Do Informative Promotional Campaigns for On-Shelf Nutritional Labeling Work?

Bryan Bollinger *
Duke University

David Hammond
University of Waterloo

Erin Hobin
Public Health Ontario

Eli Liebman
University of Georgia

Jocelyn Sacco
Public Health Ontario †

June 4, 2018

Abstract

Front-of-package and on-shelf nutrition labelling systems in supermarkets, such as Guiding Stars®, have been shown to lead to only modest increases in the purchase of more nutritious foods. Because only a small fraction of consumers report using these labels, informative promotional campaigns may increase their use. We study a large-scale, national promotion campaign for Guiding Stars conducted by a national grocery retailer in Canada who implemented the program. Using both transaction and survey data, we found only a small increase in the purchase of higher star-rated foods during the campaign, and 40-50% of the effect disappeared as the campaign concluded. To explain the limited response, surveys taken outside of stores before and after the campaign show that, although awareness and understanding of the nutrition labelling system increased marginally after the campaign, self-reported use did not. This casts doubt whether promotional campaigns can actually increase the effectiveness of on-shelf labelling.

Keywords: health, nutritional labelling, education, marketing
JEL classification codes: D12, D18, D83, I18, L81, M31

---

*Corresponding author. Duke University Fuqua School of Business, 100 Fuqua Drive, Durham, NC 27278. Email: bryan.bollinger@duke.edu. Phone: 919-660-7766.
† The authors would like to thank the national Canadian grocery firm for their data support. We also acknowledge the excellent contributions by Lana Vanderlee, Fei Zuo, Laura Rosella, Heather Manson, and Mary L'Abbe. This work was supported by the Canadian Institutes for Health Research (CIHR). Additional support was provided by the Propel Centre for Population Health Impact, Public Health Ontario, and a CIHR Applied Public Health Chair (to DH). All errors are solely the responsibility of the authors.

1 Introduction

Unhealthy diets are a leading risk factor for chronic diseases and premature death in both Canada and the United States.[1] To help consumers make more informed food purchasing decisions, Nutrition Fact Panels on most packaged foods have been mandatory in the United States since 1994 and in Canada since 2005. While evidence suggests that demand is sensitive to nutrition information, there is also evidence that consumers have difficulty interpreting and applying the information in Nutrition Facts Panels.[2] Because of this, there have been a number of attempts to simplify the manner in which nutritional information is displayed to consumers. Hawley et al. (2013) conclude in their review paper that while traffic light symbols consistently helped consumers identify healthier products, consumers still struggle to notice, comprehend, and use even simplified nutrition labelling.[3] These findings suggest that there may be benefits of large-scale campaigns that increase awareness, understanding, and use of nutrition labels.

To our knowledge, there are no studies which examine informative promotional campaigns specifically designed to facilitate awareness, understanding, and use of a nutrition labelling system. We address this gap by studying the effect of a large-scale, national campaign to improve awareness, comprehension and use of the Guiding Stars® on-shelf, nutrition labelling system that was implemented by a large, national Canadian grocery retailer in Canada in multiple banners. Guiding Stars uses a simple zero to three stars metric to summarize the healthiness of food products, the higher the star rating the higher the relative nutritional quality of the product. This intervention provides a simple, single-dimensional metric of the nutritional quality of all fresh and packaged foods that consumers may chose to use when making food purchases in a supermarket. It is one of the first large-scale nutrition labelling system for foods that is consistent with the

[1] Institute for Health Metrics and Evaluation. University of Washington. Global burden of disease study 2010 country profiles.

[2] Ippolito and Mathios (1990), Ippolito and Mathios (1995), Kiesel and Villas-Boas (2008), and Mathios (2000) all find evidence that demand is sensitive to nutrition information.

[3] Traffic light symbols use green lights to designate the healthiest products and red to designate the least healthy.

recommendations for well-designed labels by the US National Academies[4], with the notable exception that products with a 0-star rating carry no label. In other words, products with the lowest nutrition rating do not display any rating and are indistinguishable from products that are not rated.

