The Impact of NuVal Shelf Nutrition Labels on Food Purchase*

Chen Zhen† and Xiaoyong Zheng‡

September 2017

Abstract

Summary shelf nutrition labels are one of a handful practical strategies that hold the promise for improving nutrition and public health. We use a difference-in-differences approach to estimate the effect of the NuVal shelf label–an interpretive numeric score that rates the overall nutrition of foods from 1 (least healthful) to 100 (most healthful)–on consumer demand for yogurt. The results indicate that NuVal labels affect product sales through an information provision effect that increases with the value of the nutrition score on the product and a publicity (i.e., salience) effect that is unrelated to the level of the score. A one-point increase in NuVal score on a yogurt product is estimated to increase demand by 0.23%.

Keywords: Summary shelf nutrition label; NuVal; Yogurt demand

JEL Classifications: I18; L15; D12

*We thank Richard Dunn, Alessandro Gavazza, Ivan Kandilov, Brad Larsen, Rigoberto Lopez, Melinda Morrill, Wally Thurman, Charles Towe, Tomislav Vukina and seminar participants at University of Connecticut, North Carolina State University and Zhejiang University for helpful comments and suggestions. All remaining errors are our own. This research was supported by a grant from Healthy Eating Research, a national program of the Robert Wood Johnson Foundation. We also thank Mike Kruger at IRI for providing the scanner data.

†314B Conner Hall, University of Georgia, Athens, GA 30602. Email: czhen@uga.edu.

‡Campus Box 8109, North Carolina State University, Raleigh, NC 27695-8109. Email: xzheng@ncsu.edu.

1 Introduction

Information disclosure is an important part of U.S. nutrition policy. Prominent examples include the Nutrition Labeling and Education Act (NLEA) of 1990 mandating standardized Nutrition Facts labels on most packaged foods by 1994 and the required disclosure of trans fat content on Nutrition Facts labels by 2006. These labeling regulations would be most effective if people choose healthier products because they see, read, and understand the often lengthy nutrition facts hidden on the back or side of the package. However, for an average consumer who makes over 200 daily food decisions (Wansink and Sobal, 2007), reviewing and processing all of this labeling information may be challenging. In fact, over the decade following NLEA's full implementation, consumer use of most nutrition labels had declined (Todd and Variyam, 2008). The obesity epidemic that has escalated post-NLEA and other health concerns associated with unhealthy food choices motivated the search for more effective labeling strategies that supplement the Nutrition Facts label.

In 2011, the Institute of Medicine's (IOM's) Committee on Examination of Front-of-Package Nutrition Rating Systems and Symbols recommended the development of a summary multiple-level nutrition symbol that goes on the fronts of packages and provides a clear ranking of the healthfulness of the labeled product (IOM, 2012). The IOM report encourages food labeling policy to shift from the current approach of providing more nutrition facts to an interpretive one that provides simple, direct, and science-based guidance to consumers on the nutritional quality of the product.

Shelf nutrition labels are a tool that provides summary information on the overall nutrition quality of a food product. They provide nutrition cues to shoppers and may be effective in promoting healthy food choices at the point of purchase. These summary labels offer one of a small handful of practical policy tools to influence consumer nutrition behavior associated with obesity and diet-related noncommunicable diseases. There are two major summary shelf nutrition label systems in the U.S.: Guiding Stars (a four-point system introduced in 2006) and NuVal (a 1 to 100 numeric system introduced in 2008). As of August 2017, 16 chain grocers had NuVal shelf labels in their stores compared with 5 retail chains using Guiding

Stars.

Research examining the impacts of summary multiple-level nutrition labels, the type IOM recommends, on actual food purchases is scarce. U.S. studies largely focus on Guiding Stars, which has shown some effect at encouraging sales of more healthful products relative to less healthful products (Sutherland, Kaley and Fischer, 2010; Rahkovsky et al., 2013; Cawley et al., 2014). However, because Guiding Stars only classifies foods into one of four categories (zero to three stars) (Fisher et al., 2011), consumers are unable to identify and switch to products that may be on the healthier end within a star category. Moreover, foods that do not earn at least one star are not labeled at the point of purchase. This raises two concerns. First, consumers are unable to distinguish rated zero-star products from unrated products. Second, consumers may not deduce the lower nutrition of unlabeled zero-star products by observing labeled healthier products that have one to three stars.[1]

In the United States, there is a proliferation of front-of-package (FOP) label systems developed by food industry that either enhance nutrition facts disclosure or feature certain healthier products. Prominent examples include the Facts Up Front label developed by the Grocery Manufacturers Association and Food Marketing Institute and Walmart's Great For You label. The Facts Up Front label makes information on select nutrients in the Nutrition Facts label more salient on the front of the package, while the Great For You label appears on Walmart's healthier private-label products. There is concern, as was the case in the United Kingdom (Hersey et al., 2013), that the different FOP label systems could confuse consumers and mitigate their potential effectiveness. One can argue that these FOP labels may not be more effective than the Nutrition Facts label because interpreting the disclosed information is as difficult (e.g., Facts Up Front)[2] or only healthier products are labeled (e.g., Great For

[1] Mathios (2000) examined pre- and post-NLEA salad dressing sales. Before NLEA, the majority of higher-fat dressings did not have nutrition label, while most low-fat dressings did in order to voluntarily disclose the positive attribute of low fat content. The author found sales of the highest-fat brands declined significantly post-NLEA. This indicates that consumers did not fully infer the high fat content of unlabeled dressings by observing low-fat dressings that were labeled.

[2] For example, in a randomized controlled trial involving 11,981 French participants, Ducrot et al. (2016)

You).

By contrast, NuVal ranks and labels both healthier and less healthy products. It scores foods on a scale from 1 (least healthy) to 100 (healthiest) derived from the Overall Nutritional Quality Index (ONQI) algorithm that profiles the content of 21 nutrients and the quality of four nutrition factors (Katz et al., 2007). The algorithm penalizes nutrients (e.g., saturated fat, sodium, and sugar) and nutrition factors generally considered to have unfavorable health effects and rewards those (e.g., fiber, potassium) that are beneficial to health. Consequently, the higher the NuVal score, the healthier the food. Unlike the 4-point Guiding Stars program, this level of granularity in NuVal score provides an easy cue for consumers to identify healthier products even when the differences may be small. Yet, small differences compounded over repeat shopping trips can result in significant health effects. A typical NuVal label is integrated into the shelf price tag (see figure 1). This design has the benefit of making the NuVal score visually accessible for price conscious shoppers.

The ONQI algorithm that underlies NuVal was developed by an expert panel independent of food industry interests (Katz et al. 2010). To test the external validity of ONQI, Chiuve et al. (2011) used ONQI to score the diet quality of over 100,000 men and women who started as healthy individuals in two longitudinal surveys spanning more than 20 years. The authors found that baseline diets that were scored lower by the ONQI algorithm are significantly associated with higher risks of chronic disease later in the surveys. In sum, the NuVal shelf label represents a validated summary expert opinion on the healthfulness of the food product.

