior. Then, we show that prices were not affected by the tax. We use a back-of-the-envelope calculation to compare the decrease in sugar consumed through SSBs to the increase in sugar consumed by sugary foods. We test the overall effect of the policy by including both Philadelphia and neighboring localities as treatment. We find about a 4.3% increase in sugar sold from sugary foods in Philadelphia stores, which corresponds to approximately 19.1% of the decrease in sugar from SSBs sold. We also find an additional 17.5% offset from increased purchases of sugary items in neighboring localities, where the increase in purchases of sugary items is 3.7%. Finally, we also explore if there was heterogeneity among potential substitute foods items. The next section provides a brief literature review. Section 3 discusses the data, and Section 4 the methodology. Section 5 presents the results. Finally, Section 6 discusses the results and concludes.

2 | LITERATURE REVIEW

Sin taxes are taxes levied against items deemed undesirable. They serve the dual purpose of raising revenues and lowering demand for potentially harmful substances. Before the passage of SSB taxes in the US in 2015, the most common sin taxes were on tobacco and alcohol. Economic models that allow for irrational behavior due to self-control problems motivate the benefits of sin taxes (Allcott et al., 2019a; Haavio & Kotakorpi, 2011; O'Donoghue & Rabin, 2003, 2006). In the context of alcohol, tobacco, and possibly SSBs, self-control may be limited by addiction. The government can play a role by helping citizens moderate their consumption.

In the US, Berkeley, CA introduced the first SSB tax focused on altering consumption behavior in 2015. Albany, CA; Oakland, CA; Boulder, CO; and Philadelphia, Pensylvania (PA) passed their own in 2017, while San Francisco, CA and Seattle, WA followed in 2018. Cook County, IL which is home to Chicago also passed an SSB tax in 2017 but repealed it in 2018 (Cawley et al., 2020a; Chriqui et al., 2020). Washington D.C. was the last city to implement one, with an ad valorem charge of 8% over the sales value, 2% on top of the overall sales tax rate of 6%. This is different from the previous soda taxes which were per unit. Proponents of SSB taxes argue they encourage healthier behaviors from decreases in consumption of SSBs and raise substantial revenue. On the other hand, opponents believe the taxes are regressive, economically damaging, and ineffective given ample avenues of substitution (Marlow, 2019). Therefore, since 2015, there has been a growing literature examining SSB taxation. Research generally falls into four categories: changes in SSB quantity purchased/consumed, pass-through of the tax, distributional effects, and changes in consumer behavior.

Policymakers often specifically state that the goal of the tax is to reduce SSB consumption as a way to improve health (Chriqui et al., 2020; Marlow, 2019). There have been substantial differences in results depending on the city of implementation as well as choices of data and control group. For example, Falbe et al. (2016) found SSB purchases fell by 21% in Berkeley using Oakland as a control and survey data. Silver et al. (2017) found a 10% decrease in SSBs sold in Berkeley using the entire Bay Area as a control group and POS data. Lee et al. (2019) examined how often Berkeley residents were consuming SSBs using survey data and comparing to San Francisco and Oakland. They found that residents cut their consumption of soda by 0.55 times a day for a relative decline of 52.5%. In Philadelphia, Zhong et al. (2018) used a phone survey and found that within 2 months of the tax's implementation, the odds of having consumed soda that day were down by 40% and odds of having consumed energy drinks were down by 64% while finding negative but non-significant effects on the amount of consumption. Roberto et al. (2019) surveyed stores in Philadelphia and Baltimore and reported a 51% decrease in quantity SSBs sold. This effect was reduced to 38% once the authors accounted for cross-border shopping. One important limitation is that their analysis only accounted for cross-border shopping in Pennsylvania though New Jersey also borders Philadelphia. Bleich et al. (2021) similarly found a 42% decrease (6.1 fL oz decrease) in purchases of taxed beverages in Philadelphia. Cawley et al. (2019) used survey data and neighboring localities as controls and found the tax lowered sales by 31% for adults and almost no effect for children. Cawley et al. (2020a) and Cawley et al. (2020b) used self-collected data at local stores and found that there was reduced availability of SSBs and increased availability of water in Philadelphia, while there was no significant change in Oakland. Cawley et al. (2020c) specifically compared Philadelphia, Seattle, San Francisco, and Oakland and found Philadelphia experienced a far larger decline than the other four cities using a consumer panel from the marketing firm Infoscout. Finally, Seiler et al. (2021) finds that demand of taxed items falls about 46% in response to the tax, using POS data from a marketing company but that the reduction is only 22% when considering cross-border shopping.

The literature has estimated very different pass-through rates in different cities enacting SSB taxes, which is not surprising considering that the taxes are not the same. However, once again, differences may come from choices in data and methods. Falbe et al. (2015) found the pass-through rate in Berkeley to be from 47% to 67%, whereas Cawley and Frisvold (2017) found a pass-through of 43%. In Oakland, Cawley et al. (2020b) found a 60% pass-through using self-collected data store data. Meanwhile Cawley et al. (2021) studied Boulder by collecting data on at-register purchases and found that pass-through was 79.3%. In Seattle, different studies show pass-through rates of about 59% and up to 90% (Jones-Smith et al., 2020; Powell & Leider, 2020). Cawley et al. (2018) exploited the fact that the Philadelphia Airport straddles the border of Philadelphia to find that 2 months after passage, pass-through was 93%. Cawley et al. (2018) collected data in stores in Philadelphia and neighboring Pennsylvania towns and concluded that pass-through was complete (100%). Seiler et al. (2021) find a 97% pass-through, while Bleich et al. (2021) found a pass-through of 137%. The pass-through in Philadelphia appears to be much higher than in Berkeley and Oakland with Boulder falling between. Cawley et al. (2020a) expanded upon the analysis to suggest that pass-through was more common in smaller servings of SSBs, in high poverty neighborhoods, in stores farther from the border, and in independently owned stores. Other economic outcomes remains an understudied aspect of SSB taxation, although Lawman et al. (2019) used Pennsylvania Department of Labor data and concluded there was no change in employment in Philadelphia following the implementation of its SSB tax.

One concern about SSB taxation is that it has the potential to be regressive since people with lower incomes tend to spend a higher proportion of their incomes on SSBs (Allcott et al., 2019b; Cawley et al., 2019). However, the health benefits of an SSB tax are also concentrated among those with lower incomes. Cawley et al. (2019) discuss that policymakers may be less concerned if the taxes are regressive since SSBs have little to no nutritional value. Furthermore, many cities earmark revenues for specific expenditures that benefit the poor. For example, Philadelphia has earmarked revenues for prekindergarten programs and Berkeley for health education programs. In fact, Allcott et al. (2019b) includes an analysis where they factor in standard inequality aversion, regressivity of taxation, and progressivity of internality-reduction to find the optimal tax is approximately 1.5 cents per fluid ounce. It is difficult to empirically assess distributional effects because many of the benefits are long-term while the costs are short-term. Additionally, a full evaluation of the benefits might need to account for the potential success of the programs funded. With regards to changes in SSB consumption, there is some evidence of distributional effects. Bleich et al. (2021) found mixed results in sugar reduction by income. While they found that neither neighborhood income nor customer education significantly moderated calories of grams sold, they did find a significant reduction in total grams purchased from SSBs in stratified exploratory analyses.

Understanding consumer reaction to the SSB tax is vital since it will determine downstream health impacts and revenue collection. The reaction should be understood beyond changes in SSB consumption alone. In fact, Chriqui et al. (2020) used a qualitative review to understand why the Cook County SSB tax failed and found confusion of the goal (health or revenue) and the subsequent backlash, in part, led to the repeal. Alternatively, Taylor et al. (2019) found residents in Berkeley responded to media coverage and reduced consumption of SSBs prior to tax implementation. When looking at the potential effects over the untaxed margins, researchers argue that one of the most successful achievements of the policy has been the created incentive to substitute toward healthier beverages such as water. In this regard, Falbe et al. (2016) found that bottled water consumption increased 44% using self-reported data, while Silver et al. (2017) estimated the change in water consumption for bottled water at an increase of 15% for Berkeley using POS data. In Philadelphia, which has a higher tax per liquid ounce, Zhong et al. (2018) found that the probability of daily consumption of water increased by 58%, and Cawley et al. (2019) also found water consumption increases. These results are important in the sense that researchers are aware of the potential substitution effects. We will follow this line of inquiry and will examine potential substitutes that are unhealthy as they have additional sugars, margins that are required to have a complete evaluation of a policy such as a tax on SSBs.

Studying consumer responses can illuminate unintended consequences of these taxes. Cross-border shopping is an example of a potential leakage that can mitigate any positive health effects from reduced consumption and alter the sources of revenue for the cities enacting the policy and for their neighbors. Both Cawley et al. (2019) and Seiler et al. (2021) found economically and statistically significant leakage due to cross-border shopping in Philadelphia. So far, few studies tackle questions concerning additional sweetened foods. Zhen et al. (2014) estimated a structural demand system and forecasts decreases in calories and sugar intake despite an increase in consumption from additional sweetened foods that offsets nearly half of the decrease from soda. Restrepo and Cantor (2020) examined untaxed beverages in Philadelphia using survey data from National Health and Nutrition Examination Survey on adolescents and within-state variation on average sales tax rates levied on soda relative to other foods. They found that adolescents offset decreases in soda consumption with sweetened milks in up to 75% of the original decrease. Meanwhile, Bleich et al. (2021) and Gibson et al. (2021) did look directly at sugary food as part of their analyses. They found that there was no statistically significant increase in sugary food using Baltimore as a control group. Gibson et al. (2021) also did not find a statistically significant increase in spirits but did find substitution toward non-taxed beverage concentrates. More recently, Zhen et al. (2022) uses structural modeling and the instruments proposed by Allcott et al. (2019a) on survey data from FoodAPS and finds that soda taxes improve the healthfulness of families labeled as obese. This study provides one of the first causal measures of the cross-elasticities from soda over additional sugary foods.

A common statistical issue in the literature is the inference made from a unique cluster from the city taxed. Cawley et al. (2020c) urged caution when interpreting SSB results given that most analyses thus far have been unable solve this issue by using more than two geographic clusters. Our paper contributes to this literature by examining the SSB tax's effect on sugary foods using a large POS consumer panel database which allows us to select an appropriate control group. Furthermore, our analysis uses an extensive amount of data that reflect purchases of over 1000 food items. This paper establishes changes in both sugary foods sold and sugary food prices that result as a response to the soda tax in Philadelphia. In contrast to alcohol or tobacco, this focus on substitutes is relevant and unique to the context of SSBs as a “sinful” product, considering that sugar sources are abundant and found in many additional foods. Another contribution is that our data accounts for a larger set of consumers and of the total market of sugar in volume within the locality of Philadelphia and its neighbors, in comparison to self-reported data. Finally, we try to address concerns over inference with a small number of units treated by using different clustering mechanisms and ways of calculating bootstrapped standard errors as suggested by Bertrand et al. (2004) and placebo inference based on “fake” treatment assigned to non-treated units following the spirit of Conley and Taber (2011) and Hagemann (2019). Our results are robust to these choices. Our findings suggest that prices of sugary foods are not changed by the tax, but there is an increase in sugary foods sold, both in Philadelphia and in neighboring localities. This patterns offset between 21.8% and 39.9% of the reduction of sugar from SSBs we observe.

3 | DATA

We use two main sources of information. The first is the POS from the Nielsen Data, which provides consumption reports in units and prices for each of the items we study. The second is the nutrition content (sugar grams) of each item from online sources that report the Nutrition Facts. We provide a further description of these two sources in this section.

3.1 | Nielsen Data

We use information from Nielsen[1] Retail Scanner Data, provided by the Kilts Center for Marketing. This dataset captures POS retail scanner information from participating stores across the US at the three digit zip code level. For instance, Philadelphia's zip codes encompass all zips starting with 191. However, there are a few 5-digit zip codes within Philadelphia that start with 190 and 192, so they will not be included in our analysis, as most of their geographies fall outside of the range of the City of Philadelphia.[2] We then observe the Universal Product Code (UPC), the price, and the amount sold by each store.

The benefit of this dataset is its completeness and how representative it is in relation to the entire market. According to the description in its webpage, POS in the Nielsen Data captures “more than half the total sales volume of US grocery and drug stores and more than 30% of all US mass merchandizer sales volume”.[3] Since this information is automatically generated and low cost for stores to provide, there is not as much concern of under-reporting as in self-reported information sources, including surveys or even the Homescan Panel from Nielsen (Einav et al., 2010). Still, Nielsen is skewed toward bigger retailers as it only captures 2% of Convenience stores.

Furthermore, we do not observe all the purchases in the US. We miss information from restaurant sales of sugary items (unless they resupply their stock in Nielsen participating retailers), as well as information on participating stores' brands. In the case of soda, estimates for 2012 market shares reflect that close to half of the market is sold through retail stores, with an additional 12% occurring at convenience stores, 11% at vending machines, and 20% at Restaurant and Drinking places (Williams & Goldsworthy, 2012). We do observe retail stores and a smaller sample of convenience stores, while we miss vending and restaurants. We expect the concentration of the market in volume to be higher for retailers since restaurants charge a higher price per ounce. Therefore, we expect Nielsen to capture more than 50% of sales in terms of volume. We would expect to find even stronger concentrations for the market share of other sugary items in retail stores as opposed to restaurants, as consumers seldom purchased packaged snacks at restaurants with the exception of vending machines. In fact, 32% of the calories consumed by Americans come from away from home purchases, including 25% from restaurants (U.S. Department of Agriculture, 2018), and we would expect the vast majority of the remainder to come from retailers whose market share is represented in Nielsen.

We lack some information, and all of our estimates are local to the sales happening in the stores we observe over the set of UPCs we can retrieve nutritional content information on. We do not observe whether the items purchased are actually consumed. Still, we have assembled one of the most comprehensive datasets of purchases and nutritional contents to evaluate soda taxes.

We restricted our sample to the goods we were interested in studying: sugary foods and SSBs. The specific categories of foods we examined were raw sugar, jams, jellies and spreads, canned fruit, desserts and gelatin and syrup, crackers, cookies, and condiments. These groups were specifically cited by the AHA[4] for their particular importance in sugar consumption in the US. These items would address questions pertaining substitution throughout mechanisms of sugar craving and addiction, as opposed to other drinks where thirst quenching as a motive for substitution is confounded with sugar craving. Some foods seem more relevant than others as we will show. For instance, candies as sources of sugar consumption are more relevant than dressings. However, unconscious increases from different sources could compensate for the absence of sugar from SSBs, so we analyzed a broad scope of sugary foods. The reference from the AHA is comprehensive, including traditional sugary items, such as desserts, candies, cookies, jams and sweeteners; other important contributors such as bread, and cereals; and finally, less traditional items that also include sugars such as condiments and dressings.

We observed 388,178 unique UPCs in these categories for which we have 804.5 million weekly reports (observations) from January 2014 to December 2018. Out of this initial number, we discard control UPCs which are anonymized to protect stores identities, and other items for which we cannot reconstruct a full 12-digit UPC number, since we will use this information to obtain the nutritional information regarding sugar content.