While the goal of the Guiding Stars system is to simplify nutrition information and help consumers analyze tradeoffs across attributes, the effects of this intervention are generally small.[5] Hobin et al. (2017) argue that one explanation for the very modest effect is that, 6-months after implementation, fewer than 10% of consumer reported noticing and understanding the labels. The results of Hobin et al. (2017) imply that there may be scope for a national promotional campaign to improve awareness, understanding and use of the Guiding Stars system, and ultimately improving the nutritional profile of the foods consumers purchase.

The retailer launched a large-scale, national promotional campaign for the Guiding Stars program from January 9 – February 5, 2015, intended to increase awareness, understanding, and use of the Guiding Stars labels. Launching an educational and promotional campaign alongside a Front-of-Package (FOP) labelling system is consistent with the National Academies' recommendations for effective FOP labelling systems. The national promotional campaign included a 30-second television spot, radio and flyer advertisements, and enhanced in-store signage describing the Guiding Stars system and how to use the system in supermarkets. In addition, the retailer provided information through their grocery store websites to entice consumers to monitor the star ratings of their purchases over time and to compare to their fellow shoppers. In addition to serving both educational and promotional goals by providing information to help consumers understand and use the GS system, the campaign also helped in terms of corporate social responsibility (CSR) goals by associating the retailers image with healthy consumption.

We assess the response to the campaign using multiple approaches. First, we use detailed transaction data to study the effect of the national promotional campaign on foods

[4] Formerly the US Institutes of Medicine

[5] See Sutherland et al. (2010); Cawley et al. (2014); Rahkovsky et al. (2013); Hobin et al. (2017).

purchased, how this effect varies by food category, and how this effect varies by household. We estimate the effect of the national promotional campaign using a difference-in-difference approach in which we use each store's previous year as a non-concurrent control group. To attribute differences to the promotional campaign, we must assume that trends are common across years for the treated units (the stores). Similarly, we perform a cardholder analysis using each household's previous year as its non-concurrent control – now we must only assume that household trends are common across years. Given the stability of household preferences over time, this second approach replies on fewer assumptions because changes in consumer composition at stores will not affect the results. In both approaches, we find a small increase in the share of the most nutritious, three star items (with corresponding decreases in other items) during the campaign, half of which carryover to the following two months.

One explanation for the small size of the effect on purchases is the limited effect on consumer awareness and understanding. In surveys outside the supermarket banners in Ontario before and after the campaign, we find that awareness and understanding increase only marginally, and self-reported use does not change at all. Indeed, when we study heterogeneity in the household effects, we find that those households who do increase their purchase of products with more stars are those that are already purchasing higher-starred products. These findings are not encouraging for policymakers hoping to leverage informative marketing campaigns to spur healthier consumption behavior by those not already eating healthy.

The rest of the paper is organized as follows. In the next section we describe the related literature. In Section 3, we discuss the intervention in more detail and describe the data. In sections 4 and 5, we present the analysis of the store-level and household-level transactions data, respectively. In Section 6 we present survey results to test mechanisms which may drive the results. Section 7 provides a discussion and concluding remarks.

2 Background

The effectiveness of front-of-package (FOP) and on-shelf nutrition labels is of critical importance to policymakers - the US Food and Drug Administration and National Academies are both investigating their effectiveness in order to provide guidance to policymakers and the food industry (Hawley et al. 2013). It is hoped that promotional campaigns to support consumer awareness and understanding might substantially increase the use of these labels, and thus an assessment of these labelling programs in conjunction with a campaign is necessary in order to predict the true long-term effectiveness of such interventions. [6] The main benefit of the Guiding Stars label is the fact that it is a simple, at-a-glance, single index measure of the nutritional value of products. This also avoids the issue demonstrated by Schmeiser (2014) in the context of GMOs, who shows how the weight consumers place on an attribute can change with mandatory disclosure of specific attributes. The Guiding Stars system may increase the weight on nutrition, but not on one nutritional attribute to the detriment of another (except through the algorithm underpinning the Guiding Stars ratings).

