Identifying the Effect of Shelf Nutrition Labels on Yogurt Sales Using a Natural Experiment

Eric A Finkelstein
Duke-NUS Medical School
eric.finkelstein@duke-nus.edu.sg

Wenying Li
University of Georgia
wenying.li25@uga.edu

Chen Zhen
University of Georgia
czhen@uga.edu

Selected Paper prepared for presentation at the 2017 Agricultural & Applied Economics Association Annual Meeting, Chicago, Illinois, July 30-August 1

Acknowledgement and disclaimer: We thank NuVal LLC and an anonymous retailer for providing data for this research. The views of expressed here are those of the authors. Results are preliminary and subject to change in future revisions. Please do not quote without permission from the authors.

Abstract

The NuVal® shelf nutrition label rates the nutritional quality of foods on a scale of 1 (worst) to 100 (best) based on a proprietary nutrient profiling system. In 2014, NuVal updated their nutrient profiling system. We used this natural experiment to quantify the extent to which a change in the NuVal score changes consumer purchases. We focus on yogurts as this is a category that has a large range of higher and lower scores. Results reveal that a one-point increase in NuVal score increases sales by 0.36% at the median but less so for less popular yogurts. The results are consistent with the literature suggesting that shelf nutrition labels using a nutrient profiling system such as NuVal would be expected to improve food purchasing patterns and may improve health outcomes.

1. Introduction

Obesity and rates of chronic diseases remain at unhealthy levels in the US and worldwide. With unhealthy food consumption being a primary culprit, governments and industry are increasingly looking to greater use of front of package (FOP) or shelf labels on processed and packaged foods and beverages as a way to nudge consumers toward making healthier food purchases. Incorporating additional FOP or shelf nutrition labels represents an appropriate target for consideration given that the Nutrition Facts label, which has appeared on the side or rear of products for over 20 years, appears to have had little positive impact in stemming rising rates of obesity and chronic disease. This is despite the fact that more than three-fourths of energy purchased by US households comes from moderately (15.9%) and highly processed (61.0%) foods and beverages and more so among lower income and minority groups at greatest risk for obesity (Poti et al., 2016; Poti et al., 2015; Eicher-Miller et al., 2012; and Ford et al., 2014).

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 measure that goes on the front of packages (or shelf tag) and provides a clear ranking of the healthfulness of the labeled product (IOM, 2012). This recommendation is based on research which suggests that for an average person who makes over 200 daily food decisions (Wansink & Sobal, 2007) reviewing and processing all of the information contained in the Nutrition Facts label and similar multi-dimensional food labeling systems, such as the Facts Up Front label, is challenging and minimizes the effectiveness of these labels. The committee also argued that labeling only some products as healthy, as is the case with Walmart's Great For You label, is also unlikely to be effective because research has shown that consumers do not deduce the lower nutritional quality of unlabeled products by what is contained on labeled products (Mathios, 2000). Other strategies, such as the UK's stoplight system which codes foods as green, red, or yellow or the Guiding Stars four-point scoring system encourage individuals to switch across color-coded or star categories toward healthier products (Sonnenberg et al., 2013, Sutherland et al., 2010; Rahkovsky et al., 2013; Cawley et al., 2015), however, these approaches provide no signal to consumers who may wish to switch, on the margin, to a healthier product within the food category (e.g., within red foods). This is problematic given that in certain food categories, such as sugar sweetened beverages, the vast majority of foods are coded as red. For this reason, the IOM recommendation was that all foods should be ranked and labeled with an easy to interpret score that conveys the healthiness of the product.

One nutrition label used in US grocery stores that is consistent with the IOM recommendation is the NuVal® shelf nutrition label. NuVal scores foods on a scale of 1 (worst) to 100 (best) based on >30 micro- and macronutrient properties of the food. The score is driven by the Overall Nutritional Quality Index (ONQI®), a proprietary nutrient profiling algorithm that was designed to improve dietary patterns and establish weighting coefficients based on known associations between nutrients and health outcomes. In linear regression analysis of the NHANES 2003-2006 populations (n = 15,900), NuVal 1.0 scores were highly correlated with the Healthy Eating Index 2005 ($P < 0.0001$) (Katz et al., 2010). Consumption of foods that have a higher ONQI score has also been shown to be associated with modestly lower risk of chronic disease and all-cause mortality (Chiuve et al., 2011).