3.2 | Nutrition facts

Once we obtained UPCs of goods sold, we developed a web-scraping script to retrieve nutrition facts information from the US Department of Agriculture (USDA) FoodData Central website (https://fdc.nal.usda.gov/). We constructed the commercially-used UPC by including the verification number and any remaining zeros to obtain complete 12-digit UPCs. Table 1 shows what percentage of observations and unique UPCs we were able to merge with nutrition information. Nielsen provides weekly report information for 492,584 UPCs. 189,813 UPCs are control brands, and 2775 UPCs have less than 9 digits, so we cannot retrieve their actual 12-digit UPC that includes its verification code. Thus, 291,826 UPCs were identifiable in our scrape. We successfully scraped nutrition information for 11,181 UPCs. Though that is only 3.8% of identifiable UPCs, they account for

TABLE 1 Data assembling information

	Obs. (In millions)		UPCs	
	No.	**%**	**No.**	**%**
Total	7038.9	100%	492,584	100%
Control brands	1072.4	15.2%	189,813	38.5%
Other unidentifiable	269.4	3.8%	2775	2.2%
Identifiable	5697.1	80.9%	291,826	59.2%
Successfully scraped from identifiable UPCs	**3523.1**	**61.8%**	**11,152**	**3.8%**

Note: Observations from the time span January 2014 to Dec 2018. Control Brands include items whose brand description is Control Brand or No Company Listed. Other unidentifiable are UPCs that had less than nine digits or more than 12 such that constructing the verification digit was not feasible. Percentages in the top panel use the Total as a baseline, percentages in the bottom panel use the Identifiable category as reference. Our results are local to the identifiable observations in bold font.

62% of the available and identifiable information at Nielsen (weekly reports), since we are able to merge nutrition content for 3.5 billion of our rows out of the 5.7 billion identifiable reports. The reason we can scrape a low percentage of UPCs but still a high percentage of transactions is because most of the goods sold are from a few specific brands.

The top-left and top-right panels in Figure 1 shows the share of each category in the total sales in transactions from our sampled data from Nielsen for which we can retrieve information. Baked goods (including bread), candies, and cookies were the items that represent the highest sales shares, in both transactions and total sugar quantity. The top-right panel in Figure 1 shows the amount of sugar kilograms sold per item. Meanwhile, the bottom panel in Figure 1 shows the concentration of sugar by group. Sweeteners are the most concentrated followed by syrups. The fact that candy, baked goods, and cookies are the largest contributors to sugar consumption is because they represent a large number of transactions rather than having a higher density of sugar. The red lines across all the graphs highlights the importance of the first four categories that account for 80% of the total sales of sugar from sweetened foods.

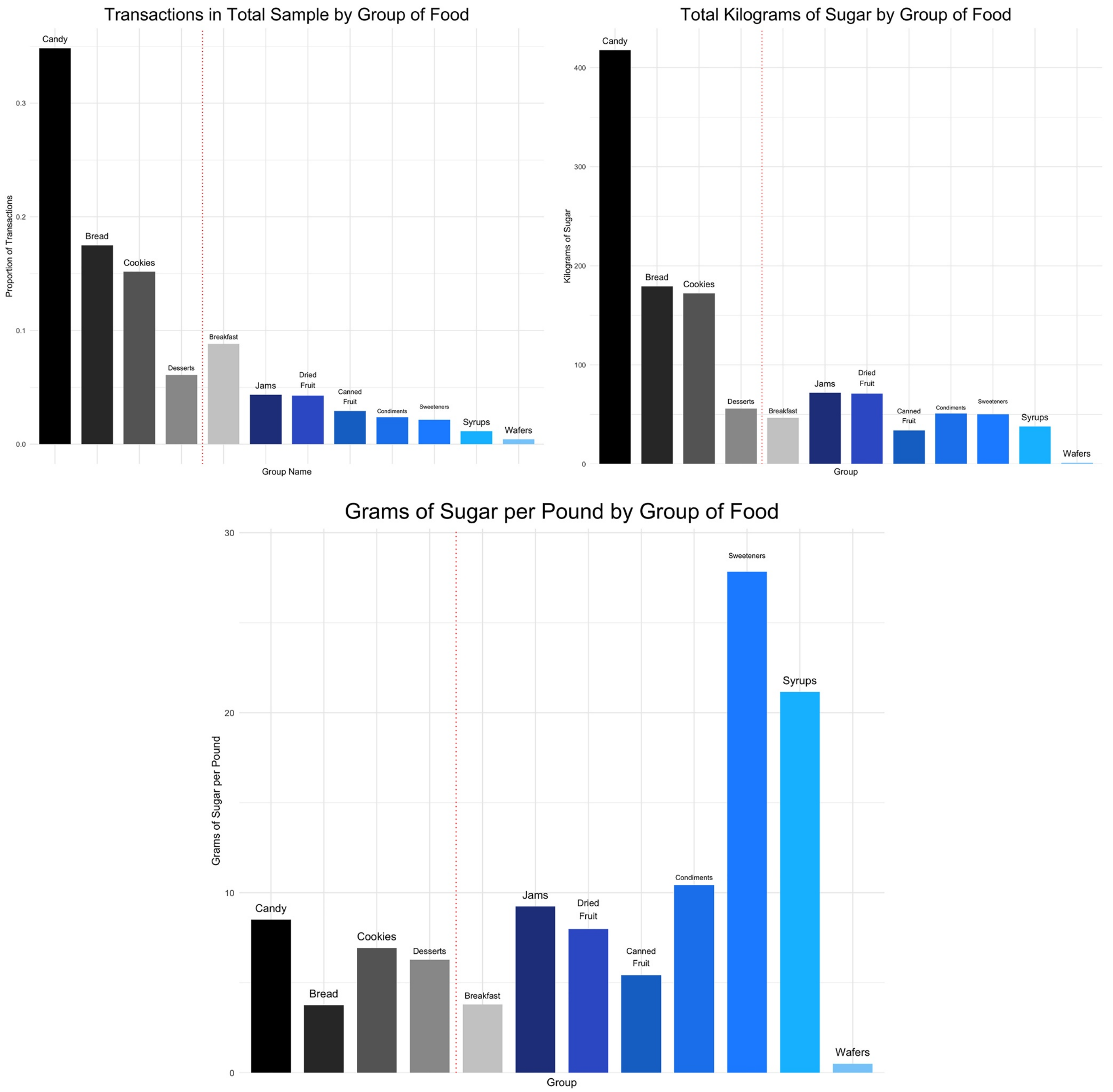

FIGURE 1 Number of transactions, total sugar, and sugar per pound by group of food [Colour figure can be viewed at wileyonlinelibrary.com]

The Nielsen data over-represent bigger retailers, which means popular, well-known brands, are over represented in our sample. Furthermore, we must drop observations which we cannot retrieve information from, either because they are not included in the online stores or in the USDA Food Central database, or because they are control brands anonymized in Nielsen. We caution that to extrapolate from our results, we must assume that substitution patterns hold for the unobserved brands. If people substituted to the unobserved brands that have higher sugar density in relation to our observed items, then our estimates would be conservative. We cannot formally test that the sugar included in unobservable items is low enough for them to cancel out our results. However, we do not expect them to differ dramatically since many brands within a food item compete in flavor and have about the same levels of sugar.[5] We are less concerned about additional sugary goods, as sugar purchases happen mostly in a handful of sub-products as shown above. Changes after the tax in the unobserved sugary foods would have to have be exorbitant to affect our estimates.

4 | METHODOLOGY

4.1 | Difference-in-differences

We use a standard difference-in-differences (DD) regression to analyze the effect of the policy on purchases of sugary foods and price changes for sugary foods as shown in Equation (1). Y_{st} is the log form of either grams of sugar or price per sugar gram depending on the specification, D_s is a dummy that is 1 if the store is located in Philadelphia, $Post_t$ is a dummy if the observation was after implementation in 2017. The model also includes store and year fixed effects and standard errors clustered at the store level. We also use this identification strategy while changing the definition of which localities are treated to estimate cross-border shopping and region affects.

$$y_{st} = \beta\,(D_s \cdot Post_t) + \eta_t + \eta_s + \epsilon_{st} \tag{1}$$

The coefficient of interest is β. Under certain assumptions, it can be interpreted as the effect of Philadelphia enacting its SSB tax on the amount of sugar from sweetened foods purchased or the price per sugar gram of sugary foods. We also conduct the same analysis for stores in neighboring localities to assess the extent that cross-border shopping can have on purchases of additional sugary foods. The fixed effects control for time invariant state trends and country-wide time varying trends. A positive β for sugar grams would be somewhat alarming for policymakers because it would indicate a substitution toward other sugary goods. Furthermore, a non-zero β for price per sugar gram would indicate that the policy affected prices of goods besides SSBs. The next subsection evaluates the main threat to validity for a DD, which is the parallel trends assumption.

4.2 | Event study analysis

Parallel trends is the primary assumption to interpret DD coefficients causally. This would imply that the differential between the series of the treated and control would have remained the same in the absence of the policy. To assess the trends pre-treatment and to examine any different effects over time post-treatment, we use event study analysis represented by Equation (2). y_{st} and D_s are defined as above. This specification estimates a coefficient of the effects for each period, and we use the period immediately before the implementation as reference. Before treatment, the coefficients should be around 0. If this is the case, it can be interpreted as follows: the distance between the series of both treatment and control is not changing, and it remains at the levels of the period before the treatment was implemented. After treatment, the coefficients will show any heterogeneity in effects over time.

$$y_{st} = \sum_{t \neq Treat-1} (\beta_t\,(D_s \cdot \eta_t)) + \eta_t + \eta_s + \epsilon_{st} \tag{2}$$

4.3 | Control groups

The Nielsen dataset is country-wide, and we can scrape nutrition facts from USDA for goods sold anywhere in the US. However, it is not clear that the rest of the country is a valid control group for Philadelphia. Furthermore, there is not a control group consistently used in the literature. For that reason, we decided to define several. First, we used stores in Pennsylvania, New Jersey, and Delaware that did not have three digit zip codes of 191 (treatment 3-digit zip) nor 190 (excluded because it overlaps Philadelphia and surrounding counties). However, Seiler et al. (2021); Cawley et al. (2018) show that there was cross-border shopping after Philadelphia passed its SSB tax. Therefore, our second control group excluded all 3-digit zip codes that neighbor the locality with 3-digit zipcode 191. Since 3-digit zip code boundaries are not the same as county boundaries, our third control group further excludes any 3-digit zip codes that neighbor excluded zip codes in control group 2. This means all 3-digit zip codes that are two borders away from 191 are excluded. We call this area ”metro” in the analysis, as most of them constitute the metropolitan area of Philadelphia and Camden. Finally, control group 4 excludes the same 3-digit zip codes as control group 3 but additionally includes all states bordering PA, New Jersey (NJ), and Delaware (DE). We fear that the neighboring localities are tainted by the policy, and to test it we run independent regressions using the neighboring localities as a treatment unit.

Figure 2 presents maps of the control groups. On the left, the dark blue states have 3-digit zip codes in each control group while the light blue states are only included in control group 4. The red rectangle shows the Philadelphia 3 digit zip codes as well as the area where 3-digit zip codes are excluded in certain control groups. The right panel is a detailed map of what is inside the red rectangle. The lightest blue 3-digit zip codes are included only in control group 1. The medium shade are included in control groups 1 and 2. The darkest blue 3 digit zip code is included in control groups 1, 2, and 3. None of the shown 3 digit zip codes are included in control group 4.

There are theoretical arguments for and against each of these control groups. The first control groups are more comparable since they are geographically close. However, they may be affected by the policy (i.e., through cross-border shopping). The latter control groups are plausibly unaffected by the policy but may not be as comparable. So, we include all four in our analysis and choose our preferred specification by picking the event study that best minimizes concerns about parallel trends and contagion from the policy. Section 5.1 presents the results of the Event Studies for each of these control groups. To further evaluate the extent to which nearby localities are affected by the policy, we estimate the effect of the policy in those zipcodes that are neighbors to Philadelphia. We use the coefficients in the pre-intervention period to assess how comparable a control group is to the treated observations. By adding the square deviation from zero, we can evaluate the different controls. This measure shows across all our estimates the trade-off between geographic proximity and comparability.

In a separate exercise available upon request, we matched Philadelphia to over 100 cities from the American Community Survey, and we find the nearest neighbors on the basis of income, population shares in race and ethnicity, and on traveling distance. In all cases, the sacrifice in comparability is substantial, with many more violations to the parallel trends assumption and worse overall comparison in the pre period. Many of these cities are geographically distant and constitute different markets. Nearby localities constitute the same market, some even as part of a metropolitan area (some zipcodes not neighboring Philadelphia are still part of the Nielsen denominated Designated Market Area). They face the same changing weather that might drive demand for SSBs and additional sugary items. Maybe more importantly, they also face the same advertisement and most likely the same set of products, as opposed to cities that are farther.

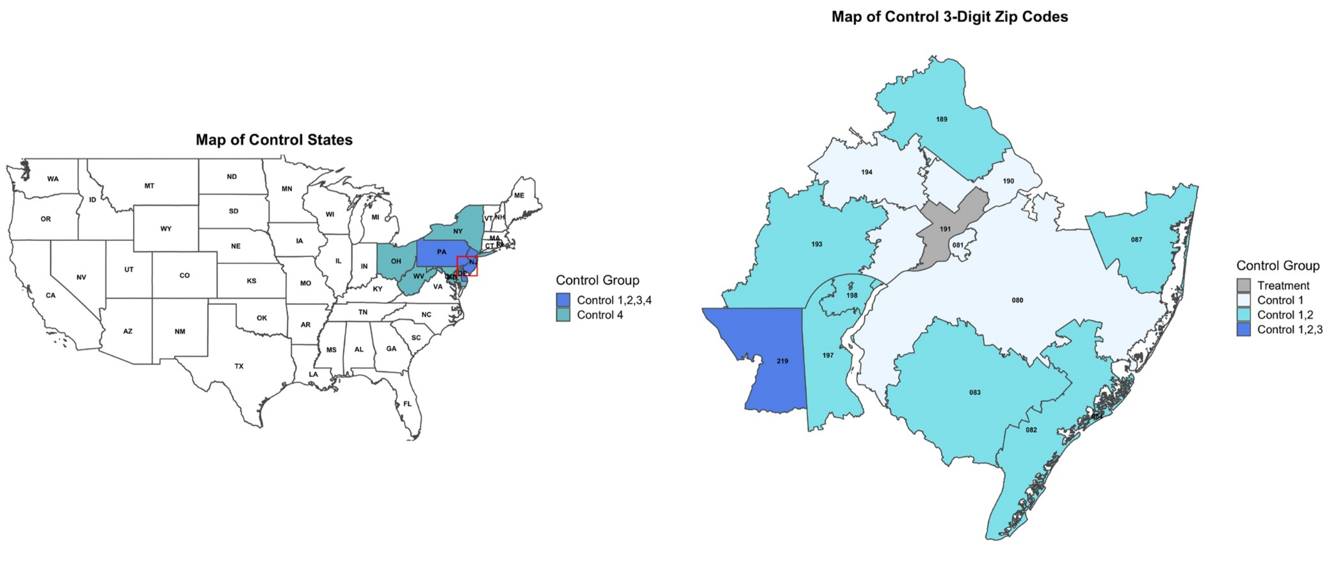

FIGURE 2 Map of control group states and 3-digit zip codes [Colour figure can be viewed at wileyonlinelibrary.com]

4.4 | Robustness tests

The first robustness exercise is the choice of control group. As we will show below, our results are qualitatively the same regardless of the choice of control group. However, the data has a limitation recognized in the literature: the fact that the inference we are making comes from the variation on one one policy lever (Bertrand et al., 2004; Conley & Taber, 2011; Hagemann, 2019). Bertrand et al. (2004) showed that DD renders inconsistent standard errors, and Conley and Taber (2011) mention how the problem is aggravated when making inference from analyzing a small number of treatment units. To address this issue, we conduct our analysis with skepticism about the choice of ideal clustering, which would usually be the 3 digit zipcode as that is how treatment is assigned in our dataset. Instead of using these results, we use clustering at the store level which renders a more conservative result. In Appendix A1, we present the same results with different clustering choices.