Researchers have been interested in the effect of nutrition information on consumer decisions for decades. Jacoby, Chestnut and Silberman (1977) present evidence that consumers tend not to seek out nutrition information or to understand it, even though they claim they would be willing to pay for more nutrition information. Despite this, Bollinger et al. (2011) find that calorie labelling on menu boards in restaurants can lead to a reduction in calories purchased, depending on consumers' prior beliefs. Liu et al. (2012) find that rank ordering products on menus by calories can further increase the effectiveness of calorie information in reducing consumption, and that including traffic-light symbols can increase it further.

One advantage of calorie labelling is the use of a single metric. Thus, other labelling systems that aggregate information on multiple product characteristics into a single, sim-

[6] Walls et al. (2009) are pessimistic about the ability of information provision alone in affecting behavioral change, partly due to the many attributes that consumers must account for in their decision making process.

ple metric (such as Guiding Stars) have been recently studied in lab contexts to assess their effectiveness. Lang et al. (2000) study a color-coded shelf label program using exit surveys (N=361), and find that awareness of the program was 28.8% but 37% of participants reported they did not use the shelf-label system. In a quasi-experimental, repeated measures study, Levy et al. (1985) found that 31% of customers in the shelf label condition reported using the labels two-years after implementation. Understanding and use of the labeling system is of course crucial if policymakers want to affect consumer health.[7]

Very few studies explore the effect of such labels on actual food purchasing behaviors. Sacks et al. (2009) find no effect of traffic lights on their categories of study, chilled pre-packaged meals and fresh pre-packaged sandwiches. Katz et al. (2010) studied the Overall Nutritional Quality Index (ONQI) which is used in the NuVal shelf-labelling system. They find that study participants view labels positively and 80% of them (N=804) said they would use the information in making purchase decisions. Levy et al. (1985) examined the "Special Diet Alert" (SDA) program using sales data across various food categories for a 2-year period. In only half of the 16 comparisons between categories in treatment and control stores was there a significant and positive health effect of the SDA program: canned fish, mayonnaise, butter/margarine, cottage cheese, and fruit juice exhibited increased growth for lower calorie products, and soft drinks, frozen vegetables, and tomato sauce increased sales for lower sodium products. There was no effect for canned fruit, fresh milk, soft drinks, cheese, nuts and snacks, dry cereals, and crackers. Finally, Sonnenberg et al. (2013) study the use of a traffic light nutrition labelling system in a hospital cafeteria using survey and receipt data and found that traffic light labels increased consideration and purchase of healthier items. Prior to the intervention, even people who stated they always purchase healthy items would still order unhealthy ones, demonstrating that consumers may have incorrect beliefs regarding the nutritional quality of foods in the absence of such programs.

There are four studies of which we are aware that examine the effectiveness of the

[7] Hawley et al. (2013) provide a nice summary of the papers on shelf and front-of-package food labels, concluding that traffic light symbols were the most effect way of conveying information.

Guiding Stars system specifically, in the absence of an accompanying campaign that would help inform the consumer regarding their use. Using store-level transaction data from 168 stores in five US states before and after implementation of the intervention, Sutherland et al. (2010) found that 24.5% of products purchased before the intervention earned a star rating. This proportion increased to 25.0% ($p < 0.001$) and 26.0% ($p < 0.0001$) at 1- and 2-year follow-ups, with the most movement in sales from 0- to 1-star products. Cawley et al. (2014) examined changes in sales following Guiding Stars implementation in 150 stores of a single supermarket chain in the US. They found that approximately 1-year after implementation, sales of foods with 0-stars declined by 8.31% ($p = 0.004$), while sales of foods with ratings of 1-, 2-, and 3-stars did not change ($p = 0.21$). The share of starred food items purchased increased by 1.39% but overall sales declined.[8] Hobin et al. (2017) studied the effect of the Guiding Stars system across all categories, assessing the impact of the program on nutritional quality of food purchases (as well as revenues) and found that the Guiding Stars system led to an increase in the mean star level of products purchased by 1.4%. Similar to this study, we augment our transaction level data with survey data in order to better understand consumers' awareness and understanding of the program.