Recent research suggests that putting the NuVal scores on store shelf tags effectively improves food purchasing patterns. Nikolova and Inman (2015) analyzed household purchase data collected through a retail chain’s loyalty card program and from a control group of households who shopped at non-NuVal retailers. They found that the share of healthier products purchased increased at the chain following rollout of NuVal. Zhen and Zheng (2017) estimated a difference-in-differences model of yogurt demand using store-level data from one NuVal store and five non-NuVal stores in a city in the Midwest. They found that a one-point increase in NuVal score increased yogurt demand by 0.3%. Although these studies are informative, the results need to be interpreted with caution given the non-random nature of the designs and potential endogeneity as more health conscious consumers may self-select into stores that offer the NuVal information.

In this study, we take advantage of a natural experiment that occurred in 2014 when NuVal’s licensing company, NuVal LLC, updated its nutrient profiling algorithm to reflect the latest US Dietary Guidelines and scientific literature. The updated algorithm, NuVal 2.0, considers the lower sodium recommendation of 2300 mg instead of 2400 mg. It also provides greater differentiation for higher quality food products. By way of example, NuVal scores increased for foods high in biological protein given supporting evidence relating it to satiety and weight outcomes. The algorithm also adjusted saturated fat to remove the influence of stearic acid which has been shown to have a benign effect on cholesterol. NuVal 2.0 does a better job of differentiating between naturally occurring and added ingredients and not adjusting scores based on trace amounts of ingredients that would not be expected to influence health outcomes. For example, unlike the earlier version, NuVal 2.0 does not penalize naturally occurring trans-fats in sources like olive oil nor does it reward foods that have trace amount of omega-3. Added ingredients such as artificial sweeteners qualify a product to be a processed food and no longer retains the same benefits as a pure product such as plain yogurt.

The updated algorithm lowered NuVal scores on some products while raising scores on others. This exogenous change allows us to evaluate the causal impact of the NuVal shelf nutrition label on purchases. We again focus on yogurts because they have a large range of higher and lower NuVal scores. We compare changes in the sale of yogurt products at a supermarket chain before and after the revision.

2. Data

Retail scanner data on weekly yogurt sales between January 1^{st}, 2013 and Aug 31^{st}, 2015 (138 weeks) were provided by a regional grocery chain consisting of 40 stores. This chain adopted NuVal labels in August 2010. NuVal LLC, provided the NuVal 1.0 (before update) and NuVal 2.0 (after update) scores for the 191 unique yogurt products, defined by having a unique Universal Product Code (UPC), available for purchase at some point during the period of analysis. For each UPC, we have information on weekly chain-level dollar and unit sales and package size (in ounces). The retailer switched from NuVal 1.0 to NuVal 2.0 in August 2014. This yields 82 weeks when NuVal 1.0 was in effect and (from January 1st, 2013 to August 1st, 2014) and 56 NuVal 2.0 weeks (from August 1st, 2014 to August 31st, 2015).

Table 1 summarizes the data during each of the two periods. During NuVal 1.0, NuVal scores for the 191 UPCs range from 23 to 100, with an average score of 48.6. After the switch to NuVal 2.0, 75% of the 191 yogurt products experienced a decrease in NuVal score, 13% experienced an increase, and 12% remained unchanged. Scores ranged from 22 to 100, with an average score of 36.8, 11.8 points lower than the mean NuVal 1.0 score.

3. Methods

To explore whether yogurt sales decreased as a result of the lower average scores in NuVal 2.0, we explored weekly unit and volume sales before and after the switch, with the expectation that overall sales would decrease given the lower average NuVal score after the change. We then divided the sample into three subgroups based on whether the NuVal score for each UPC decreased, stayed the same, or increased after the change to explore the hypothesis that those products whose score decreased would show a greater reduction in sales than those products whose score stayed the same or increased.

To further explore the effects of a change in NuVal score on purchases, we employ a quantile regression framework. Unlike standard linear regressions that provide an average effect estimate, quantile regressions are able to describe the entire conditional distribution of the dependent variable. This allows us to examine whether there are heterogeneous responses to NuVal score changes among products of different market shares.

For our data, there are also technical arguments for preferring quantile regressions over linear regression. Quantile regression minimizes the sum of the absolute residuals at each quantile rather than the sum of the squared residuals in ordinary least squares estimation. While the optimal properties of standard linear regression estimators are not robust to modest departures from normality, quantile regressions are characteristically robust to outliers and heavy-tailed distributions (Buchinsky 1994), which, as shown in Figure 1, is the case with our data.