While the previous choice can be justified on the grounds of conservative standard errors, it seems rather arbitrary. Hence, we conduct additional exercises to assess the validity of our results. First we present bootstrapped confidence intervals, clustering at both levels of choice. We draw stores/3-digit zipcodes from the treatment and the control, repeating this exercise 1000 times to obtain a non-parametric distribution of the effects. In the second exercise, we implement a placebo inference test in the spirit of Hagemann (2019). This test randomizes across the control units and assigns a "fake" treatment. This procedure is repeated across some (or all) of the feasible permutations of control units. Since this is a "fake" treatment, the expectation is that the distribution would describe the randomness of a zero treatment effect. This distribution can be used for inference, and we show results for both choices of clustering. There is a strong effect clustering at the store level and a less than marginal effect when clustering at the 3-digit zipcode. Both the bootstrapped standard errors as well as the placebo inference distribution (when clustering at the store level) strongly support the idea that following the introduction of the soda tax in Philadelphia, there was a small but perceptible and relevant increase in sugar purchased from sugary foods and no change in prices of other sugar sweetened items. The results from the bootstrap and from the placebo inference exploit the simplicity of an empirical distribution from observations where we do not expect effects as a soda tax has not been implemented, without relying on a given assumption about the statistical distribution.

5 | RESULTS

This section summarizes the results from the specifications introduced above. First, we estimate the effect of the SSB tax on purchases of sugary foods as well as their prices for the DD specification. We follow with their Event Study results. In Section 5.2 we analyze the effect of the policy in neighboring localities to measure the extent of cross-border shopping activity following the introduction of the tax. We then develop a back-of-the-envelope calculation that establishes a comparison between the effect on sugar purchases from additional sugary items relative to the ones from sodas. In Section 5.4, we present the results of different robustness checks to challenge our original hypothesis and to address the concerns from the limited number of treatment units. The last subsection presents the results assessing heterogeneity of effects for each group of sugary foods in our analysis to address concerns regarding composition of sales and the heterogeneity of the goods we follow. All DD analysis follow the specification we described in Equation (1), and all event study analysis follow Equation (2).

5.1 | Results for sugar quantities and prices

Table 2 presents our main results in terms of prices and quantities. It presents the coefficients from Equation (1) and their statistical significance based on clustering at the store level, the pre-policy monthly baseline in Philadelphia stores,[6] and either the monthly average of sugar grams sold per store or average store price in cents per sugar gram for reference of the size of the effects. The table also presents the sum of squares and the deviation from zero of the coefficients of the event study in the pre-period. We include observations back to the first quarter of 2014. This measure captures how different the series of treatment and the alternative controls really are, and allows us to make comparisons of how good counterfactuals the different control groups are.

The choice of clustering is based on the conservative estimates of the effects resulting from this level. Table A1 in Appendix A1 shows the same results and Figures A1 and A2 show the analogous event studies as those presented in this subsection but clustering at the 3-digit zipcode level instead. The statistical significance of the effects becomes stronger (see Table A1),

TABLE 2 Effect of the Soda Tax on unitary sugar gram quantities and prices

PANEL A: Dependent variable Log sugar grams sold				
	(1)	**(2)**	**(3)**	**(4)**
	(PA,DE,NJ)	**-(Neighbors)**	**-(Metro)**	**+(adj.states)**
Control group	**ln (Sg.gr)**	**ln (Sg.gr)**	**ln (Sg.gr)**	**ln (Sg.gr)**
Philly · post	0.0363** (0.0113)	0.0425*** (0.0115)	0.0467*** (0.0117)	0.0245*** (0.0109)
Baseline	378,696	378,696	378,696	378,696
R^2	0.977	0.976	0.977	0.978
N	138,574	118,874	101,161	342,230
N. Clusters	3280	2868	2495	8111
Pre-Sq.Deviation	0.0335	0.0373	0.0440	0.0375
PANEL B: Dependent variable log prices per gram of sugar				
	(1)	**(2)**	**(3)**	**(4)**
	(PA,DE,NJ)	**-(Neighbors)**	**-(Metro)**	**+(adj.states)**
Control group	**ln (cents/Sg.gr)**	**ln (cents/Sg.gr)**	**ln (cents/Sg.gr)**	**ln (cents/Sg.gr)**
Philly·Post	0.0000146 (0.00288)	−0.000868 (0.00293)	−0.00216 (0.00293)	−0.00181 (0.00275)
Baseline	1.985	1.985	1.985	1.985
R^2	0.896	0.897	0.905	0.904
N	138,574	118,874	101,161	342,230
N. Clusters	3280	2868	2495	8111
Pre-Sq.Deviation	0.0063	0.0063	0.0061	0.0281

Note: Coefficients from Equation (1). Panel A presents the results over log quantities and panel B over log unitary prices. The baseline introduces the the average monthly sales of sugar grams from additional sweetened items (Panel A), or average store price per sugar gram in cents (Panel B), in the year prior to the policy implementation. Finally, the Pre-Sq.Deviation is the sum of squared coefficients from the event study prior to policy implementation (in the pre-period), such that it provides an evaluation of control quality. Standard errors clustered at the store level in parentheses.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

and the coefficient not compliant with the parallel trend assumption in the pre-period becomes more significant as well (see Figure A1).

The columns of Table 2 differentiate across the competing control groups. Model (1) includes all observation units from the three states where the Delaware Valley is located. The second model removes the units neighboring the 3-digit zipcode 191 that identifies Philadelphia. The third model, removes the entire Delaware valley units. The final column adds adjacent states without including the metropolitan area units.

5.1.1 | Results for sugar purchases from additional sweetened foods

Panel A of Table 2 presents the results for Equation (1) when the outcome variable is sugar grams from additional sweetened foods. The magnitude of these results depend on which control group is chosen. Importantly, all of the control groups suggest a small but statistically significant increase on the consumption of sugar from these items, which oscillates from 2.5% up to 4.7%. Our preferred specification, with other Pennsylvania, Delaware, and New Jersey three digit zips as controls, shows a 4.3% increase in sugar grams consumed from these foods. Subtracting neighbors and metro area does not qualitatively change this result, and if anything increases the difference. This might happen as cross-border shopping of some of these items might be complementary to sodas. Furthermore, in the neighboring localities an increase in consumption of additional sweetened foods happens concomitantly. Finally, adding sets of adjacent states makes the effects smaller.

Figure 3 plots the four event studies corresponding to the each of the control groups. Some of the coefficients prior to tax are statistically different from zero, suggesting either an anticipatory effect during the two quarters prior to the complete implementation and enforcement of the policy, or a change in the relative difference across quarters represented by the coefficients. We remind the reader that the data is monthly, and we use monthly fixed effects in combination with quarter event coefficients,

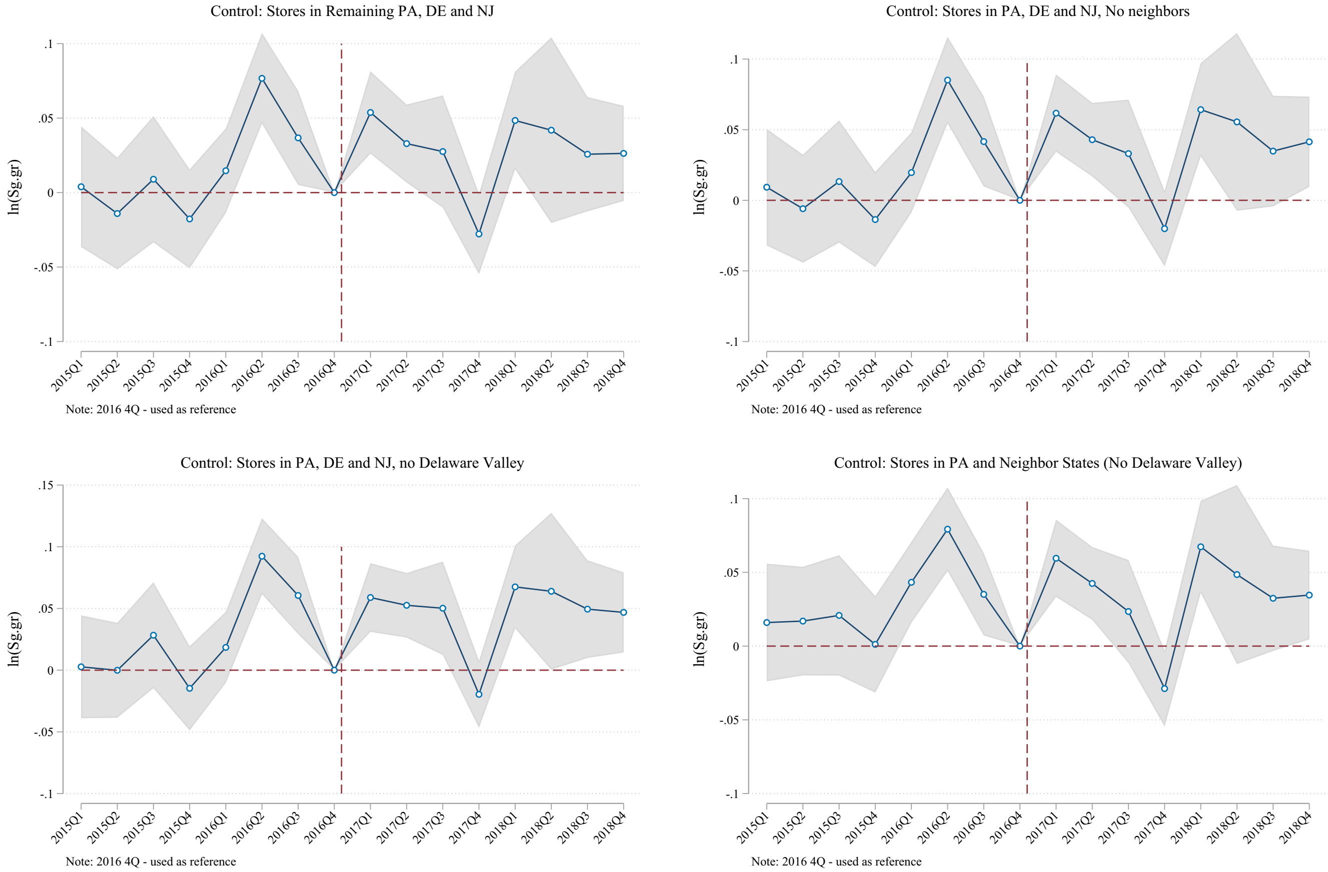

FIGURE 3 DiD Estimation - Full Event - Ln (Sugar Grams). Plot of the different coefficients, β_t from Equation (2). 95% Confidence intervals. Standard errors clustered at the store level [Colour figure can be viewed at wileyonlinelibrary.com]

allowing us to keep the richness of the data while still softening some of its noise. In separate regressions we aggregate by half year and find almost exactly identical point estimates, and no significant effect in the pre-period (see Appendix A2 where Figure A3 presents the event studies and Table A2, the coefficients). Prior to the mid 2016 reaction, most quarters do not exhibit a significant difference between treatment and control units. This remains true for the 3-digit zipcode clustering with the exception of the control that includes adjacent states, where significant differences are present across the entire pre-period. We base our selection of control on the summation of square deviations from zero in the pre-period, which again means the control that best resembles Philadelphia prior to the tax is the control comprised by stores in Pennsylvania, Delaware and New Jersey. The post-tax trends show that the effect might be different over time. The first quarter is statistically significant and positive. The effect gets smaller and even negative in the fourth quarter post-tax, but continues positive during the four quarters of 2018.

5.1.2 | Results for sugar prices from additional sweetened foods

If demand of other sugary goods increases substantially after the tax is introduced, it is possible that their prices could be affected by the tax. If this price change materializes, it will undermine the potential the soda tax has to incentivize consumers to substitute their sugar source from sodas to different sweetened foods. Panel B of Table 2 presents the results of Equation (1) where the outcome is the log of the price per sugar gram. Sugar prices from additional foods were not affected in the aggregate by the SSB tax. None of the four control groups shows a significant effect of soda taxes on the price of sugar from additional goods. This result is consistent across the clustering level and across the different robustness exercises. In term of the magnitudes of the coefficients the biggest change implies a reduction of 0.2% in prices, which interpreted through the baseline is a reduction of 0.004 cents/gram in the price of sugar from these foods. We further explore the effect on prices across the itemized sugary goods we analyze, to address the potential effects of changing of the composition of aggregate demand in each of the stores. As we will see below, there is some heterogeneity in those price changes.

Figure 4 shows the event study results for sugar gram prices. The first three exercises suggest comparable groups between Philadelphia and the pool of close localities, as we find no significant change in the differences between prices of treatment and control. For the last control, we find some statistically significant differences in the series of prices. These graphical observations is further corroborated by the measure of the summation of the squared deviation from zero, reported at the bottom of Panel A of Table 2, where the minimum difference is found on the first control groups.

5.2 | Cross-border shopping activity

We run the models described in Section 4.1 using neighboring localities as treated and excluding Philadelphia stores to assess the extent of cross-border shopping activity on additional sugary foods. The 3-digit zip codes contiguous to Philadelphia are located in both Pennsylvania and New Jersey, in the metro extension of Philadelphia and Camden, NJ. Those zipcodes are 190, 193, 194, 080 and 081. Table 3 presents the results over both, prices and quantities.

In Appendix A3, Figure A4 presents the event studies related to the estimation presented in Table 3. We observe the same trade-off we found in the previous section between geographic proximity and comparability. Localities in different states than PA, DE and NJ, present several violations to the parallel trend assumption in both prices and quantities, whereas the estimations in those states do not.

Table 3 uses the neighboring localities as treated. Thus, they are always excluded of the control rendering models 1 and 2 from the prior subsection redundant. For comparability, we label the model comparing to the remaining localities from PA, DE and NJ models 1 & 2. We find an increase in the average amount of sugar sold in stores at neighboring localities to Philadelphia between 2.1% and 3.9%, accompanied by a decrease in the prices of sugar from these items. This is consistent with a tax avoidance behavior that makes soda and sugary items substitutes within the taxed locality but compliments in the untaxed localities.