The objective of this paper is to evaluate the effectiveness of NuVal shelf labels in promoting healthier purchases. Using a supermarket's voluntary adoption of NuVal labels as a quasi-experimental design, we compare changes in the sales of yogurt products pre- and post-adoption in the adopting supermarket to those of the same products in other stores that did not have NuVal labels while controlling for product, store, time fixed effects as well as store-specific time trends. This difference-in-differences (DID) approach allows us to obtain

found the Guideline Daily Amounts label–the French counterpart of Facts Up Front label–to have no effect on the nutrition quality of food purchases, while other three interpretive labels led to lower calories purchased.

a relatively clean estimate on the effect of the NuVal score on consumer demand.

We find that posting the NuVal score increases yogurt sales and the effect is larger for products with higher scores. This indicates that consumers respond positively to the nutrition cues provided by the NuVal score. There is also evidence for a publicity (i.e., salience) effect of the shelf label as sales of lower-scoring yogurts also increased, although to a lesser degree than yogurts with higher scores. If there were no publicity effect, we would expect sales of lower-scoring products to decrease because lower NuVal scores point to lower nutritional value. Therefore, the shelf nutrition label may have made labeled products more salient and consumers to perceive even products labeled with low scores as healthier than the unlabeled ones. These findings are robust to several specification checks. We also considered but refuted an alternative explanation for the results.

The remainder of the paper is organized as follows. The next section describes the NuVal and scanner data. Our empirical strategy is detailed in Section 3. Section 4 presents and discusses the results and the final section concludes.

2 Data

This section describes the data used in the subsequent empirical analysis. NuVal scores at the Universal Product Code (UPC) level were obtained from NuVal LLC, NuVal's licensing company. Due to the large number of food items sold in grocery stores, NuVal LLC scored products in batches over time. NuVal released scores for the first batch of yogurt products in January 2009. By August 1st, 2010, it had scored 918 yogurt products and by October 25th, 2013 (the end date for the data we have access to), it had scored 2,372 yogurt products.

From the IRI Academic Data Set (Bronnenberg et al. 2008), we obtained the retail (i.e., point-of-sale) scanner data for all yogurt products sold in the universe of six grocery stores in a small town in the Midwestern United States. One of these six stores is owned by a regional grocery chain, which adopted the NuVal label in August 2010. Among the rest, two are owned by a local food co-op, one by another regional grocery chain and two each by a local independent owner. None of these five stores adopted NuVal or Guiding Stars during our sample period. Therefore, the store that adopted the NuVal label is the treatment store and the other five stores are the control stores in our empirical analysis.

As the NuVal scores, the scanner data are at the UPC level. For each UPC, we have information, at weekly frequency, on dollar sales and the number of units sold. We calculated the price for each UPC by dividing dollar sales by number of units sold in that week. In addition, the scanner data provide information on the following product characteristics for each UPC: the manufacturer, and whether the product is organic, a Greek yogurt, soymilk-based and/or a yogurt drink. We chose to focus on the yogurt category because it is a large established food category with many varieties, providing enough variation in nutrition score to identify the effects of NuVal on sales.

Because the treatment store adopted NuVal in August 2010, we define September 2010 to December 2010 as the treatment period and April 2010 to July 2010 as the control period. As food manufacturers constantly introduce new products and phase out older ones, we chose a relatively small bandwidth of 4 months each for the treatment and control periods to minimize the impact of product entry and exit on the empirical analysis. At the treatment store, 367 UPCs were sold during the treatment period, while 374 were sold during the control period. For 155 of them, the NuVal scores were available as of August 1st, 2010 and all 155 UPCs were sold in the treatment and control periods. Out of these 155 UPCs, 130 were sold in the control stores during both the treatment and control periods. These 130 UPCs are therefore the treated UPCs included in the DID model. To shed light on the differences between the treated and untreated UPCs, Table 1 presents the summary statistics for the treated and other UPCs in the treatment store during the control period. On average the treated UPCs had a lower unit price and larger quantities sold.[3] This may suggest that the most popular yogurts are scored first. Also, the treated UPCs are more likely to be produced by General Mills (owner of the Yoplait brand) and Dannon, the two leading firms in this industry, and

[3] A lower unit price does not necessarily mean the product is cheaper on a per volume basis as package sizes differ for different products.

less likely to be organic or Greek, the latter of which gained popularity after the end of our sample period.

Our DID analysis focuses on the 130 treated UPCs defined above, comparing changes in their sales quantities before and after the NuVal labels were adopted in the treatment store versus those in the control stores. From Table 1, we can see that the NuVal scores for the 130 UPCs range from 23 to 100, with an average score of 51.8. Table 1 also shows that 52% of the 130 UPCs are produced by General Mills, 33% by Dannon and the rest by other smaller manufacturers. In terms of product features, 4.6% of them are organic, 5.4% are Greek, 3.1% are soymilk-based and 5.4% are yogurt drinks. On average, about 36 units were sold per week for each UPC, with an average price of $1.43 per unit. There were very little advertising and promotion activities at the stores. Only 1.22% and 0.79% of the observations (the unit of observation is one UPC in one week in one store) come with a minor display and a medium-size ad, respectively, and there was zero occurrence of large display or ads of other sizes.

To examine the differences between the treatment store and the control stores, Table 2 provides summary statistics for the yogurts sold in these stores during the control period. The table shows that the treatment store was the market leader in yogurt sales in this small Midwestern town during our sample period. It carried 130 treated UPCs and 244 other UPCs during the control period. The five control stores carried between 83 and 124 treated UPCs and between 92 and 204 other UPCs. The treatment store also sold more units per product. For the treated UPCs, it sold about 36 units per UPC per week, while the numbers at the control stores range from 9 to 34. For other UPCs, the treatment store sold about 19 units per UPC per week, while the numbers at the control stores range from 7 to 19. The unit prices charged by the treatment store for the treated UPCs were similar to those of the control stores except control store 4, which carried a significantly less number of yogurt products. The average unit price at the treatment store for the treated UPCs was $1.43, while the unit prices at the four control stores other than control store 4 range from $1.42 to $1.46. The average price per unit at the treatment store for other UPCs was slightly higher than those of the control stores at \$2.12, while the unit prices at the control stores range from \$1.45 to \$1.89. Overall, although the treatment store was the market leader at this town, the magnitudes of differences between the treatment store and four out of the five control stores were not particularly large in terms of product variety, price, and sales.

2.1 Differences in Means

Before turning to the regression analysis, we first use summary statistics to examine the NuVal treatment effect. Table 3 provides the average weekly quantity sold and price paid for the treated UPCs before and after the NuVal labels were adopted and at both the treatment store and the control stores. The first two columns of the table report the summary statistics for all treated UPCs, regardless of their NuVal scores. In the treated store, the average quantity sold for each treated UPC increased from 36.18 units to 45.00 units after the NuVal labels were adopted. The average unit price changed very little with a slight increase of less than 1 cent. During the same time, at the control stores, the average quantity sold for each treated UPC decreased from 24.65 units to 19.34 units, with the average unit price decreasing by about 3 cents. The difference in differences in the means suggests that posting NuVal labels increased demand for the treated UPCs by 14.13 units, or about 39% of 36.18, the average quantity sold in the treatment store before the NuVal labels were adopted.