We estimate quantile regressions of log chain-level volume sales on log unit price and the interaction of a dummy variable for the post-NuVal 2.0 period with the score difference between NuVal 2.0 and NuVal 1.0. To control for product heterogeneity and time fixed effects, we include week- and UPC-specific dummy variables as covariates. The estimating equation is specified as

$$(1) \quad y_{i,t} = \alpha + \sum_{i=1}^{190} \beta_{1,i} D_{1,i} + \sum_{t=1}^{137} \beta_{2,t} D_{2,t} + \beta_3 D_{3,t} \times (S_{i2} - S_{i1}) + \beta_4 P_{i,t} + U_{i,t}$$

where $y_{i,t}$ is the logarithm of chain-level volume sales for UPC i in week t. Three dummies variables are included in the model. $D_{1,i}$ is the dummy variable for UPC i, $D_{2,t}$ is the week dummy, and $D_{3,t}$ is the dummy variable equal to 1 if t is after NuVal revision and 0 otherwise. S_{i1} and S_{i2} are NuVal 1.0 and 2.0 score for UPC i, respectively. $P_{i,t}$ is logarithm of unit price. $U_{i,t}$ represents the error term. α is the intercept. The β's are coefficients. β_3 is the coefficient of interest.

The quantile regression coefficients can be interpreted as the partial derivative of the conditional quantile of $y_{i,t}$ with respect to the corresponding explanatory variables. For example, if β_3 is estimated to be positive and statistically significant, then it will confirm our hypothesis that an increase in NuVal score increases sales. The quantile regression also allows us to test the hypothesis that the impact of a change in NuVal score will be greater (smaller) for foods whose market shares were larger (smaller). This results because a change in score is more likely to be noticed on more popular (i.e., greater volume) products. The sample used to estimate equation (1) covers 191 UPCs from 82 NuVal 1.0 weeks and 56 NuVal 2.0 weeks with a total of 23,758 observations. We present results at the 0.1, 0.25, 0.50, 0.75, and 0.9 quantiles.

4. Results

Based on table 1, consistent with the hypothesis that a reduction in scores would reduce demand, on average, about 336 units were sold per week per UPC in the NuVal 1.0 period, whereas this figured dropped to 289 units sold per week in the NuVal 2.0 period. This reduction occurred despite average yogurt prices ($0.16/ounce) remaining stable over the two periods. Table 2 summarizes the sales data separately for the three yogurt groups that experienced an increase, decrease, and no change in score. As expected, weekly volume dropped by 261, 247, and 206 ounces per UPC for products whose NuVal scores decreased, stayed the same, or increased, respectively. Average weekly unit sales also decreased more for the group whose NuVal score decreased than for the group whose score increased (49.6 vs. 32.4 unit reduction per UPC). However, unit reduction was least for the products whose score remained unchanged, at 22.1 units per UPC.

Quantile regression results are presented in Table 3. Consistent with our hypothesis, at the 0.10 quantile (i.e., the lower end) of the distribution, the coefficient estimate of β_3 is smallest at 0.0021, suggesting that the influence of the score change is least influential for the least popular products. The marginal effect increase to above 0.0035 and remains statistically significant at the median, the 0.75 and 0.95 quantiles, where the coefficient is largest. For example, at the 0.9 quantile, β_3 is estimated to be 0.0037. As such, if a product's NuVal score decreases by 11.8 points, which is the average score change for yogurt products in our sample, then weekly sales volume would be expected to decrease by 4.4%.

5. Discussion

This study took advantage of the natural experiment that occurred when NuVal underwent a change in the algorithm used to score nutritional quality of foods. This natural experiment allowed us to identify the causal effect of NuVal scores on yogurt sales, a product with a large range of more and less healthy alternatives. At the median, a ten-point increase in NuVal score is estimated to increase sales of the corresponding yogurt product by 3.6%. This is remarkably close to the 3.0% estimate for the yogurt category reported in Zhen and Zheng (2017) even though that study used a different methodology, data from a different retailer, and only had data during the NuVal 1.0 period

Unlike that study, we also showed that there is a moderate degree of heterogeneity in the impact of NuVal scores on consumer demand, with the greater influence occurring for foods with greater volume. This should be expected given that consumers are more inclined to look at the NuVal score on foods they have previously purchased. Given that some of the most commonly purchased foods are of poor nutritional quality, this suggests that adding the NuVal score to these foods would have a significant (negative) influence on demand and would be expected to positively influence health outcomes of shoppers.

This analysis has several limitations. This includes the limitation to one food category (yogurt), the inability to control for potential price endogeneity that may result from demand changes due to labeling, and the inability to control for changes in demand for non-yogurt products as a result of the move to NuVal 2.0. We also are only able to estimate the influence of a change in NuVal score on purchases, but are unable to document the extent to which those changes influence health outcomes.