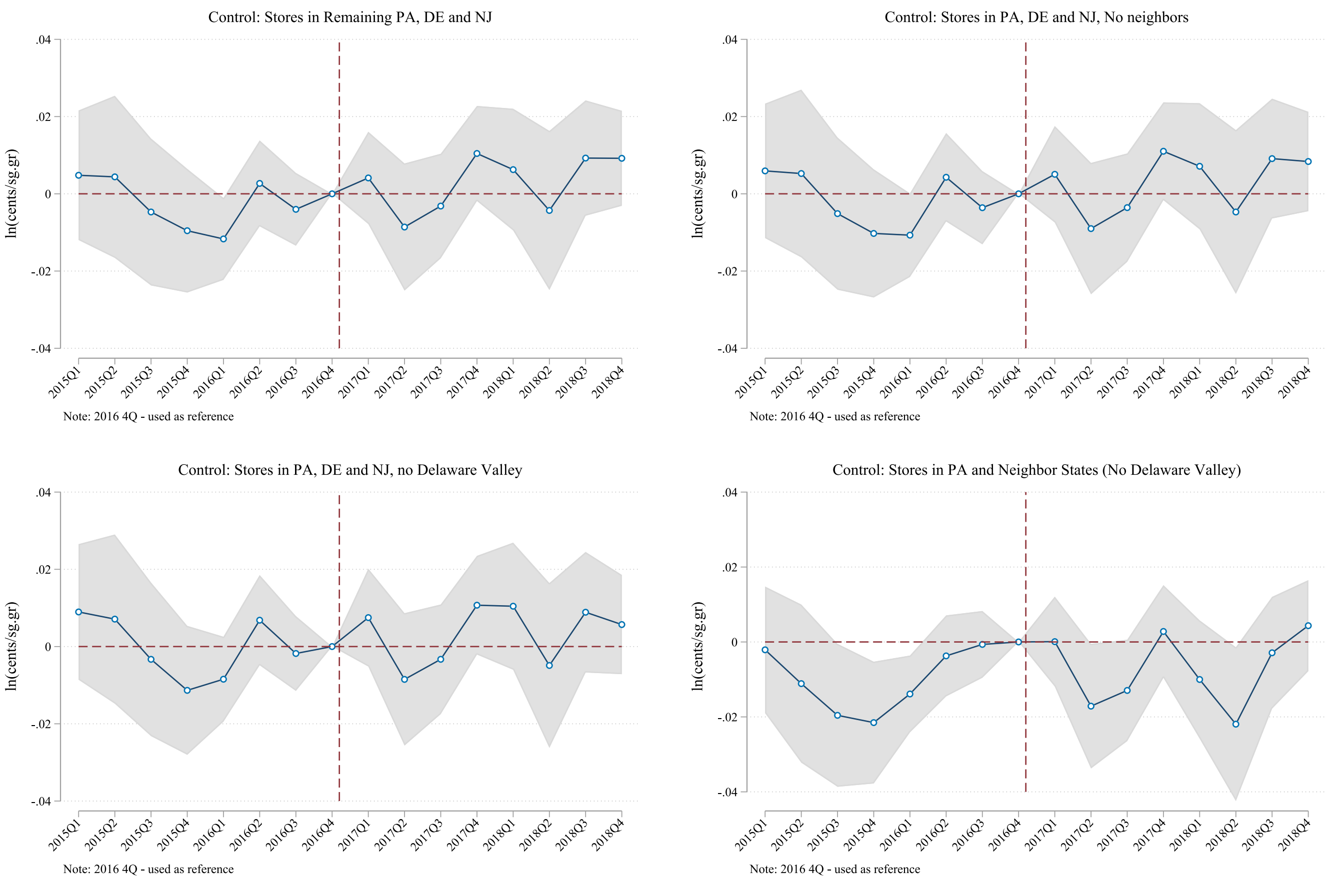

FIGURE 4 DiD Estimation - Full Event - Ln (Cents/Sugar Gram). Plot of the different coefficients, β_t from Equation (2). 95% Confidence intervals from standard errors clustered at the store level [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Effect of the Philadelphia Soda Tax in neighboring localities

PANEL A: Dependent variable Log sugar grams sold			
	(1 & 2)	(3)	(4)
	(PA,DE,NJ)	-(Metro)	+(adj.states)
Control group	ln (Sg.gr)	ln (Sg.gr)	ln (Sg.gr)
Philly neighbor · post	0.0365** (0.0090)	0.0393*** (0.0092)	0.0205*** (0.0084)
Baseline	405,668	405,668	405,668
R^2	0.977	0.979	0.978
N	137,288	118,413	374,542
N. Clusters	3087	2714	8330
Pre-Sq.Deviation	0.0149	0.0200	0.0274
PANEL B: Dependent variable log prices per gram of sugar			
	(1 & 2)	(3)	(4)
	(PA,DE,NJ)	-(Metro)	+(adj.states)
Control group	ln (cents/Sg.gr)	ln (cents/Sg.gr)	ln (cents/Sg.gr)
Philly*Post	−0.00556** (0.00194)	−0.0066** (0.00194)	−0.00535*** (0.00172)
Baseline	1.996	1.996	1.996
R^2	0.896	0.903	0.903
N	137,288	118,413	374,542
N. Clusters	3087	2714	8330
Pre-Sq.Deviation	0.00565	0.0067	0.0126

Note: Coefficients from Equation (1) including localities neighboring Philadelphia as treatment instead. Panel A presents the results over log quantities and panel B over log unitary prices. The baseline introduces the average monthly sales of sugar grams from additional sweetened items (Panel A), or the average store price per sugar gram in cents (Panel B), in the year prior to the policy implementation. Finally, the Pre-Sq.Deviation is the sum of squared coefficients from the event study prior to policy implementation (in the pre-period), such that it provides an evaluation of control quality. Standard errors clustered at the store level in parentheses.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

5.3 | Interpreting the results

How do we interpret these results? From Table 2, we know that the representative store in Philadelphia sells monthly 378.7 kilos of sugar in the 12 months prior to policy implementation. A back-of-the-envelope calculation suggests that the change of 4.3% represents an increase of about 16 kilos of sugar in the representative store.[7] How relevant are these changes in the context of the decrease in consumption and sugar purchases from sodas and other SSBs? To address this question we first run the specifications described in Equation (1) for SSBs, and assess their validity in the context of the parallel trends they exhibit using Equation (2) (see Appendix A4, where Figure A5 presents the event studies for soda purchases in Philadelphia). The coefficients from Equation (1) on sugar from SSBs in Philadelphia are presented in Panel C of Table 4 and all are statistically significant at the 0.1% confidence level. We find a decrease in purchased sugar from beverages of 31.2%, for the comparable control group to our preferred specification. This is a robust finding in the level of the coefficient across all different controls. We cannot compare directly with the previous literature since most papers report reductions in ounces rather than sugar grams (Cawley et al., 2019; Seiler et al., 2021). However, our percentage change is somewhat smaller than previous findings.

When we interpret the results we find that the reduction in sugar purchased from taxed items is 86.2 kilos in the representative store in Philadelphia. Accordingly, when comparing the changes in consumption from sugary beverages and sugary items, the substitution behavior toward higher consumption of sugary items different than soda offsets the desired objective of the tax on decreased sugar purchased by 19.1% since there is an increase of 16 kilos in sugar sales from additional sweetened items in the representative store. If we add on top of this increase the average increase in sugar sales at the representative neighboring stores, the offset is 36.6%, and the numbers for each control group are reported in the row labeled $|(A+B)/C|$.

The comparison assumes that we can extrapolate our findings, and that the changes in sugar purchases on the items we do not observe are on the order of the changes we present here. As stated in Section 3, we do not have information of brands not

TABLE 4 Back-of-the-Envelope Calculations - Net Change in Sugar purchases

PANEL A: Philadelphia Additional Sugary Items Coefficients and Baseline					
		(1)	**(2)**	**(3)**	**(4)**
		(PA,DE,NJ)	**-(Neighbors)**	**-(Metro)**	**+(adj.states)**
	Control group	**ln (Sg.gr)**	**ln (Sg.gr)**	**ln (Sg.gr)**	**ln (Sg.gr)**
(A)	Add. Items	0.0363**	0.0425***	0.0467***	0.0245***
	Baseline (grms)	378,696	378,696	378,696	378,696
	Change (grms)	13,999	16,441	18,105	9393
	(|A/C|)	16.2%	19.1%	21.0%	11.5%
PANEL B: Neighbors additional sugary items coefficients and baseline					
		(1)	**(2)**	**(3)**	**(4)**
		(PA,DE,NJ)	**-(Neighbors)**	**-(Metro)**	**+(adj.states)**
	Control group	**ln (Sg.gr)**	**ln (Sg.gr)**	**ln (Sg.gr)**	**ln (Sg.gr)**
(B)	Add. Items	0.0365**	0.0365**	0.0393***	0.0206***
	Baseline (grms)	405,668	405,668	405,668	405,668
	Change (grms)	15,080	15,080	16,260	8443
	(|B/C|)	17.5%	17.5%	18.9%	10.3%
PANEL C: Philadelphia sugar sweetened beverages - coefficients and baseline					
		(1)	**(2)**	**(3)**	**(4)**
		(PA,DE,NJ)	**-(Neighbors)**	**-(Metro)**	**+(adj.states)**
	Control group	**ln (Sg.gr)**	**ln (Sg.gr)**	**ln (Sg.gr)**	**ln (Sg.gr)**
(C)	Beverages	−0.312***	−0.312***	−0.312***	−0.294***
	Baseline (grms)	321,750	321,750	321,750	321,750
	Change (grms)	−86,235	−86,235	−86,235	−81,957
	|(A + B)/C|	33.7%	36.6%	39.9%	21.8%

Note: Coefficients from Equation (1) for separate regressions for additional sweetened foods and sweetened beverages. Panel A presents the results for additional sugar sweetened items different than beverages, Panel B presents the results for neighboring stores, and panel C presents the results for beverages. The baseline introduces the monthly sales of sugar grams for both additional items and beverages in the representative store, in the year prior to the policy implementation (A + B)/C result establishes a comparison across unitary representative stores in Philadelphia and in the neighboring localities.

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

included in the Nielsen data nor from UPCs for which we can not retrieve their nutritional content. This happens in both SSBs and additional sugary items.

5.4 | Robustness checks

As mentioned in Section 4, the consideration of competing control groups and allowing for two levels of clustering establishes our first attempt to account for the skepticism we have toward the choice of standard errors and the comparisons we make. However, we perform additional tests to further assess the robustness of our analysis. We start with bootstrapped standard errors, following Bertrand et al. (2004). Figure 5 presents the distributions of coefficients for bootstrapped estimations of Equation (1) using store clustering, while 3-digit zipcode clustering is presented in Appendix A5 Figure A6. For store level clustering we draw with replacement stores from treatment and control. For 3-digit zipcode clustering, we only draw from the control group as there is only one cluster for Philadelphia.

We present results for two of the controls we have followed this far. On the left panels of the Figure 5, we present results for stores in the remaining of Pennsylvania, Delaware and New Jersey without neighboring localities (Model two in the previous subsection), our preferred specification, and on the right panel, the same remaining stores but without the entire metro area (Model 3 previously). These groups suffer less from contagion effects, and offer better counterfactuals than the control with

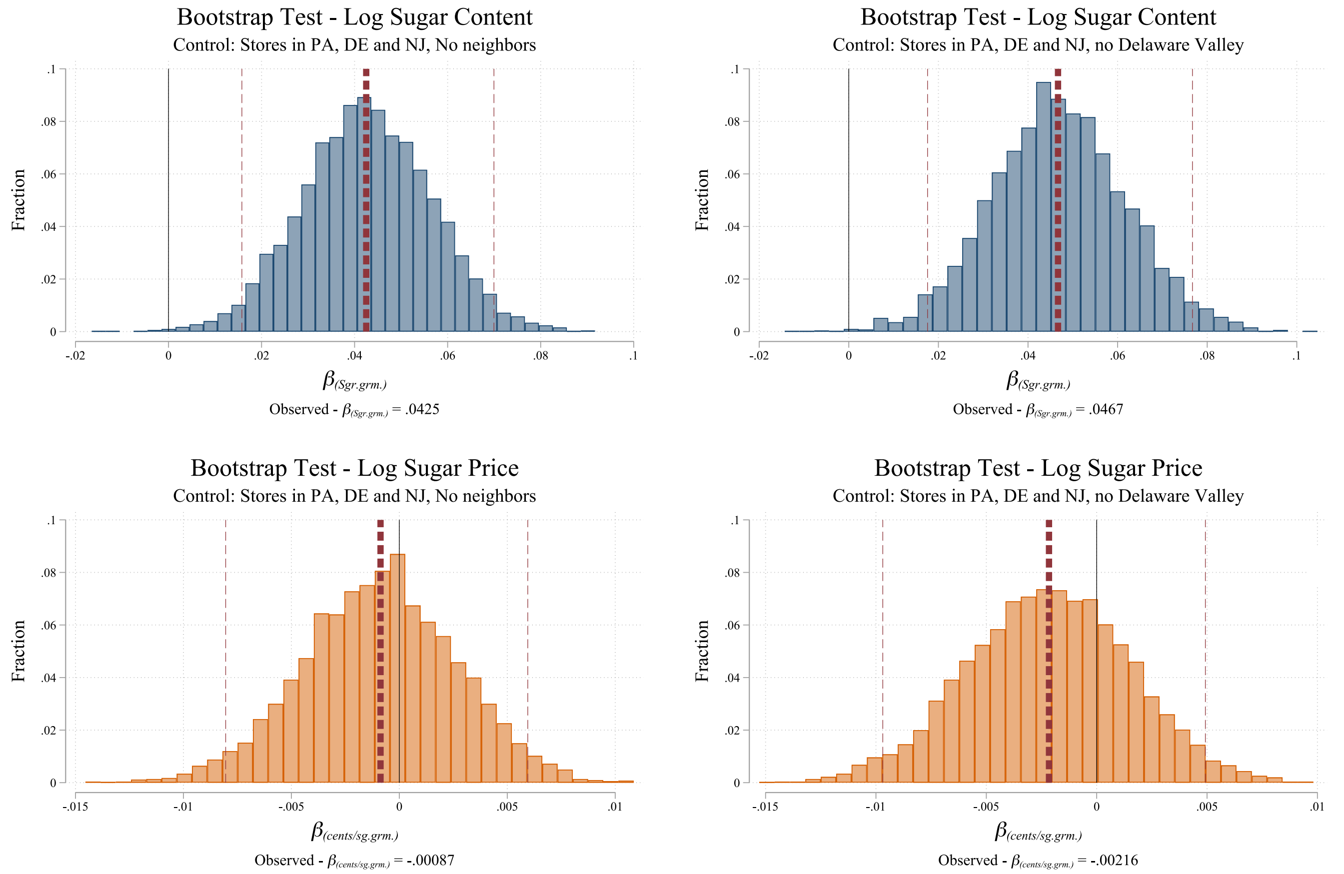

FIGURE 5 Bootstrap Distribution of Coefficients. Bootstrapped estimation of the effect of the introduction of the Philadelphia Soda Tax on prices and quantities of additional sweetened items. Bootstrap draws based on store clustering. Each unit of the histogram is the estimation of the effect using a bootstrapped sample (without replacement) for the treated and control units. For the treated group, we find 195 stores in Philadelphia. For the control sample (2) Remaining PA, DE and NJ, without neighbors there are 3085 stores to draw from. For the control sample (3) Remaining PA, DE and NJ, without Philadelphia metro area, there are 2300 stores [Colour figure can be viewed at wileyonlinelibrary.com]

adjacent states. The top two panels of Figure 5 present the histograms for the log of quantities, and the bottom panels for the log of prices. The thick red line portrays our estimated effect, as presented in Table 2. The dashed red lines are the bootstrapped 95% confidence intervals, and the black continuous line presents the zero effect over the variable in question. For quantities, we observe that the distributions are at the right of zero and both are significant as the zero line falls on the rejection zone. For prices, we observe a narrowly calculated zero, as both exercises the zero is included in the cnfidence interval range and well toward the center of the distribution. These results hold for both levels of clustering: store-level and 3-digit zipcode level (an analogous graph to Figure 5 is presented in Appendix A5). Again, the store clustering renders more conservative confidence intervals.

The second robustness exercise is based in placebo inference, trying to obtain a distribution of effects based on the control units. For the store level clustering, following the results from Table 2, we have 3280 clusters of which 195 are in Philadelphia. From the remaining stores, we draw 5000 random draws without replacement based on some of the possible permutations of more than 3000 donor stores from the control and 195 places among the real treatment units. For each random draw we estimate the "fake" treatment effect, and in Figure 6 we present the distribution for the two controls we have followed this far.

On the top panel, we present the result for the change in sugar purchased from sweetened foods. In addition to the 95% confidence interval, the figure also presents the *p*-value for the one-side tail test. Accordingly, when looking at the controls units for the rest of PA, DE and NJ with no neighbors, out of 5000 draws for placebo inference, only 0.1% of the estimated "treatment effects" were greater than the estimated treatment effect for Philadelphia as presented in Table 2, of 0.043. Once again, we confirm that the introduction of the soda tax did not lead to a change in prices of sugar from additional sweetened foods. The bottom panel of the Figure 6 presents the placebo distribution for the treatment effects over price, and the actual value we find for Philadelphia is in the middle of the distribution as close to zero as possible.