National health promotional campaigns have been shown to sometimes be effective. Davis et al. (2008) summarizes the findings for anti-smoking campaigns, some of which have been very effective, and in the health domain, Barragan et al. (2014) showed that a campaign illustrating the number of sugar packs contained in sugary drinks led to increased intention to decrease sugary beverage consumption and an increase in awareness of health messages. In contrast, in the domain of sexually transmitted diseases, many campaigns have been shown to be ineffective (Stekler et al. 2013; Ross et al. 2016). Noar

[8]The issue with booth of these studies is the lack of controls for seasonality and other macro trends affecting food purchases. To overcome this issue, Rahkovsky et al. (2013) compared sales of breakfast cereals in supermarkets with and without Guiding Stars (using propensity score matching) with the use of Nielson Scantrack data, 1-year before, and 20-months after implementation. They found that the market share of cereals with 1-, 2-, and 3-stars significantly increased by 1.15%, 0.89%, and 0.54%, respectively, while market share of 0-star cereals decreased by 2.58% ($p < 0.01$). The authors found that purchasing of least nutritious (0-star) cereals was higher among lower income consumers, and this association was found in stores with and without the Guiding Stars system.

(2006) provide a summary of the research in national health campaigns and find that campaigns may have moderate impact but only when carefully designed. The extent of the impact depends heavily on the desired behavioral outcome.

There is no study of which we are aware to look at campaigns specifically designed to facilitate awareness and understanding of a nutrition labelling program. This is surprising given both the importance of consumers' awareness and understanding of nutrition labelling, as well as the increased use of multiple policy tools in concert to obtain desired firm of policy maker outcomes. Further research is needed to examine the design of promotional campaigns supporting nutritional labelling systems to determine the most effective elements and messages.

3 Campaign Details and Data

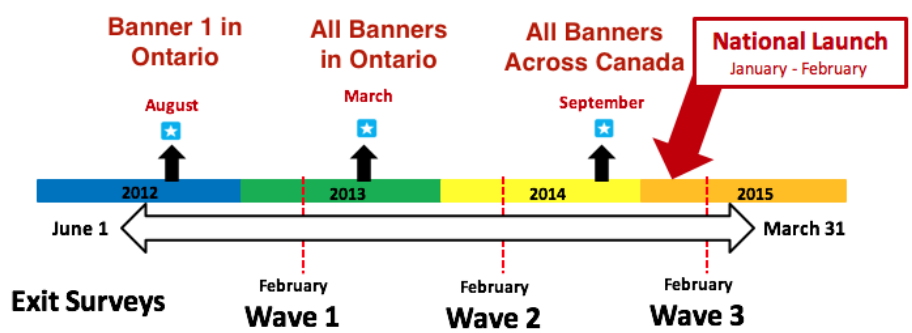


Figure 1: Labelling and Campaign Timeline

This figure presents the timing of the labeling, education campaign, and exit surveys. The labeling was implemented by the first banner in Ontario in August 2012. The other banners introduced Guiding Stars in March of 2013. The national campaign we study in this paper occurred in January and February of 2015. We administered surveys in February 2013, 2014 and 2015.

The national campaign was run from January 9, 2015 to February 5, 2015. Figure 1 shows the exact timing of the Guiding Stars implementation and education campaign. In addition, we label the times at which we collect survey data on customer awareness, understanding, and use of the Guiding Stars labels. Because implementation of the Guiding Stars program occurred nationally before the educational campaign, we don't have a contemporaneous control group. We therefore use the year April 1, 2013 - March 31, 2014 as a control for the year April 1, 2014 -March 31, 2015. This allows us to account for seasonality and other time-varying factors which may not be controlled for by week-of-year dummy variables. Because we need to use the prior year as the control group, we confine our analysis to the first banner in Ontario, which had implemented Guiding Stars system prior to the start of our control period, April 1, 2013.

The campaign itself consisted of 30-second English and French ads on national television, flyers in weekly ads, magazine ads, in-store demonstrations, and permanent in-store signage. The retailer worked through their PR team to engage national vendors, and feature print articles appeared in Chatelaine, Canadian Health & Lifestyle and Best Health. Of note, the national campaign was not followed by any large, sustained marketing endeavors which might have helped to maintain or increase awareness and use of the program.