Despite these limitations, the results are consistent with the theory and empirical literature suggesting that front of package labeling using a scoring system such as NuVal would be expected to improve food purchasing patterns and may improve health outcomes. The latter should be an area of future research.

References

Buchinsky, M., 1994. Changes in the U.S. wage structure 1963-1987: Application of quantile regression. *Econometrica, 62*, 405–458.

Cawley, J., Sweeney, M. J., Sobal, J., Just, D. R., Kaiser, H. M., Schulze, W. D., Wethington, E., & Wansink, B. (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., Sampson, L., & Willett, W. C. (2011). The association between a nutritional quality index and risk of chronic disease. *American Journal of Preventive Medicine, 40*, 505–513.

Eicher-Miller, H. A., Fulgoni, V. L., & Keast, D. R. (2012). Contributions of processed foods to dietary intake in the US from 2003-2008: A report of the Food and Nutrition Science Solutions Joint Task Force of the Academy of Nutrition and Dietetics, American Society for Nutrition, Institute of Food Technologists, and International Food Information Council. *The Journal of Nutrition*, 142, 2065S–2072S.

Ford, C.N., S.W. Ng, and B.M. Popkin. (2014). Are food and beverage purchases in households with preschoolers changing?: A longitudinal analysis from 2000 to 2011. *American Journal of Preventive Medicine, 47*, 275–282.

Institute of Medicine. (2012). *Front-of-package nutrition rating systems and symbols: Promoting healthier choices*. Washington, DC: The National Academies Press.

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

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*, 651–678.

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.

Poti, J. M., Mendez, M. A., Ng, S. W., & Popkin, B. M. (2015). Is the degree of food processing and convenience linked with the nutritional quality of foods purchased by US households? *American Journal of Clinical Nutrition, 101*, 1251–1262.

Poti, J. M., Mendez, M. A., Ng, S. W., & Popkin, B. M. (2016). Highly processed and ready-to-eat packaged food and beverage purchases differ by race/ethnicity among US households. *The Journal of Nutrition*, 146, 1722–1730.

Rahkovsky, I., Lin, B. H., Lin, C. T., & Lee, J. Y. (2013). Effects of the Guiding Stars Program on purchases of ready-to-eat cereals with different nutritional attributes. *Food Policy, 43*, 100–107.

Sonnenberg, L., Gelsomin, E., Levy, D. E., Riis, J., Barraclough, S., & Thorndike, A. N. (2013). A traffic light food labeling intervention increases consumer awareness of health and healthy choices at the point-of-purchase. *Preventive Medicine*, *57*, 253–257.

Sutherland, L. A., Kaley, L. A., & Fischer, L. (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.

Wansink, B., & Sobal, J. (2007). Mindless eating: The 200 daily food decisions we overlook. *Environment and Behavior*, *39*, 106–123.

Zhen, C., & Zheng, X. (2017). The impact of NuVal shelf nutrition labels on food purchase. Working Paper.