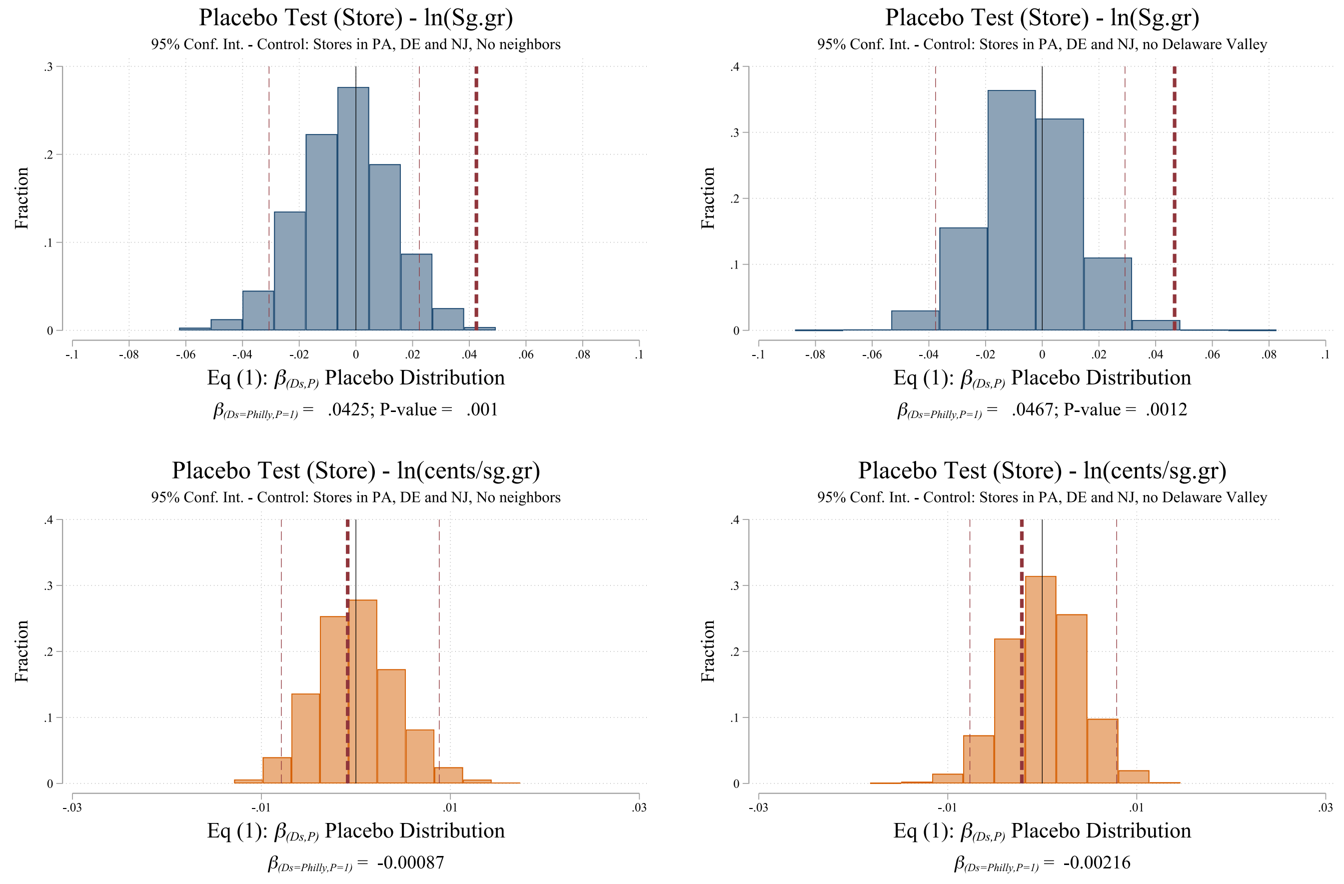

FIGURE 6 Placebo Distribution of Treatment Effects. Placebo inference distribution of the effect on prices and quantities of additional sweetened items of the introduction of a Soda Tax, as the one in Philadelphia, in 5000 different permutations of control units clustered at the store level. Each unit of the histogram is the estimation of a "fake" treatment effect over a random draw without replacement of units from the control sample used as donors, and compared to the remaining of the sample. The control sample (2) Remaining PA, DE and NJ, without neighbors has 3280 stores 195 are treated, the permutation sample draws from the permutation of the remainder 3085 donor stores on 195 positions for "fake" treatments. The control sample (3) Remaining PA, DE and NJ, without metro, uses 2300 stores from non-Philadelphia stores to accommodate in 195 treated stores [Colour figure can be viewed at wileyonlinelibrary.com]

In Figure A7 of Appendix A6, we present the histograms for taking random draws clustering at the 3-digit zipcode. We present the results for completeness, but we see some limitations to this level of clustering. The *p*-values for the two control groups we have followed are close to 25%. This means that 25% of the estimated effects for "fake" treatment are greater than the effects we find for Philadelphia. This level of clustering coarsens the distribution, and there are observations with increases well beyond 10% that are substantial. In fact, we can observe that if we remove the neighbors or other localities in the Delaware valley, the number of units in the donor sample is of only 57 or 49 units, and several placebo localities are in NJ surrounding the New York area. On the other hand, going further, for the control group that includes adjacent states but drops the Delaware Valley, we observe that the biggest increases in sugar purchased from sweetened foods comes from the New York city area and some 3-digit zipcodes near Albany. The fact that concomitant exogenous variation affects our control units casts doubt about the feasibility of inferring from this sample of clustered units. These two arguments cast doubt about the ability to conduct inference using 3-digit zipcode variation. The coarsening of the distribution demands huge treatment effects, and the potential additional exogenous variation in other cities that might drive the results on them, whether from the soda tax in itself through cross-border shopping, or from the other potential measures that affect consumption behavior. This problem of lack of power given few donors and potential exogenous variation has been pointed before for placebo inference (Buchmueller et al., 2011; Cunningham & Shah, 2018; Hagemann, 2019).

5.5 | Heterogeneity across different goods

We ran the same DD specifications as in Equation (1) for each category for total sugar and price. Since the control group that includes all stores in PA, DE, and NJ best satisfied the parallel trends assumption for prices and total sugar, it is the one we used for this analysis. Table 5 presents the results for total sugar consumed while Table 6 presents the results for prices. Results are presented in ascending order of the effects across the different items. There are some significant and some non-significant coefficients, and there is not a clear pattern of increased consumption or price changes across the different items. There is no systematic substitution of sugar purchased from sodas to a different sugary good obvious in our data. Finally, Figure 7 order the items by amount of sugar content to present the changes in sugar and unitary prices of sugar per each individual item. Of the top four (which make up 72.6% of the total sugar consumed), two coefficients are non-significant, one is significantly positive, and one is significantly negative. Therefore, there is not an obvious main sugary food substitute for SSB.

6 | DISCUSSION AND CONCLUSIONS

Using data from POS from Nielsen Scanner Data, supplemented with Nutrition Facts from the USDA Food Central and additional online sources, we found that the tax on SSBs in Philadelphia did not raise the prices of sugary foods but slightly increased the purchase of some additional sweetened items different than beverages. Our preferred specification indicated that there was a 4.3% increase in sales of sugar sweetened foods in the representative Philadelphia store. We also find that the amount of sugar purchased from additional sugary items increased 3.7% in stores neighboring Philadelphia. We test these results across different control groups and across different ways of clustering, concerned that we cannot trust the standard errors due to a small number of treatment sample units. We compare with different geographies based on the proximity to Philadelphia and to the degree of contagion that these control units could have experienced due to the introduction of the policy. Accordingly, we analyze clustering at the geographic unit at which the policy was implemented (3-digit zipcode that represents the geographic area of Philadelphia). However, we present store level standard errors as our preferred clustering as they are more conservative. In addition, we produce bootstrapped standard errors and placebo tests using both levels of clustering. They largely support our results.

We compare the results in a back-of-the-envelope setting to the changes in sugar purchased from beverages after the introduction of the tax. We find that the increase in sugar purchased from additional sweetened foods inside Philadelphia offsets approximately 19.1% the reduction in sugar purchased from beverages. Furthermore, purchases from additional sugary items in neighboring localities offset up to 18.9% of the change in sugar from SSBs. Combined the reduction offsets between 21.8% and 39.9% of the decrease in sugar purchases from SSBs. We also analyze the changes in consumption at individual categories of sugary foods. We find a substantial heterogeneity in changes of consumption and unitary sugar prices in these items. However, out of the four main sources of sugar intake, only desserts consumption is increasing. This result is consequential with the recommendation of taxing SSBs, instead of a broader range of sugary items, as SSBs only have the nutritional value of providing energy with less satiation compared to other foods with the same caloric content (Roache & Gostin, 2017).

Our conclusions are not free of caveats, and all our results are local to the information we observe. We make inference on a subsample of the Nielsen data for which we can retrieve nutritional information. We lose information on retailer brands that, by design, we cannot query as it is anonymized. Also, the Nielsen data over-represents bigger retailers and, as in those retailers it is easier to find items of popular, well-known brands, these products are over represented in our sample. While there should

TABLE 5 Effect on sugar purchased from additional sugary goods, in each category

	(Fruit)	(Baked)	(Sugars)	(Cookies)	(Candy)	(Dressings)	(Breakfast)	(Desserts)	(Jams)
	ln (Sg.gr)	ln (Sg.gr)	ln (Sg.gr)	ln (Sg.gr)	ln (Sg.gr)	ln (Sg.gr)	ln (Sg.gr)	ln (Sg.gr)	ln (Sg.gr)
Philly*Post	−0.148***	−0.0801**	−0.0665*	−0.0469	0.0133	0.0417	0.0632*	0.0664*	0.110***
	(0.0437)	(0.0263)	(0.0268)	(0.0839)	(0.0126)	(0.0267)	(0.0286)	(0.0270)	(0.0321)
Baseline	93.5	243.1	314.5	102.8	425.4	68.0	114.6	77.0	73.4
R^2	0.808	0.950	0.944	0.212	0.942	0.965	0.946	0.938	0.916
N	215,121	116,610	155,543	169,854	118,593	110,548	118,089	115,995	107,629
N Clusters	2711	2727	2708	2740	2743	2691	2734	2706	2659

Note: Control group including stores in PA, DE & NJ. Store level clustered standard errors in parentheses.

+ $p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

TABLE 6 Effect on unitary sugar gram prices of each sugary good category

	(Desserts)	(Sugars)	(Breakfast)	(Baked)	(Jams)	(Dressings)	(Cookies)	(Candy)	(Fruit)
	ln (ct/sg.gr)	**ln (ct/sg.gr)**	**ln (ct/sg.gr)**	**ln (ct/sg.gr)**	**ln (ct/sg.gr)**	**ln (ct/sg.gr)**	**ln (ct/sg.gr)**	**ln (ct/sg.gr)**	**ln (ct/sg.gr)**
Philly*Post	−0.0473**	−0.0370	−0.0252	−0.0241*	−0.00852	−0.00619	0.0255	0.0303***	0.0565***
	(0.0163)	(0.0260)	(0.0241)	(0.0113)	(0.0170)	(0.0102)	(0.0268)	(0.00492)	(0.0115)
Baseline	0.012	0.014	0.073	0.003	0.085	0.208	0.026	0.002	0.029
R^2	0.495	0.274	0.451	0.779	0.470	0.759	0.247	0.715	0.278
N	115,995	155,543	118,089	116,610	107,629	110,548	169,854	118,593	215,121
N. Clusters	2706	2708	2734	2727	2659	2691	2740	2743	2711

Note: Control group including stores in PA, DE & NJ. Store level clustered standard errors in parentheses.

+ $p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

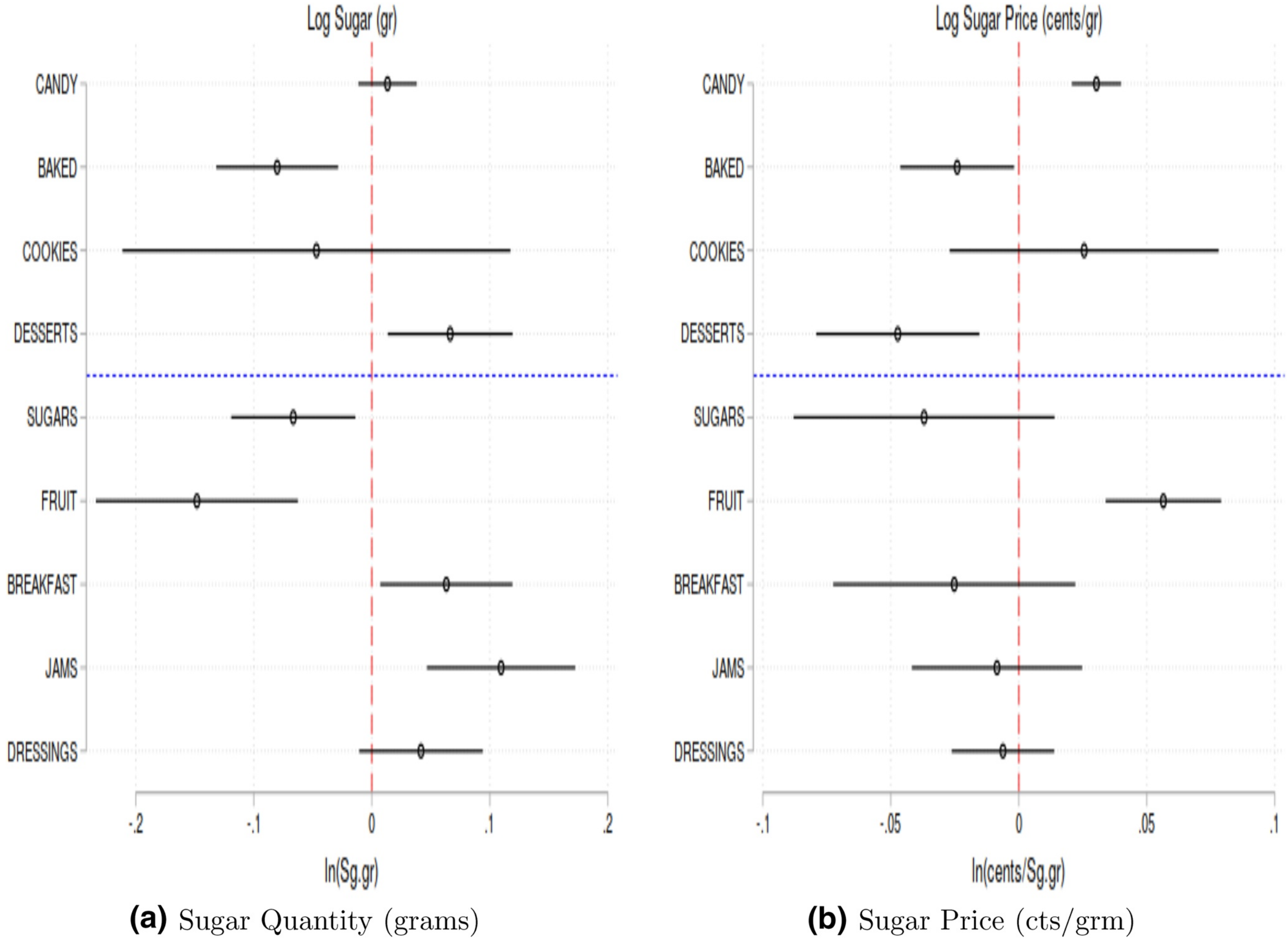

FIGURE 7 Change in sugar consumption and prices by food group [Colour figure can be viewed at wileyonlinelibrary.com]

be differences that we cannot test in the substitution patterns on the items we do not observe, there are reasons for us to believe that they might not entirely offset our results. First, we do not expect the difference in sugar for additional sugary items to be dramatically different, as within good items several competing brands have similar levels of sugar. If people substituted to the unobserved brands that have higher sugar density in relation to our observed items, then our estimates would be conservative. If the opposite is true that would bias our estimates down.