The primary data for this study includes all transactions for supermarkets in the first banner in Ontario, Canada between April 1, 2013 and March 31, 2015. Each line in the raw data corresponds to a particular UPC on a given shopping trip or transaction, where a shopping trip is defined by a receipt. Each line contains information on the price, quantity (by weight or number of packages), the UPC code, date and time of purchase, and a store identifier. In addition, a subset of the data has a unique household identifier related to their loyalty card that is affiliated with the supermarket banners. The dataset includes a total of 77.2 million transactions, 9.7 million of which are by shoppers who use the PC credit card which allow us to track their purchases over time.[9] We merge the transaction data with the Guiding Stars data from the Guiding Stars Licensing Company, who, for each UPC offered by the banners determine a final star rating, as well as nutrition infor-

[9]We trim the purchase data by excluding the shopping trips in the top percentile of items, where depending on the specification items are unweighted across UPCs, by quantity marked on the receipt or by servings.

mation for each product.[10] The Guiding Stars nutrition information dataset includes more than sixty thousand unique UPCs.

For the main analysis, we aggregate from UPC-receipt level to store-day level. Within each store-day, we compute a quantity weighted average of the nutrition characteristics for each UPC within a transaction, then a quantity weighted average across transactions. Therefore, each observation in the dataset used in the analysis is a store-day, with a corresponding quantity-weighted average for that store-day for each variable.

Table 1: Summary Stats for Store-Level Analysis

	mean	sd
Items/Trans	7.88	1.85
Fat	4.24	0.32
Sat Fat	1.42	0.13
Trans Fat	0.05	0.01
Sodium	143.21	15.42
Sugar	6.55	0.44
Calories	113.65	5.23
Dietary Fiber	1.53	0.12
Protein	4.63	0.23
Omega 3 mg	51.36	7.82
Mean Stars	1.35	0.14
Share of 0 Star	0.45	0.05
Share of 1 Star	0.10	0.01
Share of 2 Star	0.09	0.01
Share of 3 Star	0.36	0.04
Num Transactions	2480.81	1038.43
Total Items	19542.80	9232.41
Observations	31121	

This table presents summary statistics for each dependent variable, where each observation is a store-day weighted average of that variable, weighted by the quantity sold. There are 31121 store-day observations. The bottom panel presents transactions per day and total number of items per day, which are not weighted by quantity sold.

Table 1 presents summary statistics for the nutrition characteristics for the entire sample period. The upper panel displays summary statistics of calories and nutrients. The

[10] A small fraction of products are not rated.

lower panel which displays summary statistics for the number of transactions and total number of items.[11] On average, there are 2480 transactions per store-day in our data. The average item purchased contains 113 calories and 4.2 grams of fat per serving, and has 1.35 stars. The average number of items per transaction is 7.88 meaning there are 19,542 items purchased per day in our data.

4 Transaction Data Analysis

4.1 Average Effects

In order to estimate the effect of the national promotional campaign we use a difference-in-differences approach in which the control group is the previous year. Figure 2 plots the average star rating per item (weighted by quantity) purchased, for the transaction data in the first graph, and for the cardholder data in the second. The treatment group is defined as the second year of our sample, April 1, 2014 to March 31, 2015, represented by the darker line, while the control group is, April 1, 2013 to March 31, 2014, represented by the thinner gray line. The solid vertical line at 13 denotes January, which in the treatment year was the month the promotional campaign ran, while the dashed vertical line denotes February which is the month the promotional campaign ended.

Figure 2 also helps to illustrate the variation we use and the identification strategy. The gray line, which represents the control group, is almost always higher than the darker line which represents the treatment group, indicating that the overall healthiness of baskets declined between the two years. The only time the second year is higher than the first year is in January, the primary month during which the campaign ran. After the campaign's conclusion, the black line again dips below the gray line, indicating some of the campaign effects immediately disappear.

It should be noted that there is considerable seasonality evident in the graphs, with a

[11] Because the quantities in the lower panel are summed across days, rather than means, these two statistics are constructed without weighting across transactions within a store day.