In the remaining columns of Table 3, we investigate whether the treatment effect is different for UPCs with higher and lower NuVal scores. Columns 3 and 4 report summary statistics in the treatment store for the treated UPCs with scores higher than or equal to 50, while columns 5 and 6 report the same descriptive statistics for treated UPCs with scores lower than 50.[4] Table 3 illustrates that in the treated store, the increase in the average quantity sold of the treated UPCs is driven by an increased demand for high-scoring UPCs,

[4] We selected the cutoff point of 50 for two reasons. First, it is close to the mean score of 51.8 for the treated UPCs reported in Table 2. Second, as the scores are designed to range from 1 to 100, 50 would be a natural cutoff point if consumers believe, as a rule of thumb, that UPCs with scores higher than 50 are healthy products and those below 50 are unhealthy ones.

which increased from 51.17 units to 70.40 units. Sales of the low-scoring UPCs, however, decreased from 21.91 to 20.83 units on average. In the control stores, high-scoring UPCs decreased from 31.20 to 24.54, while low-scoring UPCs decreased from 17.55 to 13.59. In terms of the difference in differences, higher-scoring UPCs had a treatment effect of an increase of 25.89 units, whereas lower-scoring UPCs experienced an average increase of 2.88 units.

Although the differences in means results are informative, these simple comparisons do not take into account the effects of many observed as well as unobserved factors on consumer demand. Our empirical strategy below uses regression analysis to control for these factors and analyze the treatment effect in a more rigorous way.

3 Empirical Strategy

We use DID regressions to examine the effect of NuVal labels on the treated UPCs, controlling for many other observed and unobserved factors that affect consumer demand for yogurt products. We run the following regression using data on the treated UPCs in both the treatment and the control stores and during both the control and the treatment periods,

$$\log Q_{ist} = \beta_0 + \beta_1 D_s + \beta_2 D_t + \beta_3 D_s * D_t + \beta_4 \log p_{ist} + \alpha_i + \gamma_s + \delta\gamma_s * t + \lambda_t + \varepsilon_{ist}, \quad (1)$$

where $\log Q_{ist}$ is the logarithm of the number of units sold in store s during week t for UPC i, D_s is a dummy variable indicating the treatment store, D_t is a dummy variable indicating the treatment period and $\log p_{ist}$ is the log unit price for UPC i in store s during week t. α_i is the time- and store-invariant UPC fixed effect controlling for unobserved product characteristics that are likely to affect consumer demand. Among other things, it captures the part of knowledge consumers may already have about the nutrition of product i through past experience, the Nutrition Facts label, and any voluntary FOP labels and labeling claims that exist prior to NuVal adoption. γ_s is the store fixed effect controlling for the UPC- and time-invariant store-specific factors (e.g., amenity) that may affect consumer demand. λ_t is the week fixed effect controlling for store- and UPC-invariant time-specific factors (e.g., seasonality) that may affect consumer demand, and ε_{ist} is the error term. Finally, t is the time (weekly) trend variable and hence $\gamma_s * t$ represents store-specific time trends in consumer demand. The reason we include this term is explained in detail in the identification subsection below. The coefficient of interest is β_3, which can be interpreted as the average treatment effect on the treated. Note that in (1), the term $\beta_1 D_s$ is subsumed in the store fixed effects γ_s and the term $\beta_2 D_t$ is subsumed in the time fixed effects λ_t. Hence, coefficient estimates for β_1 and β_2 will not be reported.

Although useful for examining the average treatment effect of the NuVal label on the treated UPCs, equation (1) does not address the extent to which the NuVal label effect is related to new nutrition information provided by the score versus a salience effect from the label. To distinguish between the two effects, we interact NuVal score with the treatment indicator. If consumers consider the information embedded in the score as new and have higher preferences for healthier yogurts, then treated UPCs with higher scores should experience a higher increase in quantity sold compared with treated UPCs with lower scores. Alternatively, if the only effect of the NuVal label is to make the labeled product more salient, then the treatment should have the same impact across products regardless of their NuVal scores.

To distinguish between the two effects, we estimate the following equation

$$\begin{aligned} \log Q_{ist} &= \beta_0 + \beta_1 D_s + \beta_2 D_t + \beta_3 D_s * D_t + \theta S_i * D_s * D_t \\ &\quad + \beta_4 \log p_{ist} + \alpha_i + \gamma_s + \delta \gamma_s * t + \lambda_t + \varepsilon_{ist}, \end{aligned} \tag{2}$$

where S_i is the NuVal score for UPC i and $\log p_{ist}$ is the log unit price for UPC i in store s during week t. The θ parameter allows us to examine how the treatment effect varies with the NuVal score. If θ is estimated to be positive and statistically significant, then we can conclude that the treatment effect is higher for UPCs with higher NuVal scores. Such a finding would suggest the NuVal score contains new information valuable to consumers.

3.1 Identification

The validity of our empirical strategy above depends critically on two identification assumptions. First, estimation of the average treatment effect depends on the assumption that treatment is independent of potential outcomes (Imbens, 2004), that is, the treatment variables in (1) and (2) are exogenous. This is also referred to as the "common trends" assumption in the literature, which in our context means the sales trends of the treated UPCs would be the same in the treatment and control stores in absence of the NuVal label adoption. This assumption would be violated if the treatment store timed when to introduce the NuVal labels based on some unobserved demand factors. We believe this is a highly unlikely scenario for two reasons. First, the decision to adopt the NuVal labels was a strategic decision made at the retail chain level to cater more to health conscious consumers, not by the manager of the particular treatment store. Second, even if the decision was made in response to changes in unobserved demand factors, the response was to changes in the overall demand across all stores owned by the chain, not to the idiosyncratic demand shocks at the treatment store. Nonetheless, we follow Acemoglu, Autor and Lyle (2004) and Hoynes and Schanzenbach (2009) to include interactions of pre-treatment variables with time trends in the regressions to make this assumption more likely to hold. Specifically, we include store-specific time trends in the regressions, following the recommendation by Angrist and Pischke (2009, pp. 239.). We will also further examine this assumption below after estimation.

Second, the price variable in demand equations (1) and (2) could potentially be endogenous. Endogeneity bias could arise if stores condition on time-varying and store-specific unobservable (to the econometrician) factors that affect demand when they set prices. However, Dube, Hitsch and Rossi (2010) argue that price variation in scanner data is mainly across brands and only a small percentage of variation is explained by store and time effects, which are all controlled for by the UPC, store, and time fixed effects. Any remaining endogeneity bias is likely to be trivial. Hendel and Nevo (2013) also argue that potential correlation between prices and the error term is not a major concern in their demand estimation using retail scanner data. Nonetheless, we follow Nevo (2001) and others in the literature to use "Hausman" type instrumental variables (Hausman, 1997). Specifically, we use average prices for the same UPCs during the same week in nearby cities in the same Census region as the Midwestern town as our instrument for the price variable and estimate (1) and (2) using two-stage least squares (2SLS).