While the data are not entirely perfect, it is one of the most representative information sources that has been used to evaluate soda taxes. The Nielsen Data captures "more than half the total sales volume of US grocery and drug stores and more than 30% of all US mass merchandizer sales volume".[8] This information from administrative records has a high degree accuracy and concentrates a very relevant share of the total market. According to our results, the data provide evidence that there is an important offset to the 2017 SSB tax in Philadelphia through substitution to sugary food.

CONFLICT OF INTEREST

The author declares that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from The Nielsen Company (US), LLC, provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. Restrictions apply to the availability of these data, which were used under license for this study. Data are available at https://www.chicagobooth.edu/research/kilts/datasets/nielseniq-nielsen%20with%20the%20permission%20of%20The%20Nielsen%20Company%20(US),%20LLC,%20and%20the%20Kilts%20Center%20for%20Marketing%20Data%20Center.

ORCID

Felipe Lozano-Rojas https://orcid.org/0000-0002-9300-6094
Patrick Carlin https://orcid.org/0000-0002-4284-1946

ENDNOTES

1 Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

2 According to the crosswalk provided by HUD found here, the zip codes within Philadelphia that were excluded are PO box zip codes only. So, we are not dropping any addresses within Philadelphia.

3 Retail Scanner Data - NIELSENIQ Description. Available at: https://www.chicagobooth.edu/

4 See: Added Sugar is not so Sweet - Infographic. Available at: https://www.heart.org

5 For instance a can of Pepsi, one of Coca-Cola and one of Big Cola have 41, 39, and 38 sugar grams per can 12 lq oz respectively. Something similar happens with cookies, wafers and other items.

6 Calculated as the exponential of the average log, given that the variables are highly skewed to the left, from September 2015 to September 2016.

7 We interpret coefficient estimates as growth rates. Change is obtained from: $\Delta Y = Y_1 - Y_0$, where Y_0 is the baseline and $Y_1 = Y_0 \cdot e^{\beta}$, using the coefficient from Equation (1).

8 Retail Scanner Data - NIELSENIQ Description. Available at: https://www.chicagobooth.edu/

REFERENCES

Allcott, H., Lockwood, B. B., & Taubinsky, D. (2019a). Regressive sin taxes, with an application to the optimal soda tax. *The Quarterly Journal of Economics*, *134*(3), 1557–1626. https://doi.org/10.1093/qje/qjz017

Allcott, H., Lockwood, B. B., & Taubinsky, D. (2019b). Should we tax sugar-sweetened beverages? An overview of theory and evidence. *Journal of Economic Perspectives*, *33*(3), 202–227. https://doi.org/10.1257/jep.33.3.202

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, *119*(1), 249–275. https://doi.org/10.1162/003355304772839588

Bleich, S. N., Dunn, C. G., Soto, M. J., Yan, J., Gibson, L. A., Lawman, H. G., Mitra, N., Lowery, C. M., Peterhans, A., Hua, S. V., & Roberto, C. A. (2021). Association of a sweetened beverage tax with purchases of beverages and high-sugar foods at independent stores in Philadelphia. *JAMA Network Open*, *4*(6), e2113527. https://doi.org/10.1001/jamanetworkopen.2021.13527

Buchmueller, T. C., DiNardo, J., & Valletta, R. G. (2011). The effect of an employer health insurance mandate on health insurance coverage and the demand for labor: Evidence from Hawaii. *American Economic Journal: Economic Policy*, *3*(4), 25–51. https://doi.org/10.1257/pol.3.4.25

Cawley, J., Frisvold, D., Hill, A., & Jones, D. (2019). The impact of the Philadelphia beverage tax on purchases and consumption by adults and children. *Journal of Health Economics*, *67*, 102225. https://doi.org/10.1016/j.jhealeco.2019.102225

Cawley, J., Frisvold, D., Hill, A., & Jones, D. (2020a). The impact of the Philadelphia beverage tax on prices and product availability. *Journal of Policy Analysis and Management*, *39*(3), 605–628. https://doi.org/10.1002/pam.22201

Cawley, J., Frisvold, D., Hill, A., & Jones, D. (2020b). Oakland's sugar-sweetened beverage tax: Impacts on prices, purchases and consumption by adults and children. *Economics & Human Biology*, *37*, 100865. https://doi.org/10.1016/j.ehb.2020.100865

Cawley, J., Frisvold, D., & Jones, D. (2020c). The impact of sugar-sweetened beverage taxes on purchases: Evidence from four city-level taxes in the United States. *Health Economics*, *29*(10), 1289–1306. https://doi.org/10.1002/hec.4141

Cawley, J., Frisvold, D., Jones, D., & Lensing, C. (2021). The pass-through of a tax on sugar-sweetened beverages in Boulder, Colorado. *American Journal of Agricultural Economics*, *103*(3), 987–1005. https://doi.org/10.1111/ajae.12191

Cawley, J., & Frisvold, D. E. (2017). The pass-through of taxes on sugar-sweetened beverages to retail prices: The case of Berkeley, California. *Journal of Policy Analysis and Management*, *36*(2), 303–326. https://doi.org/10.1002/pam.21960

Cawley, J., Thow, A. M., Wen, K., & Frisvold, D. (2019). The economics of taxes on sugar-sweetened beverages: A review of the effects on prices, sales, cross-border shopping, and consumption. *Annual Review of Nutrition*, *39*(1), 317–338. https://doi.org/10.1146/annurev-nutr-082018-124603

Cawley, J., Willage, B., & Frisvold, D. (2018). Pass-through of a tax on sugar-sweetened beverages at the Philadelphia international Airport. *Journal of the American Medical Association - JAMA*, *319*(3), 305–306. https://doi.org/10.1001/jama.2017.16903

Chriqui, J. F., Sansone, C. N., & Powell, L. M. (2020). The sweetened beverage tax in Cook County, Illinois: Lessons from a failed effort. *American Journal of Public Health*, *110*(7), 1009–1016. https://doi.org/10.2105/ajph.2020.305640

Colchero, A., Popkin, B., Rivera, J., & Ng, S. (2016). Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: Observational study. *BMJ*, *352*, 6704. https://doi.org/10.1136/bmj.h6704

Conley, T. G., & Taber, C. R. (2011). Inference with "difference in differences" with a small number of policy changes. *The Review of Economics and Statistics*, *93*(1), 113–125. https://doi.org/10.1162/rest_a_00049

Cunningham, S., & Shah, M. (2018). Decriminalizing indoor prostitution: Implications for sexual violence and public health. *The Review of Economic Studies*, *85*(3), 1683–1715. https://doi.org/10.1093/restud/rdx065

Einav, L., Leibtag, E., & Nevo, A. (2010). Recording discrepancies in nielsen homescan data: Are they present and do they matter? *Quantitative Marketing and Economics*, *8*(2), 207–239. https://doi.org/10.1007/s11129-009-9073-0

Falbe, J., Rojas, N., Grummon, A., & Madsen, K. (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. https://doi.org/10.2105/ajph.2015.302881

Falbe, J., Thompson, H., Becker, C., Rojas, N., McCulloch, C., & Madsen, K. (2016). Impact of the Berkeley excise tax on sugar-sweetened beverage consumption. *American Journal of Public Health*, *106*(10), 1865–1871. https://doi.org/10.2105/ajph.2016.303362

Gibson, L. A., Lawman, H. G., Bleich, S. N., Yan, J., Mitra, N., LeVasseur, M. T., Lowery, C. M., & Roberto, C. A. (2021). No evidence of food or alcohol substitution in response to a sweetened beverage tax. *American Journal of Preventive Medicine*, *60*(2), e49–e57. https://doi.org/10.1016/j.amepre.2020.08.021

Grummon, A. H., Lockwood, B. B., Taubinsky, D., & Allcott, H. (2019). Designing better sugary drink taxes. *Science*, *365*(6457), 989–990. https://doi.org/10.1126/science.aav5199

Haavio, M., & Kotakorpi, K. (2011). The political economy of sin taxes. *European Economic Review*, *55*(4), 575–594. https://doi.org/10.1016/j.euroecorev.2010.06.002

Hagemann, A. (2019). Placebo inference on treatment effects when the number of clusters is small. *Journal of Econometrics*, *213*(1), 190–209. https://doi.org/10.1016/j.jeconom.2019.04.011

Han, E., & Powell, L. (2013). Consumption patterns of sugar-sweetened beverages in the United States. *Journal of the Academy of Nutrition and Dietetics*, *113*(1), 43–53. https://doi.org/10.1016/j.jand.2012.09.016

Jones-Smith, J. C., Walkinshaw, L. P., Oddo, V. M., Knox, M., Neuhouser, M. L., Hurvitz, P. M., Saelens, B. E., & Chan, N. (2020). Impact of a sweetened beverage tax on beverage prices in Seattle, WA. *Economics & Human Biology*, *39*, 100917. https://doi.org/10.1016/j.ehb.2020.100917

Lawman, H. G., Bleich, S. N., Yan, J., LeVasseur, M. T., Mitra, N., & Roberto, C. A. (2019). Unemployment claims in Philadelphia one year after implementation of the sweetened beverage tax. *PloS One*, *14*(3), e0213218. https://doi.org/10.1371/journal.pone.0213218

Lee, M. M., Falbe, J., Schillinger, D., Basu, S., McCulloch, C. E., & Madsen, K. A. (2019). Sugar-sweetened beverage consumption 3 years after the Berkeley, California, sugar-sweetened beverage tax. *American Journal of Public Health*, *109*(4), 637–639. https://doi.org/10.2105/ajph.2019.304971

Lozano-Rojas, F. (2018). An inquiry on the heterogeneous effects of sin taxes levied on soda. Available at SSRN 3225901.

Marlow, M. L. (2019). The dismal record of soda taxes. *Journal of American Physicians and Surgeons*, *24*(2), 53–56.

O'Donoghue, T., & Rabin, M. (2003). Studying optimal paternalism, illustrated by a model of sin taxes. *American Economic Review*, *93*(2), 186–191. https://doi.org/10.1257/000282803321947029

O'Donoghue, T., & Rabin, M. (2006). Optimal sin taxes. *Journal of Public Economics*, *90*(10–11), 1825–1849. https://doi.org/10.1016/j.jpubeco.2006.03.001

Ogden, C., Carroll, M., Kit, B., & Flegal, K. (2014). Prevalence of childhood and adult obesity in the United States, 2011-2012. *Journal of the American Medical Association*, *311*(8), 806–814. https://doi.org/10.1001/jama.2014.732

Powell, L. M., & Leider, J. (2020). The impact of Seattle's sweetened beverage tax on beverage prices and volume sold. *Economics & Human Biology*, *37*, 100856. https://doi.org/10.1016/j.ehb.2020.100856

Restrepo, B. J., & Cantor, J. H. (2020). The effects of soda taxes on adolescent sugar intake and blood sugar. *Health Economics*, *29*(11), 1422–1434. https://doi.org/10.1002/hec.4142

Roache, S. A., & Gostin, L. O. (2017). The untapped power of soda taxes: Incentivizing consumers, generating revenue, and altering corporate behavior. *International Journal of Health Policy and Management*, *6*(9), 489–493. https://doi.org/10.15171/ijhpm.2017.69

Roberto, C. A., Lawman, H. G., LeVasseur, M. T., Mitra, N., Peterhans, A., Herring, B., & Bleich, S. N. (2019). Association of a beverage tax on sugar-sweetened and artificially sweetened beverages with changes in beverage prices and sales at chain retailers in a large urban setting. *Journal of the American Medical Association - JAMA*, *321*(18), 1799–1810. https://doi.org/10.1001/jama.2019.4249

Seiler, S., Tuchman, A., & Yao, S. (2021). The impact of soda taxes: Pass-through, tax avoidance, and nutritional effects. *Journal of Marketing Research*, *58*(1), 22–49. https://doi.org/10.1177/0022243720969401

Silver, L., Ng, S. W., Ryan-Ibarra, S., Taillie, L. S., Induni, M., Miles, D., Poti, J., & Popkin, B. (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. https://doi.org/10.1371/journal.pmed.1002283

Taylor, R. L., Kaplan, S., Villas-Boas, S. B., & Jung, K. (2019). Soda wars: The effect of a soda tax election on university beverage sales. *Economic Inquiry*, *57*(3), 1480–1496. https://doi.org/10.1111/ecin.12776

U.S. Department of Agriculture, A. R. S. (2018). What we eat in America: Percentages of selected nutrients contributed by food and beverages consumed away from home, NHANES 2017–2018.

Williams, J., & Goldsworthy, P. (2012). *Breaking down the chain: A guide to the soft drink industry*. National Policy & Legal Analysis Network - NPLAN.

Zhen, C., Chen, Y., Lin, B. H., Karns, S., Mancino, L. M., & Ver Ploeg, M. (2022). Do obese and non-obese consumers respond differently to price changes? Implications of preference heterogeneity for using food taxes and subsidies to reduce obesity. MPRA Working Paper 112697, University Library of Munich, Germany.