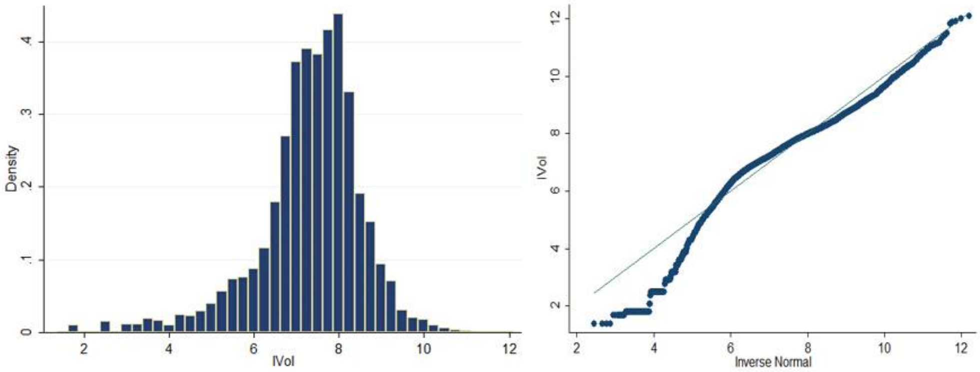

Figure 1. The histogram and Quantile-Quantile Plot of the dependent variable

Table 1. Descriptive statistics by NuVal version

	Mean			Std. Dev.		Min		Max	
Variables	**NuVal 1.0**	**NuVal 2.0**	**Diff.**	**NuVal 1.0**	**NuVal 2.0**	**NuVal 1.0**	**NuVal 2.0**	**NuVal 1.0**	**NuVal 2.0**
NuVal score	48.55	36.75	-11.80	25.39	16.72	23.00	22.00	100.00	100.00
Weekly units sold per UPC	335.88	288.71	-47.17	324.38	300.14	0.07	0.04	1779.26	1360.20
$/unit	2.14	2.36	0.22	1.46	1.69	0.41	0.39	6.55	7.22
$/ounce	0.16	0.16	0.00	0.08	0.08	0.04	0.07	0.92	0.81
Weekly volume (ounce) sold per UPC	2571.39	2280.01	-291.38	2039.74	2121.34	0.71	0.21	10675.54	10261.71

Table 2. Descriptive statistics by direction of NuVal score change

	Mean			Std. Dev.		Min		Max	
Increased Group									
Variables	**NuVal 1.0**	**NuVal 2.0**	**Diff.**	**NuVal 1.0**	**NuVal 2.0**	**NuVal 1.0**	**NuVal 2.0**	**NuVal 1.0**	**NuVal 2.0**
Weekly volume sold per UPC	2372.74	2166.61	-206.14	3115.63	3574.59	0.00	0.00	36490.50	34487.10
Weekly units sold per UPC	330.31	297.88	-32.43	553.81	634.36	0.00	0.00	6885.00	6507.00
$/unit	2.12	2.15	0.03	1.64	1.69	0.50	0.50	6.92	6.58
$/ounce	0.20	0.20	0.00	0.06	0.05	0.08	0.08	0.29	0.27
NuVal score	47.08	50.21	3.13	29.78	29.79	23.00	23.00	96.00	96.00
Decreased Group									
	Mean			**Std. Dev.**		**Min**		**Max**	
Variables	**NuVal 1.0**	**NuVal 2.0**	**Diff.**	**NuVal 1.0**	**NuVal 2.0**	**NuVal 1.0**	**NuVal 2.0**	**NuVal 1.0**	**NuVal 2.0**
Weekly volume sold per UPC	2866.99	2605.68	-261.30	5354.22	4194.18	0.00	0.00	179772.00	66804.00
Weekly units sold per UPC	378.17	328.53	-49.64	875.83	663.73	0.00	0.00	29962.00	11134.00
$/unit	1.72	1.74	0.02	1.47	1.51	0.27	0.25	6.91	6.58
$/ounce	0.17	0.17	0.00	0.13	0.14	0.04	0.04	1.00	1.00
NuVal score	52.40	35.49	-16.91	24.52	10.65	24.00	22.00	99.00	81.00
Unchanged Group									
	Mean			**Std. Dev.**		**Min**		**Max**	
Variables	**NuVal 1.0**	**NuVal 2.0**	**Diff.**	**NuVal 1.0**	**NuVal 2.0**	**NuVal 1.0**	**NuVal 2.0**	**NuVal 1.0**	**NuVal 2.0**
Weekly volume sold per UPC	1834.48	1587.39	-247.09	2995.99	2838.17	0.00	0.00	46523.40	39230.60
Weekly units sold per UPC	270.85	248.74	-22.11	569.05	549.58	0.00	0.00	8778.00	7402.00
$/unit	2.14	2.36	0.22	1.46	1.69	0.41	0.39	6.55	7.22
$/ounce	0.24	0.25	0.01	0.18	0.19	0.04	0.06	0.74	0.78
NuVal score	30.09	30.09	0.00	15.27	15.27	23.00	23.00	100.00	100.00

Table 3. Quantile regression results using weekly volume sales as the dependent variable

	Quantile Regression (%)				
Regressors	**10**	**25**	**50**	**75**	**90**
$D_{3,t} \times (S_2 - S_1)$	0.0021**	0.0031**	0.0036**	0.0035**	0.0037**
	(0.0004)	(0.0004)	(0.0004)	(0.0003)	(0.0004)
$P_{i,t}$	-2.7292**	-3.0451**	-3.4002**	-3.8415**	-4.4230**
	(0.0790)	(0.0464)	(0.0349)	(0.0288)	(0.0413)
$D_{3,t}$	-0.2600**	-0.0521	0.0971**	0.2329**	0.1784**
	(0.0583)	0.0897	(0.0472)	(0.0306)	(0.0356)
$[Pseudo-]R^2$	0.6750	0.6198	0.5993	0.6166	0.6419

Notes: clustered sandwich standard errors (bootstrapped for quantiles) are given in parentheses. ** denotes statistical significance at 5%; Other regressors include dummy variable indicating single UPC and the week dummies.