4 Results

The final sample of dataset used to estimate equations (1) and (2) come from data on the 130 treated UPCs from the 4-month control period and the 4-month treatment period at both the treatment and control stores. The total number of observations is 23,758. The unit of observation is one UPC in one store during one week. Definitions and the summary statistics for the variables used in estimation are presented in Table 4. About 20% of the observations are from the treatment store and 50% of the observations are from the treatment period. As a result, about 10% of the observations are from the treatment store during the treatment period.

Estimation results for equation (1) are reported in Table 5. Both the OLS and 2SLS methods are used. Week, store and UPC fixed effects as well as store-specific time trends are included in the regressions. The standard errors are adjusted for clustering at the both the UPC and week levels. Therefore, the reported standard errors are robust to correlations and heteroskedasticity among observations for the same UPC or in the same week.

As results from both methods are similar, we focus our discussion on the 2SLS results. First, the R^2 is close to 0.75, indicating we have included the most relevant explanatory variables for yogurt sales as they together explain a large fraction of the variation in the dependent variable. Second, the first-stage F statistic is 34.10, rejecting the hypothesis that our instrumental variable is weak. Third, the estimated price elasticity is -2.99 and the estimate is statistically significant. This indicates that demand for yogurts at the UPC level is quite price-elastic. Fourth, the estimated average treatment effect on the treated is 25.16% and statistically significant, meaning that on average, posting the NuVal label boosts sales of the treated UPCs at the treatment store by 25.16% relative to the control stores. This effect is equivalent to that of a 8.44% price reduction for treated UPCs.

4.1 Heterogeneity of the Treatment Effect

To test whether the treatment effect varies across UPCs with different NuVal scores and prices, we interact the NuVal score variable with the treatment variable. Regression results for equation (2) are displayed in Table 6. As in Table 5, both OLS and 2SLS are used. But before we discuss the results, we first use the estimation results to further examine the "common trends" assumption discussed above, which is critical for the validity of our analysis here. We do not expect the quantities sold ($\log Q_{ist}$, the dependent variable in our regressions) from the control stores and the treatment store to have a common trend, even during the control period. This is because (1) and (2) show that consumer demand depends on many factors such as price, unobserved store characteristics, seasonal effects and store-specific time trends, in addition to the labeling effect we are after and these factors are different for different stores during different weeks. However, if our model is specified correctly and the "common trends" assumption holds, then there should be no clear different patterns among the regression residuals for those observations from the control stores and those from the treatment store because we have already controlled for the treatment variables as well as other demand shifters in the regressions. By contrast, the extreme opposite case where all the residuals for observations from the control stores are first negative and then become positive while all the residuals for observations from the treatment store are first positive and then become negative over time (such that the overall sum is still zero as by construction due to one of the properties of OLS and 2SLS regressions) clearly indicates the underlying trends are not the same, even after controlling for various demand shifters. We use residuals from the 2SLS estimation of (2) to examine whether this is true. Figure 2a shows the time series plot of the residuals for observations from the control stores while Figure 2b shows the same plot for observations from the treatment store. The figures do not reveal any clear pattern differences between the two groups of residuals. All the residuals, for observations both from the control stores, scatter randomly around zero during all weeks, lending support to our specification and the associated "common trends" assumption.

Again, as results from both the OLS and 2SLS are quite similar, we focus our discussion below on the results from 2SLS. The estimated treatment effect is

$$0.1420 + 0.0023 * S_i. \tag{3}$$

Eq. (3) shows that the treatment effect is larger for UPCs with a higher NuVal score. The lowest NuVal score for the treated UPCs in our dataset is 23. The estimated treatment effect for a UPC with a score of 23 is 0.1949. This means that compared with the demand in the control stores, demand for this UPC increases by 19.49% with the posting of the NuVal label. The highest NuVal score for the treated UPCs in our dataset is 100. The estimated treatment effect for a UPC with a score of 100 is 0.372. This means that compared with the demand in the control stores, demand for this UPC increases by 37.2% with the posting of the NuVal label.

Therefore, the treatment effect is positive for all treated UPCs and is positively associated with the NuVal score. This result implies that NuVal contains new information for consumers and preferences for yogurts increase with the nutrition value of the product. The publicity effect is at work as well. If there were no publicity effect, demand for the lower-scoring UPCs would decrease as these labels signal low product nutrition to consumers. The NuVal label may have made the treated UPCs more salient and thus increased demand for even the lowest-scoring yogurts. Alternatively, consumers may believe, albeit incorrectly, that even the lower-scoring products are better in nutrition than unscored products. Consequently, demand for the lower-scoring products also increases. This explanation, if correct, would be consistent with one of the fundamental assumptions underlying the famous information unraveling theory in the quality disclosure literature (e.g., Grossman 1981; Milgrom 1981)– that rational consumers would infer nondisclosure as having the lowest quality.[5]

[5] The information unraveling theory predicts that firms have the incentive to voluntary disclose product quality on all but the worst products. However, voluntary unraveling of information may not be complete if

4.2 Robustness Checks

To check the sensitivity of our findings to alternative specifications, we perform several robustness checks. The key identifying assumption in our DID regression analysis is that trends in the sales of the treated UPCs would have been the same in both the treatment store and the control stores during the treatment period if NuVal labels were not adopted. If the trends were not the same, then part of the estimated treatment effect could simply reflect the different sales trends in the treatment and control stores and, hence, our estimate of the treatment effect would be biased. We have included store-specific time trends in our regressions to make this assumption more likely to hold in our empirical analysis. Here, we perform an additional robustness check to examine the validity of this assumption.

In the DID analysis above, we include all grocery stores other than the treatment store in the control group. From Table 2, it is clear that some control stores are more similar to the treatment store than others in terms of the number of treated UPCs offered, number of units sold per week and unit price. The advantage of including all control stores is that we have a larger sample and the parameters are more precisely estimated. The disadvantage of doing so is that by including those control stores that are not very similar to the treatment store, the key identifying assumption that sales trends are similar in the treatment store and the control stores in absence of treatment may fail. In our first robustness check, we only include one store in the control group. The selected store is control store 2. From Table 2, it is clear that this is the control store that is most similar to the treatment store during the control period. Only specification (2) is estimated and results are reported in Table 7. Compared with the results in Table 6, we find the value of the coefficient estimate for the base treatment status variable $D_s * D_t$ is now negative and the magnitude of the coefficient

manufacturers do not have a credible method of labeling (Shavell, 1994), disclosure is costly (Farrell, 1986), not enough consumers value the aspect of quality being disclosed (Fishman and Hagerty, 2003), or consumers do not equate nondisclosure with inferior product quality (Grossman, 1981). Although adoption of NuVal is a voluntary decision on the part of the retailer, food manufacturers do not have a say in which products should or should not be scored. This is different from most cases considered in the information unraveling literature where producers make the labeling decisions.

estimate for the heterogenous treatment effect variable, $S_i * D_s * D_t$, is larger. The estimated treatment effect is,

$$-0.2393 + 0.0042 * S_i, \tag{4}$$

which implies that the treatment effect is only positive for those UPCs with a NuVal score that is at least 57. This further strengthens our conclusion that NuVal scores contain new information to consumers and consumers respond to the higher nutrition quality signaled by higher scores with higher demand.