Zhen, C., Finkelstein, E. A., Nonnemaker, J. M., Karns, S. A., & Todd, J. E. (2014). Predicting the effects of sugar-sweetened beverage taxes on food and beverage demand in a large demand system. *American Journal of Agricultural Economics*, *96*(1), 1–25. https://doi.org/10.1093/ajae/aat049

Zhong, Y., Auchincloss, A. H., Lee, B. K., & Kanter, G. P. (2018). The short-term impacts of the Philadelphia beverage tax on beverage consumption. *American Journal of Preventive Medicine*, *55*(1), 26–34. https://doi.org/10.1016/j.amepre.2018.02.017

How to cite this article: Lozano-Rojas, F., & Carlin, P. (2022). The effect of soda taxes beyond beverages in Philadelphia. *Health Economics*, *31*(11), 2381–2410. https://doi.org/10.1002/hec.4586

APPENDIX

A1 Main Results Clustering at the 3-Digit Zipcode Level

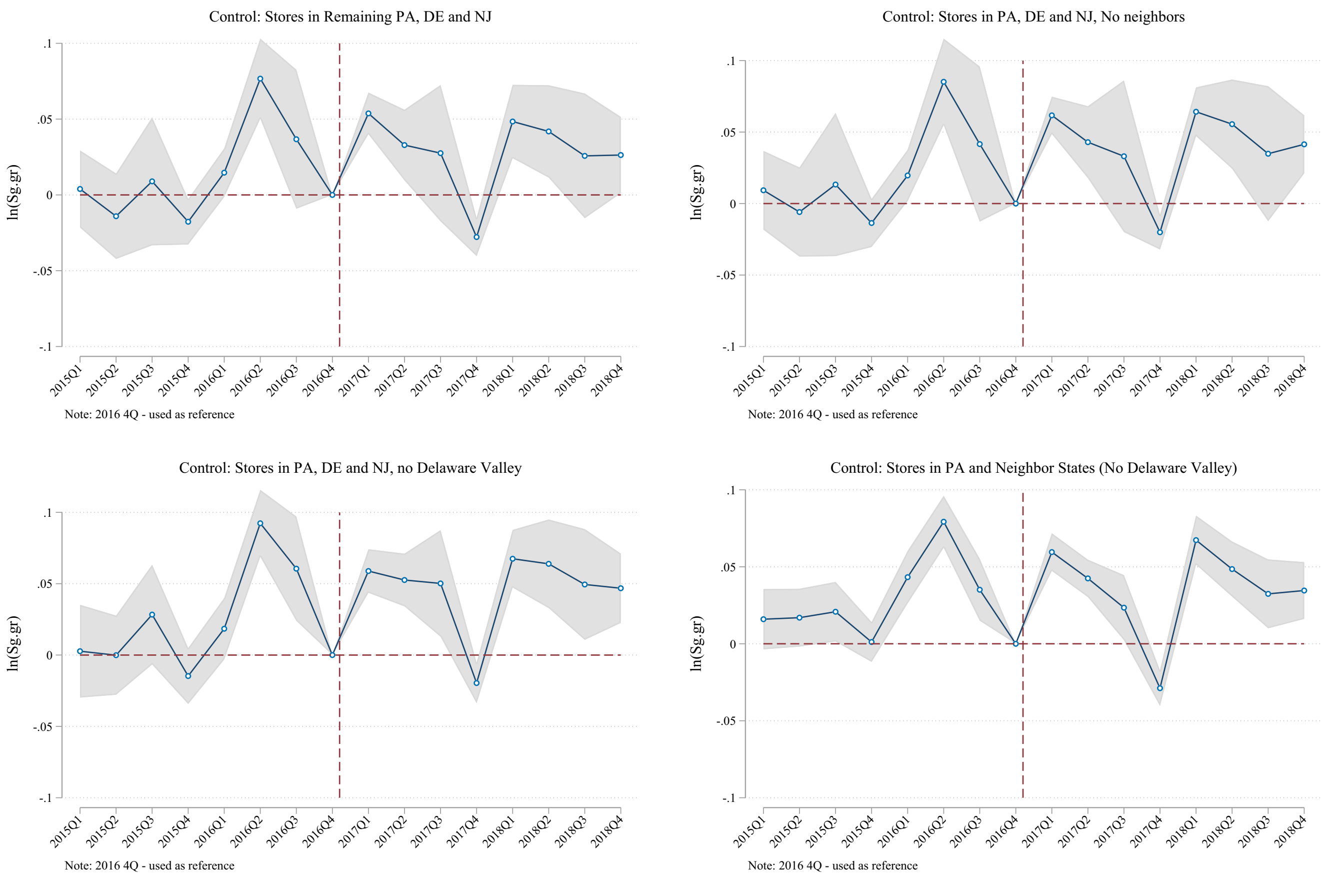

FIGURE A1 DiD Estimation - Event Studies - Ln (Sugar Grams). Plot of the different coefficients, β_t from Equation (2). 95% Confidence intervals. Standard errors clustered at the 3-digit zipcode level [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE A1 Effect of the Soda tax - 3-Digit Zipcode Clustering

PANEL A: Dependent variable log sugar grams sold				
	(1)	(2)	(3)	(4)
	(PA,DE,NJ)	-(Neighbors)	-(Metro)	+(adj.states)
Control group	ln (Sg.gr)	ln (Sg.gr)	ln (Sg.gr)	ln (Sg.gr)
Philly·Post	0.0363*** (0.0085)	0.0425*** (0.0093)	0.0467*** (0.0112)	0.0245*** (0.0085)
Baseline	378,696	378,696	378,696	378,696
R^2	0.977	0.976	0.977	0.978
N	138,574	118,874	101,161	342,230
N. Clusters	62	58	50	155
Pre-Sq.Deviation	0.0335	0.0373	0.0440	0.0375
PANEL B: Dependent variable log prices per gram of sugar				
	(1)	(2)	(3)	(4)
	(PA,DE,NJ)	-(Neighbors)	-(Metro)	+(adj.states)
Control group	ln (cents/Sg.gr)	ln (cents/Sg.gr)	ln (cents/Sg.gr)	ln (cents/Sg.gr)
Philly·Post	0.0000146 (0.00148)	−0.000868 (0.00169)	−0.00216 (0.00184)	−0.00181 (0.00125)
Baseline	1.985	1.985	1.985	1.985
R^2	0.896	0.897	0.905	0.904
N	138,574	118,874	101,161	342,230
N. Clusters	62	58	50	155
Pre-Sq.Deviation	0.0063	0.0063	0.0061	0.0123

Note: Coefficients from Equation (1). Panel A presents the results over log quantities and panel B over log unitary prices. The baseline introduces the average monthly sales of sugar grams from additional sweetened items (Panel A), or the average store price per sugar gram in cents (Panel B), in the year prior to the policy implementation. Finally, the Pre-Sq.Deviation is the sum of squared coefficients from the event study prior to policy implementation (in the pre-period), such that it provides an evaluation of control quality. Standard errors clustered at the 3-digit zipcode level in parentheses.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

A2 Half year aggregation estimates

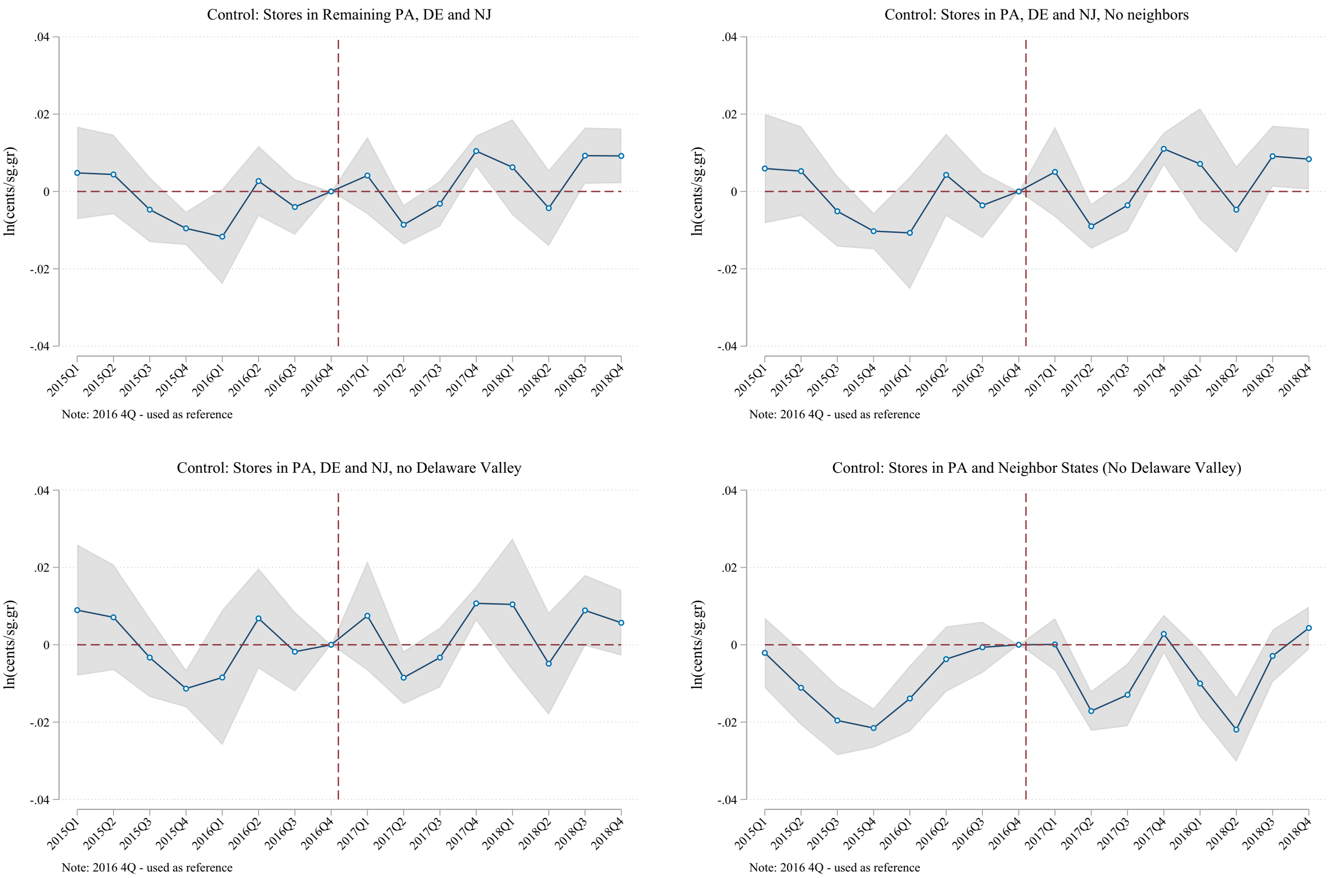

FIGURE A2 DiD Estimation - Event Studies - Ln (Cents/Sugar Gram). Plot of the different coefficients, β_t from Equation (2). 95% Confidence intervals from standard errors clustered at the 3-digit zipcode level [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE A2 Effect of the Soda tax - Half-Year Regressions

PANEL A: Dependent variable log sugar grams sold				
	(1)	**(2)**	**(3)**	**(4)**
	(PA,DE,NJ)	**-(Neighbors)**	**-(Metro)**	**+(adj.states)**
Control group	**ln (Sg.gr)**	**ln (Sg.gr)**	**ln (Sg.gr)**	**ln (Sg.gr)**
Philly·Post	0.0355** (0.0176)	0.0412** (0.0177)	0.0438** (0.0179)	0.0291* 0.0182)
R^2	0.981	0.980	0.978	0.971
N	24,732	21,220	18,055	61,566
N. Clusters	3247	2843	2470	8034
Pre-Sq.Deviation	0.0343	0.0335	0.0382	0.0305
PANEL B: Dependent variable log prices per gram of sugar				
	(1)	**(2)**	**(3)**	**(4)**
	(PA,DE,NJ)	**-(Neighbors)**	**-(Metro)**	**+(adj.states)**
Control group	**ln (cents/Sg.gr)**	**ln (cents/Sg.gr)**	**ln (cents/Sg.gr)**	**ln (cents/Sg.gr)**
Philly·Post	−0.00260 (0.00356)	−0.00279 (0.00360)	−0.00267 (0.00360)	0.00236 (0.00348)
R^2	0.960	0.961	0.965	0.963
N	24,732	21,220	18,055	61,566
N. Clusters	3247	2843	2470	8034
Pre-Sq.Deviation	0.0133	0.0141	0.0147	0.0090

Note: Coefficients from Equation (1). Panel A presents the results over log quantities and panel B over log unitary prices. Data aggregated at the half year level, and half year fixed effects included. The Pre-Sq.Deviation is the sum of squared coefficients from the event study prior to policy implementation (in the pre-period), such that it provides an evaluation of control quality. Standard errors clustered at the store level in parentheses.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

A3 Neighboring localities Event Studies

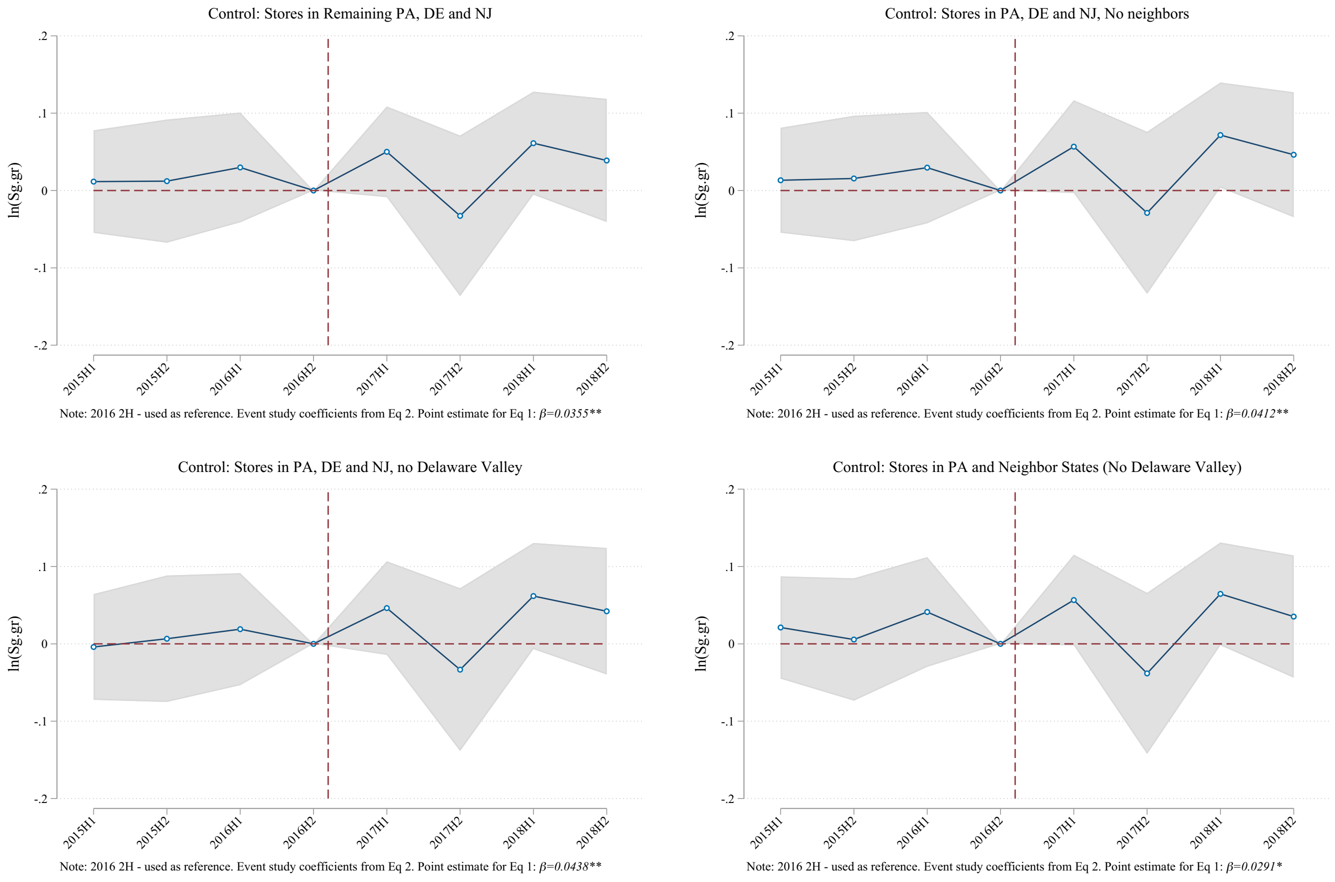

FIGURE A3 DiD Estimation Half Year Aggregation - Event Studies - Ln (Sugar Grams). Plot of the different coefficients, β_t from Equation (2). 95% Confidence intervals. Data aggregated at the half year level, and half year fixed effects included. Standard errors clustered at the store level. Each panel presents the point estimates from the coefficient in Equation (1) in its respective note below [Colour figure can be viewed at wileyonlinelibrary.com]

A4 Event Studies Sugar Sweetened Beverage measures in the Nielsen Data

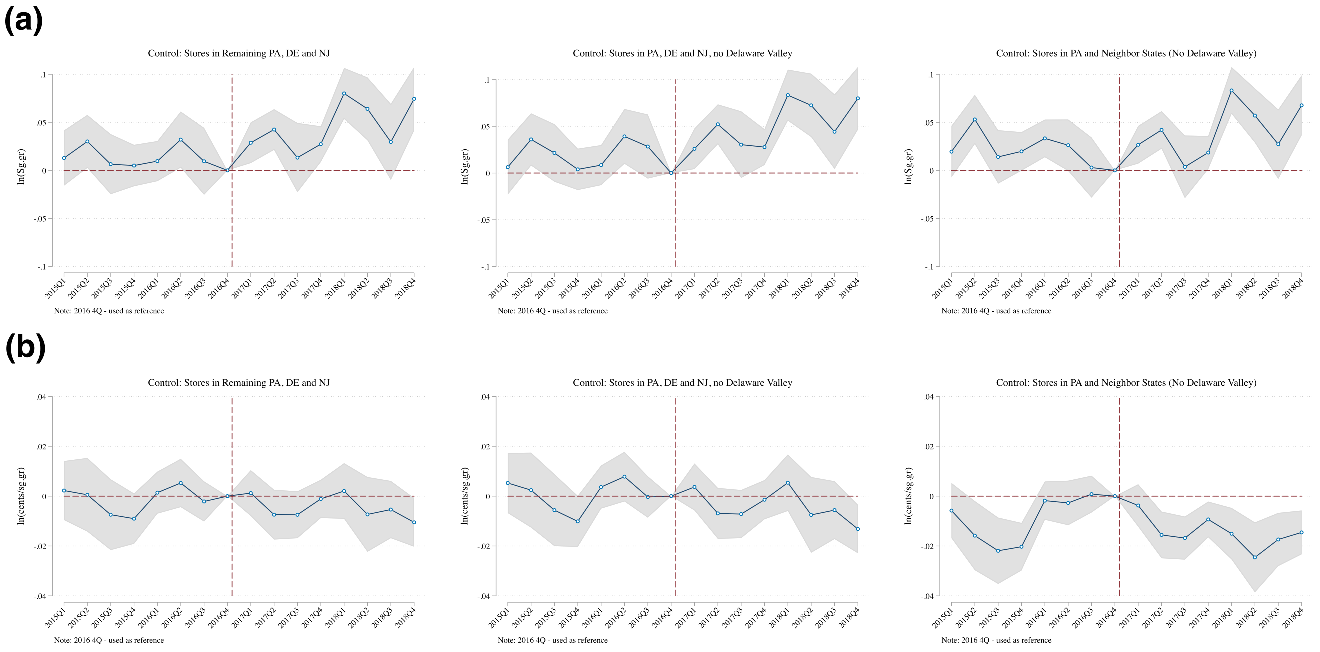

FIGURE A4 DiD Estimation - Event Studies - Neighboring Localities. Panel (a) Quantities Ln (Sg.grs.). Panel (b) Prices Ln (cts/Sg.gr.). Plot of the different coefficients, β_t from Equation (2) using localities neighboring Philadelphia as treatment. 95% Confidence intervals. Standard errors clustered at the store level [Colour figure can be viewed at wileyonlinelibrary.com]

A5 Bootstrapped Results for 3-Digit Zipcode Clustering

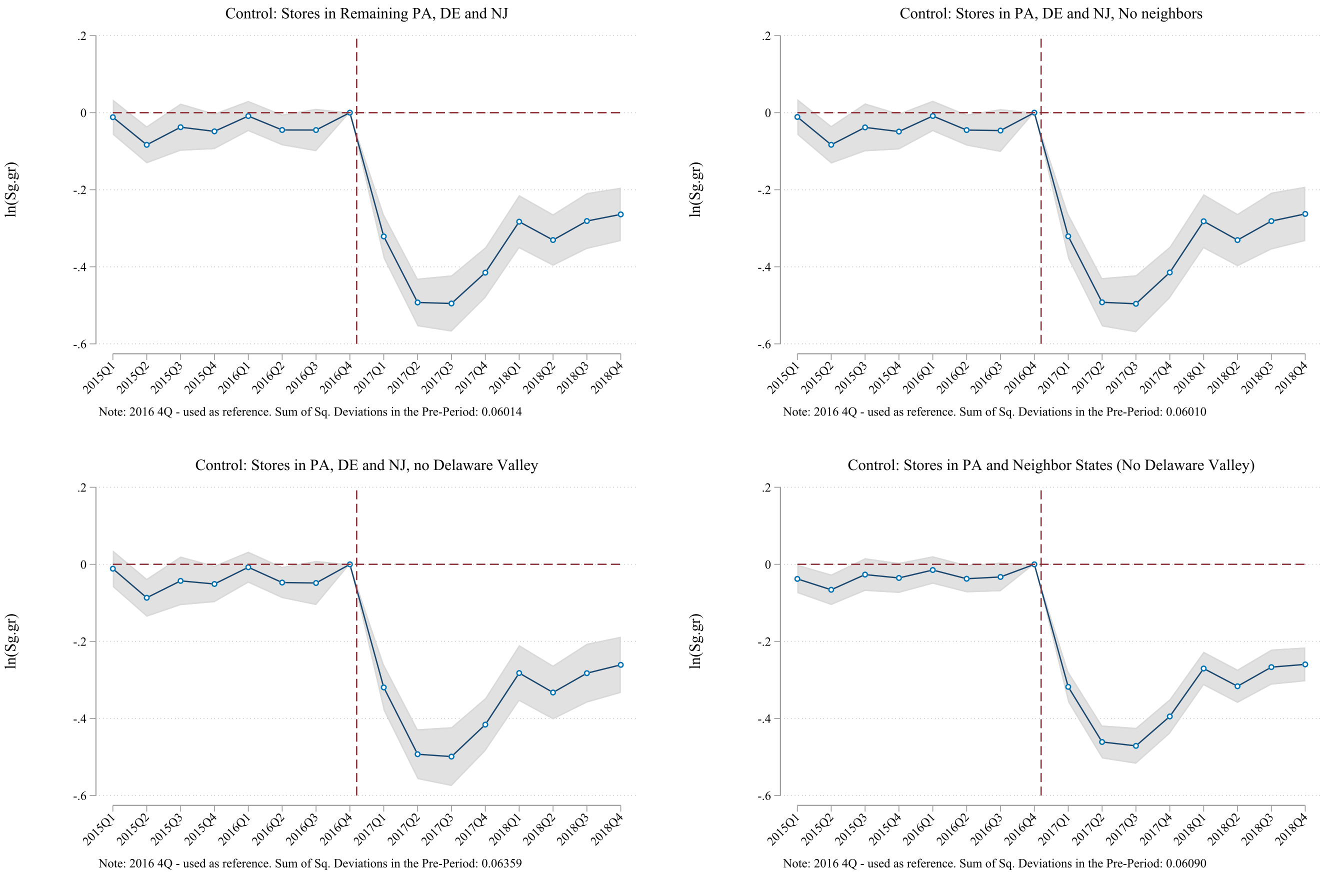

FIGURE A5 DiD Estimation - Event Studies - sugar sweetened beverage (SSB) Ln (Sugar Grams). Plot of the different coefficients, β_t from Equation (2). 95% Confidence intervals. Standard errors clustered at the store level [Colour figure can be viewed at wileyonlinelibrary.com]

A6 Placebo Inference Distribution for 3-Digit Zipcode Clustering

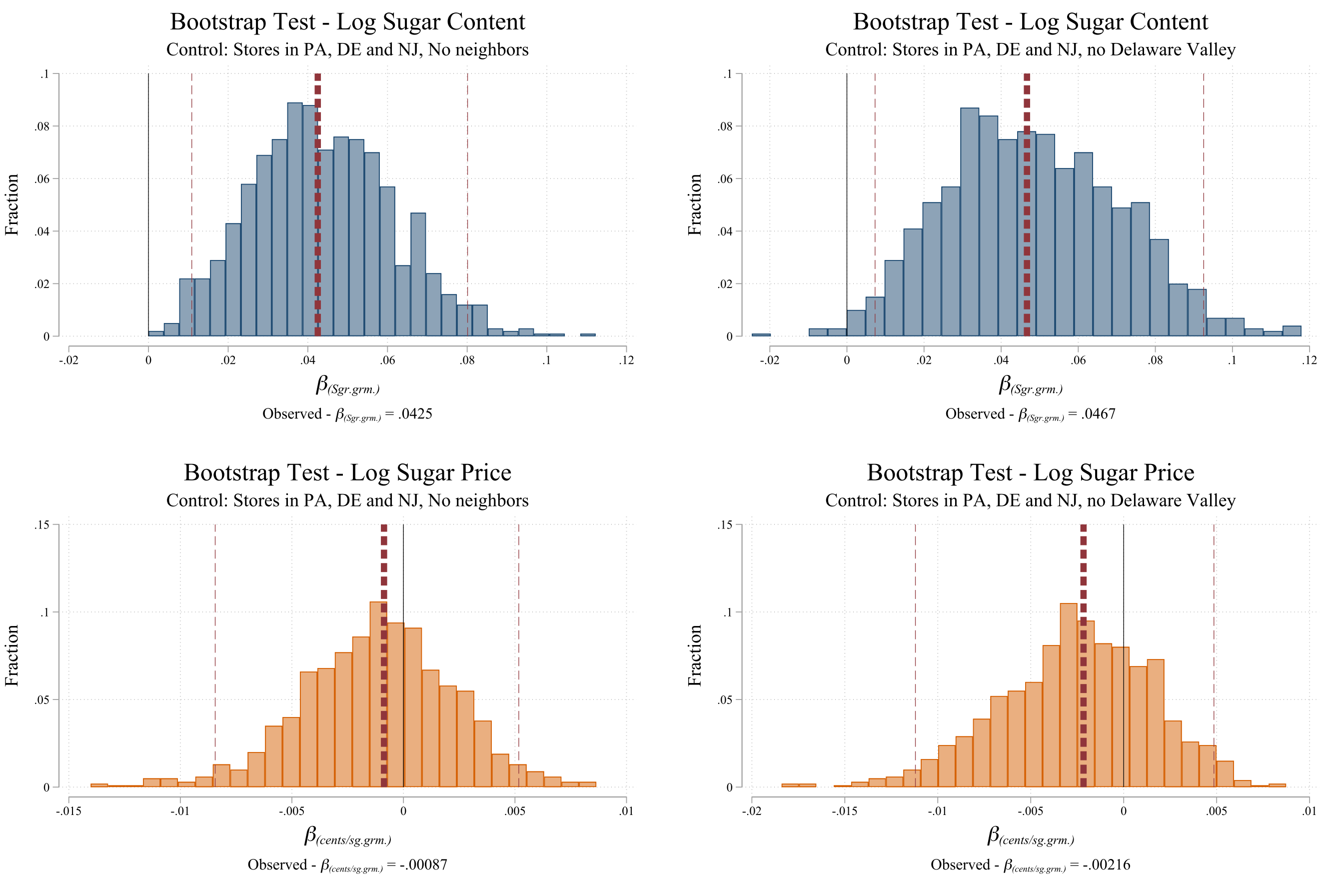

FIGURE A6 Bootstrap Distribution of Coefficients. Bootstrapped estimation of the effect of the introduction of the Philadelphia Soda Tax on prices and quantities of additional sweetened items. Bootstrap draws based on store clustering. Each unit of the histogram is the estimation of the effect using a bootstrapped sample (without replacement) for the treated and control units. For the treated group, only one cluster is available Philadelphia, the zipcode 191. For the control sample (2) Remaining PA, DE and NJ, without neighbors there are 57 clusters to draw from. For the control sample (3) Remaining PA, DE and NJ, without Philadelphia metro area, there are 49 clusters [Colour figure can be viewed at wileyonlinelibrary.com]

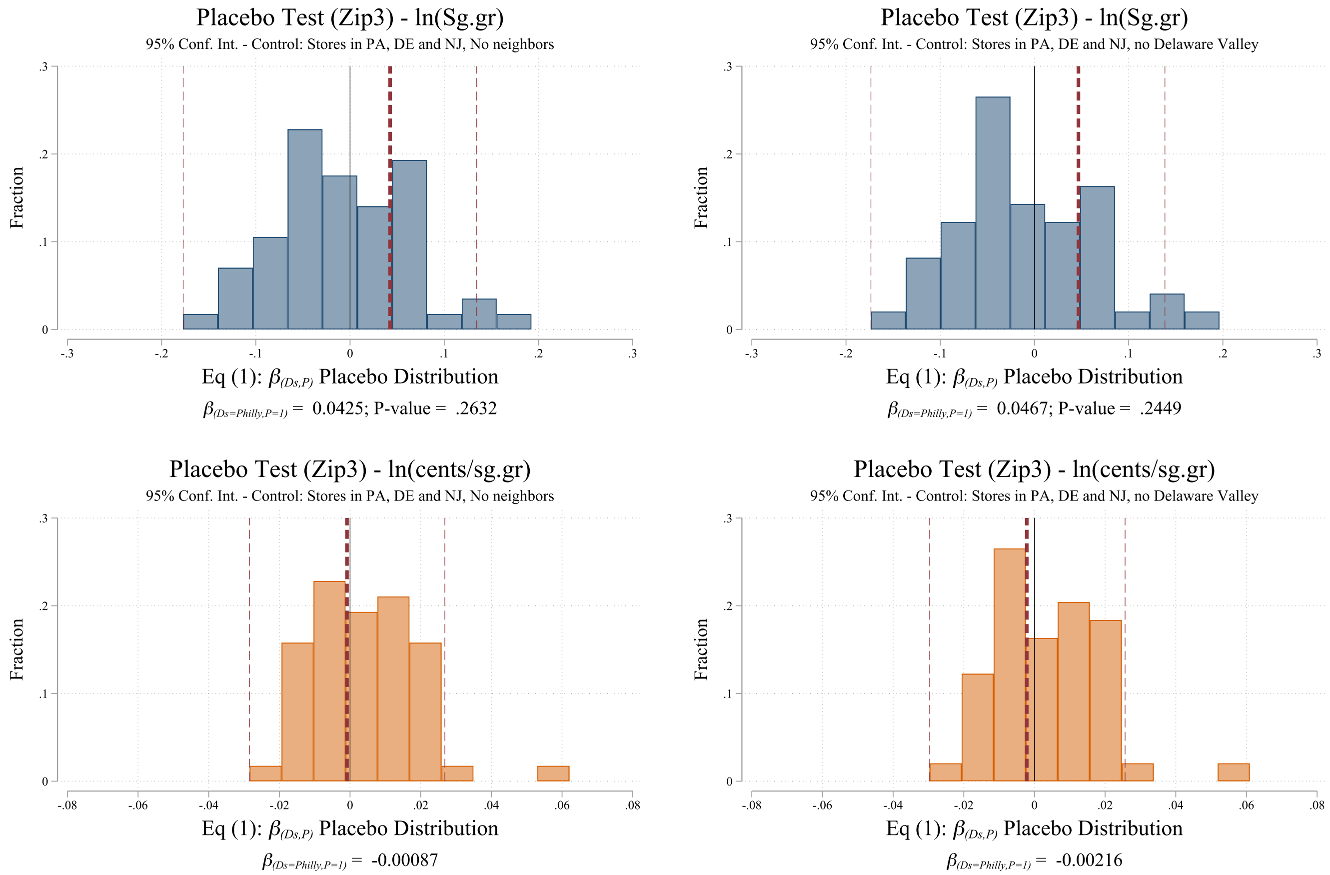

FIGURE A7 Placebo Distribution of Treatment Effects. Placebo inference distribution of the effect on prices and quantities of additional sweetened items of the introduction of a Soda Tax, as the one in Philadelphia, in different permutations of control units clustered at the zipcode level. Each unit of the histogram is the estimation of a "fake" treatment effect over all potential 3-digit zipcodes different than the one used as treatment. For the control sample (2) Remaining PA, DE and NJ, without neighbors there are 57 clusters. For the control sample (3) Remaining PA, DE and NJ, without Philadelphia metro area, there are 49 clusters [Colour figure can be viewed at wileyonlinelibrary.com]