Our second robustness check is a set of falsification or placebo tests. In each of the placebo test, a single control store is used as a placebo for the treatment store and the remaining four stores are used as the control stores. Both specifications (1) and (2) are estimated for each placebo test and the estimation results are displayed in Table 8. Results show there is no clear and coherent evidence for the false NuVal treatment effect in each of the placebo test, lending support to the supposition that our main results above are not driven by some other unobserved (by the econometrician) demand shifters. In tests where the first three control stores are used as the placebo, the estimated treatment effect does not vary with the NuVal score in the sense that the estimated coefficient for the treatment status and NuVal score interaction variable is not statistically significant. In the test where the fifth control store is used as the placebo, the estimated treatment effect decreases with the NuVal score. In the test where the fourth control store is used as the placebo, the estimated overall treatment effect is negative, driven by the large, negative and significant estimate of the coefficient on the treatment status variable.

4.3 Potential Biases

Although our results are robust to several specification checks, it is important to bear in mind that they are still subject to two potential biases. First, if consumers shop at multiple stores and learn about (and memorize) the nutritional value of some products in the treatment store, they may take this into account when buying the same products in the control stores. If this is true, then our estimate would be a combination of the NuVal effect as well as this multi-store shopping effect. While the NuVal effect makes sales of the treated products in the treatment store and in the control stores diverge, the multi-store shopping effect closes the gap. As a result, our estimate would be an underestimate of the true effect of the label. To examine how much multi-store shopping was happening, we take advantage of another dataset that tracks grocery purchases for a panel of households from the same town where the grocery stores are located. In our sample, 608 consumers purchased yogurt at the treatment store during the treatment period. Out of the 608 consumers, 182, 111, 91, 28 and 158 also purchased yogurt at control store 1, 2, 3, 4 and 5, respectively, during the same period. This shows that multi-store shopping is fairly common for consumers in this market. This is not surprising as this is a small town with an area of about 30 square miles and the grocery stores are not far away from one another. To further examine how the multi-store shopping effect may affect our estimates, we re-estimate our DID regressions using only control stores 3 and 4 in the control group. These two control stores had the least number of customers who also shopped at the treatment store. Results are collected in Table 9. Compared with the results in Tables 5 and 6, the estimated treatment effect on the treated is only slightly different. For example, for specification (2), the average treatment effect is estimated to range from 17% to 36.25%, compared with 19.49% to 37.2% as results from Table 6 imply. Thus, we conclude that the NuVal effect is slightly underestimated due to the multiple store shopping effect.

Second, our results show that the positive effect of NuVal on yogurt demand is larger for products with higher scores. If the effect on demand is so large for some products such that they become out of stock, then our estimate would again underestimate the true effect of NuVal. We do not believe stockout due to NuVal-induced demand spike is likely for two reasons. First, our results show that the largest estimated effect of NuVal is a 37% increase in quantity demanded, which is not a very large percentage. Second, modern supply chain management allows grocery stores to restock rather quickly in a stockout event.

4.4 An Alternative Interpretation

Our main result above shows that posting NuVal labels leads to an increase in consumer demand for yogurt products, especially yogurts with higher NuVal scores. Our interpretation of this result is that NuVal has both a salience effect and an information provision effect that promote demand for healthier yogurts among the treatment store's existing customers. A competing interpretation of our result is that NuVal labels at the treatment store attracted health conscious consumers from the control stores to the treatment store. That is, there may have been a change in the composition of consumer types at the treatment store after NuVal adoption. As a result, demand for yogurt products, especially healthier ones, increases relative to that of the control stores because more health conscious consumers now shop at the treatment store. This store switching effect, if significant, would bias our estimate upward. Unlike the aforementioned multi-store shopping effect that would only change the magnitude of the estimated treatment effect, the store switching effect would lead to a qualitatively different conclusion. Therefore, it is important to examine whether this alternative interpretation is empirically plausible.

To do this, we again use the dataset that tracks grocery purchases by a panel of households. First, we define the treatment period as the 4 months after the introduction of NuVal labels. Four months is a relatively short amount of time and many consumers who only shopped at the control stores during the control period probably had not learned about the NuVal labels by the end of our sample period. Indeed, we notice that in our sample, out of the 1,006 households that ever purchased yogurt but never at the treatment store during the control period, 2.28% (23 households) started purchasing yogurt only in the treatment store during the treatment period. These are the consumers who switched from the control stores to the treatment store. On the other hand, out of the 188 households that purchased yogurt only at the treatment store during the control period, 6.38% (12 households) started purchasing yogurt only at the control stores during the treatment period. These are the consumers who switched from the treatment store to the control stores. If the main reason behind store switching were the availability of NuVal at the treatment store, then we would expect to see a significantly larger percentage of consumers switching from control stores to the treatment store than vice versa. The statistics here do not support such a hypothesis. Second, as this dataset tracks household-level purchases, we know who shopped at which stores each week. Using these information, we created store and week-specific average household characteristics, averaging across all households that shopped at a given store in a given week. The statistics include average income, average household size, average number of children, and the percentages of households headed by a person with college degree, less than 35 years old, between 35 and 64, and 65 years old and above. These variables are used as additional controls in the regressions to control for any effect from store switching by consumers. Regression results are collected in Table 10. As we can see, the results are very similar to those in Tables 5 and 6. Thus, we reject store switching by the consumers as an alternative interpretation for our main results.

5 Conclusion

In the absence of a federal mandate on FOP nutrition labels, NuVal is one of the few commercially available multiple-level interpretive summary nutrition labels in the U.S. that rate both healthier and less healthy food products. Given the breadth of NuVal's coverage in retail partners (16 retail chains as of August 2017) and number of products (over 100,000 UPCs), it is important to evaluate its performance in achieving its stated public health goal of promoting healthy food purchase.

Using yogurt as a case study, we found NuVal to have an information provision effect and a publicity effect on yogurt demand. The information provision effect promotes demand for healthier yogurts by offering information through the NuVal score that is new to consumers. A one-point increase in the NuVal score is estimated to increase demand for the rated yogurt by about 0.23% (Table 6). The publicity effect increases demand for healthier and less healthy yogurts equally by making the labeled products more salient.

Moving forward, several questions remain to be addressed. First, as we only focus on the 4-month window around the adoption of NuVal, our analysis is short-run in nature. In the long run, consumers who shop at the treatment store may learn about the healthfulness of the products through NuVal and pass the information to shoppers at other stores. In this case, even if other stores do not adopt the NuVal label, there will be a spillover effect. Empirically quantifying this effect is challenging because product entry and exit can more easily contaminate identification of the main effects of interest in the longer run. Second, it will interesting to know if the salience effect is sustained over time as the percentage of scored and labeled products grows and novelty of the label fades. Third, any interpretive summary nutrition label is necessarily a one-size-fits-all approach to health promotion. However, the population is not consisted of a homogenous group of consumers. Rather, consumers are heterogeneous in health status, nutritional need, education, financial resources, and food preferences. Some consumers stand to benefit more from summary nutrition labels than others. It will be useful to formally account for these heterogeneities and examine how different consumer segments may respond differently to NuVal labels. These questions are beyond the scope of this paper and are left for future research.

REFERENCES

Acemoglu, D., D. H. Autor and D. Lyle (2004): "Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury," *Journal of Political Economy*, 112, 3, 497-551.

Bronnenberg, B. J., M. W. Kruger and C. F. Mela (2008): "The IRI Marketing Data Set." *Marketing Science*, 27, 745-748.

Cawley, J., M. J. Sweeney, J. Sobal, D. R. Just, H. M. Kaiser, W. D. Schulze, E. Wethington and B. Wansink (2014): "The Impact of a Supermarket Nutrition Rating System on Purchases of Nutritious and Less Nutritious Foods." *Public Health Nutrition*, 18, 8-14.

Chiuve, S. E., L. Sampson, and W. C. Willett (2011): "The Association between a Nutritional Quality Index and Risk of Chronic Disease." *American Journal of Preventive Medicine*, 40, 505–513.

Dube, J.-P., G. J. Hitsch and P. E. Rossi (2010): "State Dependence and Alternative Explanations for Consumer Inertia." *Rand Journal of Economics*, 41, 3, 417-445.

Ducrot, P., C. Julia, C. Méjean, E. Kesse-Guyot, M. Touvier, L. K. Fezeu, S. Hercberg and S. Péneau (2016): "Impact of Different Front-of-Pack Nutrition Labels on Consumer Purchasing Intentions: A Randomized Controlled Trial." *American Journal of Preventive Medicine*, 50, 5, 627-636.

Farrell, J. (1986): "Voluntary Disclosure: Robustness of the Unraveling Result, and Comments on Its Importance." In *Antitrust and Regulation*, edited by R. E. Grieson. Lexington, Mass: Lexington Books, 1986.

Fisher, L. M., L. A. Sutherland, L. A. Kaley, T. A. Fox, C. M. Hasler, J. Nobel, M.A. Kantor and J. Blumberg (2011): "Development and Implementation of the Guiding Stars Nutrition Guidance Program." *American Journal of Health Promotion*, 26, 2, 55-63.

Fishman, M. J. and K. M. Hagerty (2003): "Mandatory Versus Voluntary Disclosure in Markets with Informed and Uninformed Customers." *Journal of Law, Economics, & Organization*, 19, 1, 45-63.

Grossman, S. J. (1981): "The Informational Role of Warranties and Private Disclosure about Product Quality." *Journal of Law and Economics*, 24, 3, 461-483.

Hendel, I. and A. Nevo (2013): "Intertemporal Price Discrimination in Storable Goods Markets." *American Economic Review*, 103, 7, 2722-2751.

Hersey, J. C., K. C. Wohlgenant, J. E. Arsenault, K. M. Kosa and M. K. Muth (2013): "Effects of Front-of-Package and Shelf Nutrition Labeling Systems on Consumers." *Nutrition Reviews*, 71, 1, 1-14.

Hoynes, H. W. and D. W. Schanzenbach (2009): "Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program," *American Economic Journal: Applied Economics*, 1, 4, 109-139.

Hausman, J. (1997): "Valuation of New Goods Under Perfect and Imperfect Competition," in *The Economics of New Goods*, Studies in Income and Wealth Vol. 58, ed. by T. Bresnahan and R. Gordon. Chicago: National Bureau of Economic Research.

Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *Review of Economics and Statistics*, 86, 4-29.

Institute of Medicine (2012): Front-of-package Nutrition Rating Systems and Symbols: Promoting Healthier Choices. Washington, DC: The National Academies Press.

Katz, D. L., V. Y. Njike, D. Kennedy, and J. Treu (2007): Overall Nutritional Quality Index, version 1: reference manual. Derby CT: Yale University School of Medicine. http://www.nuval.com/images/upload/file/ONQI%20Manual%205_5_09.pdf

Katz, D. L., V. Y. Njike, L. Q. Rhee, A. Reingold and K. T. Ayoob (2010): "Performance Characteristics of NuVal and the Overall Nutritional Quality Index (ONQI)." *American Journal of Clinical Nutrition*, 91, 1102S–1108S.

Kruger, M.W. and D. Pagni (2012): Academic Data Set Description, Version 2.0, Analytics Research and Development, IRI. November 29, 2012.

Mathios, A. D. (2000): "The Impact of Mandatory Disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market." *Journal of Law and Economics*, 43, 2, 651-677.

Milgrom, P. R. (1981): "Good News and Bad News: Representation Theorems and Applications." *Bell Journal of Economics*, 12, 2, 380-391.

Nikolova, H. D. and J. J. Inman (2015): "Healthy Choice: The Effect of Simplified Point-of-Sale Nutritional Information on Consumer Food Choice Behavior." *Journal of Marketing Research*, 52, 817-835.

Rahkovsky, I., B.-H. Lin, C.-T. J. Lin and J.-Y. Lee (2013): "Effects of the Guiding Stars Program on Purchases of Ready-to-eat Cereals with Different Nutritional Attributes," *Food Policy*, 43, 100-107.

Shavell, S. (1994): "Acquisition and Disclosure of Information Prior to Sale," *RAND Journal of Economics*, 25, 20-36.

Sutherland, L. A., L.A. Kaley and L. Fischer (2010): "Guiding Stars: The Effect of a Nutrition Navigation Program on Consumer Purchases at the Supermarket." *American Journal of Clinical Nutrition*, 91(suppl), 1090S–1094S.

Todd, J. E. and J. N. Variyam (2008): "The Decline in Consumer Use of Food Nutrition Labels, 1995-2006." Economic Research Report Number 63. Washington, DC: Economic Research Service, U.S. Department of Agriculture.

Wansink, B. and J. Sobal (2007): "Mindless Eating: The 200 Daily Food Decisions We Overlook," *Environment and Behavior*, 39, 106–123.

Table 1 Descriptive Statistics for Treated and Other UPCs in the Treatment Store during the Control Period

Variable	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max	Obs
	Treated UPCs					Untreated UPCs				
NuVal Score	51.80	25.58	23	100	130					
Units sold	36.18	49.14	1	679	2290	18.50	38.62	1	711	4076
Unit price	1.43	1.12	0.35	4.79	2290	2.12	1.44	0.38	7.39	4076
General Mills	0.52	0.50	0	1	130	0.20	0.40	0	1	244
Dannon	0.33	0.47	0	1	130	0.32	0.47	0	1	244
Soy	0.03	0.17	0	1	130	0.03	0.17	0	1	244
Greek	0.05	0.23	0	1	130	0.10	0.30	0	1	244
Yogurt Drink	0.05	0.23	0	1	130	0.06	0.23	0	1	244
Organic	0.05	0.21	0	1	130	0.20	0.40	0	1	244

Table 2 Descriptive Statistics for UPCs in Different Stores during the Control Period

Variable	#	NuVal Score	Units Sold	Unit Price	#	Units Sold	Unit Price
			Treated UPCs			Other UPCs	
Treatment Store	130	51.80	36.18	1.43	244	18.50	2.12
		(25.58)	(49.14)	(1.12)		(38.62)	(1.44)
Control Store 1	124	52.00	21.82	1.42	193	13.41	1.91
		(25.51)	(33.16)	(1.06)		(36.38)	(1.57)
Control Store 2	123	52.61	34.48	1.43	204	18.52	1.99
		(25.93)	(62.53)	(1.09)		(53.70)	(1.58)
Control Store 3	118	53.42	26.40	1.46	204	13.15	1.95
		(26.14)	(65.46)	(1.09)		(29.33)	(1.57)
Control Store 4	83	48.94	9.00	1.18	92	6.92	1.45
		(22.73)	(8.11)	(1.00)		(6.94)	(1.09)
Control Store 5	119	52.94	25.96	1.42	200	10.48	1.89
		(25.60)	(49.62)	(1.11)		(19.80)	(1.53)

Note: Standard deviations are in parentheses.

Table 3 Differences in Means

	Treatment Store	Control Stores	Treatment Store & Score >=50	Control Stores & Score>=50	Treatment Store & Score<50	Control Stores & Score<50
Units (Control Period)	36.18 (49.14)	24.65 (50.91)	51.17 (65.13)	31.20 (61.61)	21.91 (16.06)	17.56 (34.52)
Units (Treatment Period)	45.00 (105.03)	19.34 (37.89)	70.40 (145.26)	24.54 (46.12)	20.83 (16.22)	13.59 (24.71)
Price (Control Period)	1.43 (1.12)	1.40 (1.08)	1.07 (0.86)	1.09 (0.81)	1.77 (1.22)	1.73 (1.23)
Price (Treatment Period)	1.43 (1.11)	1.37 (1.05)	1.06 (0.85)	1.07 (0.78)	1.79 (1.21)	1.70 (1.19)
Number of UPCs	130	130	63	63	67	67

Note: Standard deviations are in parentheses.

Table 4 Summary Statistics for the Regression Sample

	Mean	Std. Dev.	Min	Max
Log of Units Sold	2.44	1.22	0	7.19
Log of Unit Price	0.08	0.68	-1.27	1.61
D_s	0.20	0.40	0	1
D_t	0.50	0.50	0	1
$D_s * D_t$	0.10	0.30	0	1
IV for log price	-0.07	0.68	-0.87	1.41

Notes: total number of observations: 23,578. D_s: treatment store dummy variable. D_t: treatment period dummy variable. IV for log price: average price for the same UPC in the same week in nearby cities in the same Census region.

Table 5 Estimation Results for Specification (1)

	OLS		2SLS	
	Estimate	Std. Err.	Estimate	Std. Err.
$D_s * D_t$	0.1563**	0.08	0.2516*	0.15
Log of Unit Price	-2.3973***	0.19	-2.9828***	0.90
UPC Fixed Effects	included		included	
Store Fixed Effects	included		included	
Week Fixed Effects	included		included	
Store-specific Trends	included		included	
Adjusted R-squared	0.75		0.75	
1st-Stage F statistic			34.10	

Notes: ** and *** denote statistical significance at 5% and 1%, respectively. D_s: treatment store dummy variable. D_t: treatment period dummy variable.

Table 6: Estimation Results for Specification (2)

	OLS		2SLS	
	Estimate	Std. Err.	Estimate	Std. Err.
$D_s * D_t$	0.0443	0.10	0.1420	0.17
Log of Unit Price	-2.3970***	0.19	-3.0304***	0.90
$D_s * D_t$* Score	0.0021***	0.0011	0.0023***	0.0011
UPC Fixed Effects	included		included	
Store Fixed Effects	included		included	
Week Fixed Effects	included		included	
Store-specific Trends	included		included	
Adjusted R-squared	0.75		0.74	
1st-Stage F statistic			34.70	

Notes: ** and *** denote statistical significance at 5% and 1%, respectively. D_s: treatment store dummy variable. D_t: treatment period dummy variable.

Table 7: Robustness Check: Using Only One Control Store

	OLS		2SLS	
	Estimate	Std. Err.	Estimate	Std. Err.
$D_s * D_t$	-0.1421	0.11	-0.2393	0.34
Log of Unit Price	-2.3343***	0.17	-1.9085	3.11
$D_s * D_t$* Score	0.0026***	0.0011	0.0042***	0.0012
UPC Fixed Effects	included		included	
Store Fixed Effects	included		included	
Week Fixed Effects	included		included	
Store-specific Trends	included		included	
Adjusted R-squared	0.80		0.81	
1st-Stage F statistic			3.91	

Notes: ** and *** denote statistical significance at 5% and 1%, respectively. D_s: treatment store dummy variable. D_t: treatment period dummy variable.

Table 8a: Instrumental Variable Estimation (2SLS) Results from the Placebo Store 1

	Specification (1)		Specification (2)	
	Estimate	Std. Err.	Estimate	Std. Err.
$D_s * D_t$	0.3052	0.24	0.1722	0.26
Log of Unit Price	-3.1850***	0.80	-3.1861***	0.80
$D_s * D_t$* Score			0.0027	0.0026
UPC Fixed Effects	included		included	
Store Fixed Effects	included		included	
Week Fixed Effects	included		included	
Store-specific Trends	included		included	
Adjusted R-squared	0.71		0.71	
1st-Stage F statistic	34.41		34.41	

Notes: *, ** and *** denote statistical significance at 10%, 5% and 1%, respectively. D_s: treatment store dummy variable. D_t: treatment period dummy variable.

Table 8b: Instrumental Variable Estimation (2SLS) Results from the Placebo Store 2

	Specification (1)		Specification (2)	
	Estimate	Std. Err.	Estimate	Std. Err.
$D_s * D_t$	0.0958	0.14	-0.0230	0.18
Log of Unit Price	-3.2176***	0.83	-3.2584***	0.83
$D_s * D_t$* Score			0.0024	0.0015
UPC Fixed Effects	included		included	
Store Fixed Effects	included		included	
Week Fixed Effects	included		included	
Store-specific Trends	included		included	
Adjusted R-squared	0.71		0.71	
1st-Stage F statistic	33.31		34.25	

Notes: *, ** and *** denote statistical significance at 10%, 5% and 1%, respectively. D_s: treatment store dummy variable. D_t: treatment period dummy variable.

Table 8c: Instrumental Variable Estimation (2SLS) Results from the Placebo Store 3

	Specification (1)		Specification (2)	
	Estimate	Std. Err.	Estimate	Std. Err.
$D_s * D_t$	-0.2669	0.21	-0.2206	0.22
Log of Unit Price	-3.2142***	0.82	-3.1985***	0.81
$D_s * D_t$* Score			-0.0009	0.0014
UPC Fixed Effects	included		included	
Store Fixed Effects	included		included	
Week Fixed Effects	included		included	
Store-specific Trends	included		included	
Adjusted R-squared	0.71		0.71	
1st-Stage F statistic	33.50		33.73	

Notes: *, ** and *** denote statistical significance at 10%, 5% and 1%, respectively. D_s: treatment store dummy variable. D_t: treatment period dummy variable.

Table 8d: Instrumental Variable Estimation (2SLS) Results from the Placebo Store 4

	Specification (1)		Specification (2)	
	Estimate	Std. Err.	Estimate	Std. Err.
$D_s * D_t$	-0.2806	0.18	-0.4809**	0.21
Log of Unit Price	-3.2082***	0.82	-3.2582***	0.84
$D_s * D_t$* Score			0.0038***	0.0013
UPC Fixed Effects	included		included	
Store Fixed Effects	included		included	
Week Fixed Effects	included		included	
Store-specific Trends	included		included	
Adjusted R-squared	0.71		0.71	
1st-Stage F statistic	33.59		33.43	

Notes: *, ** and *** denote statistical significance at 10%, 5% and 1%, respectively. D_s: treatment store dummy variable. D_t: treatment period dummy variable.

Table 8e: Instrumental Variable Estimation (2SLS) Results from the Placebo Store 5

	Specification (1)		Specification (2)	
	Estimate	Std. Err.	Estimate	Std. Err.
$D_s * D_t$	0.2184	0.15	0.3742***	0.15
Log of Unit Price	-3.2278***	0.83	-3.1767***	0.80
$D_s * D_t$* Score			-0.0030**	0.0014
UPC Fixed Effects	included		included	
Store Fixed Effects	included		included	
Week Fixed Effects	included		included	
Store-specific Trends	included		included	
Adjusted R-squared	0.71		0.71	
1st-Stage F statistic	33.21		34.41	

Notes: *, ** and *** denote statistical significance at 10%, 5% and 1%, respectively. D_s: treatment store dummy variable. D_t: treatment period dummy variable.

Table 9: Robustness Check: Only Control Stores 3 and 4 in the Control Group (2SLS)

	Specification (1)		Specification (2)	
	Estimate	Std. Err.	Estimate	Std. Err.
$D_s * D_t$	0.2245	0.19	0.1125	0.22
Log of Unit Price	-2.6479**	1.28	-2.7522**	1.28
$D_s * D_t$* Score			0.0025*	0.0013
UPC Fixed Effects	included		included	
Store Fixed Effects	included		included	
Week Fixed Effects	included		included	
Store-specific Trends	included		included	
Adjusted R-squared	0.75		0.75	
1st-Stage F statistic	22.55		23.69	

Notes: *, ** and *** denote statistical significance at 10%, 5% and 1%, respectively. D_s: treatment store dummy variable. D_t: treatment period dummy variable.

Table 10: Instrumental Variable Estimation (2SLS) Results with Additional Controls for Average Household Characteristics

	Specification (1)		Specification (2)	
	Estimate	Std. Err.	Estimate	Std. Err.
$D_s * D_t$	0.2307	0.15	0.1206	0.17
Log of Unit Price	-2.9696***	0.91	-3.0176***	0.91
$D_s * D_t$* Score			0.0023**	0.0011
Average Income	-1.01e-5	1.01e-5	-1.01e-5	1.02e-5
Average Household Size	0.6489	0.62	-0.6505	0.63
Average # of Children	0.2672	0.42	0.2595	0.43
% of Households Headed by A Person with College Degree	-0.0550	0.58	-0.0641	0.60
% of Households with Head Less than 35 years old	-0.3366	2.10	-0.2767	2.15
% of Households with Head Aged between 35 and 64	0.1282	0.83	0.1419	0.85
UPC Fixed Effects	included		Included	
Store Fixed Effects	included		included	
Week Fixed Effects	included		included	
Store-specific Trends	included		included	
Adjusted R-squared	0.75		0.75	
1st-Stage F statistic	33.66		34.26	

Notes: *, ** and *** denote statistical significance at 10%, 5% and 1%, respectively. D_s: treatment store dummy variable. D_t: treatment period dummy variable.

Figure 1: Examples of Price Tags with NuVal Scores

Note: rating from 1 (the least healthy) to 100 (the healthiest).

Figure 2a: Times Series Plots of Residuals for Observations from the Control Stores

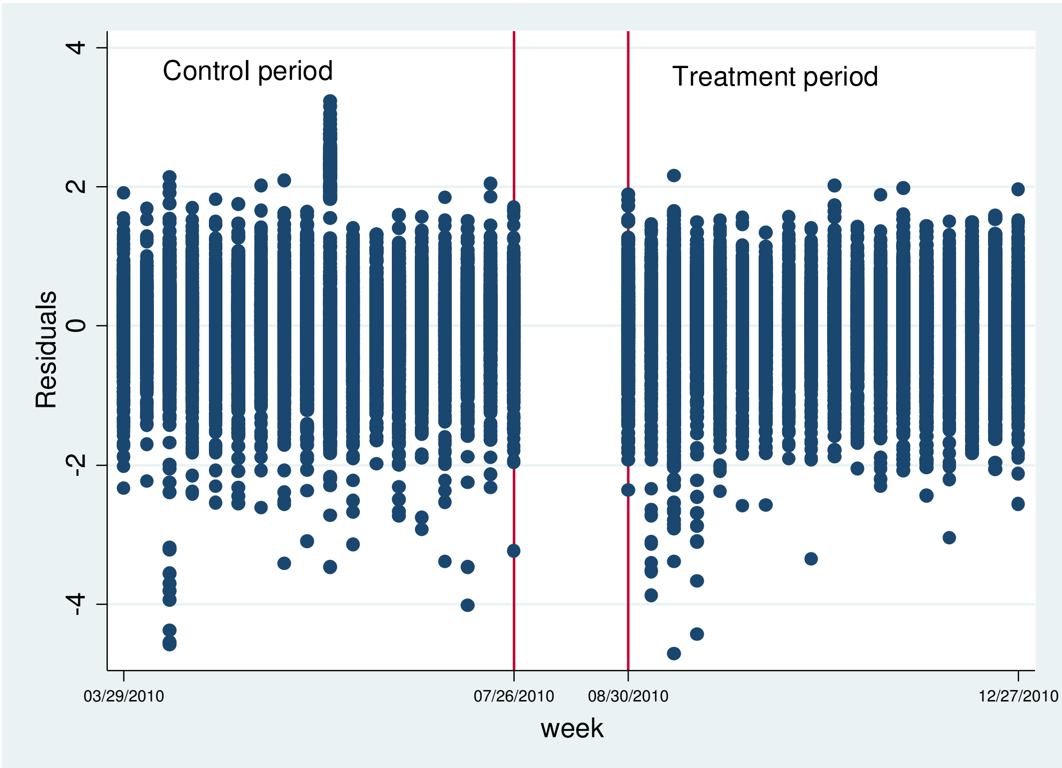

Figure 2b: Times Series Plots of Residuals for Observations from the Treatment Store