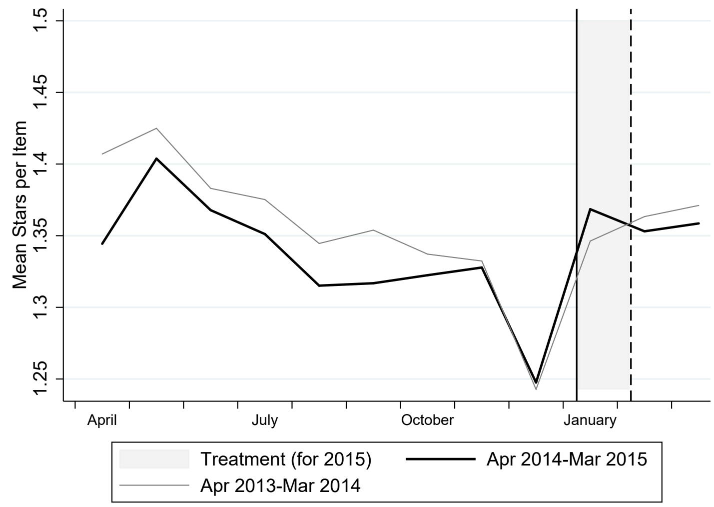
large decline in the star ratings of items purchased over the winter holidays. It also should be noted that there is discrepancy in the mean star ratings of purchased products in the holiday months, relative to the overall decline in healthiness between year one and two. In year 2, the mean star rating is almost exactly the same across the two years, despite the general decline. The fact that we see this for both the transaction and cardholder data can help rule out a consumer composition effect. Our preferred explanation is that consumers have specific holiday items that they purchase that are the same across both years and thus not subject to the same general macro trend that leads to lower star levels in year 2. Thus we believe the more relevant comparisons for the campaign and post campaign periods (January through March) are to the preceding non-holiday months. For our main analysis, we include all observations in all months. That said, if the higher star level in year 2 in November and December are due to some other trend only relevant to that year, than our campaign estimates would be overestimates of the actual effects.

To do a more formal analysis, we estimate the following difference-in-difference regression:

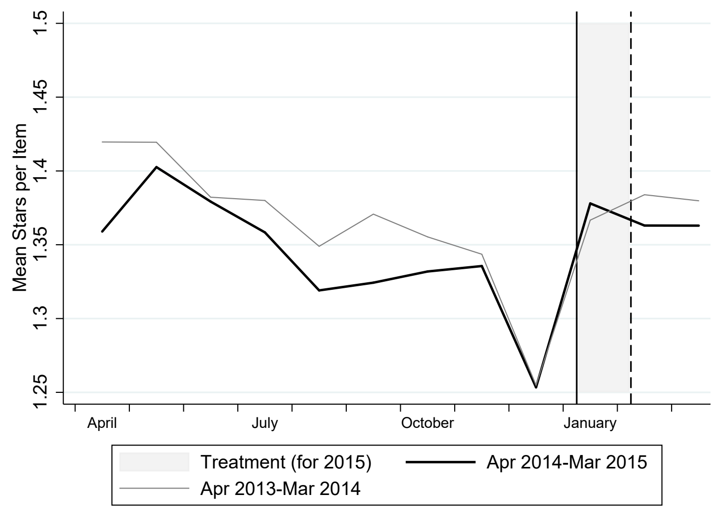
$$log(y_{jt}) = \alpha(D_t \times \zeta_T) + \beta(P_t \times \zeta_T) + \zeta_T + \gamma_{jw} + \xi_d + \delta_{n_{jt}} + \epsilon_{jt}. \quad (1)$$

We index store by j, day of year by t, week of year by w and day of week by d. Because we expect each dependent variable, indicated by y_{jtr}, to change proportionally, we use log values for our dependent variables of interest.[12] This lets us interpret the coefficients as proportional changes due to the implementation of Guiding Stars. ζ_T is a dummy variable indicating the treatment year, i.e. it is set to one on April 1, 2014.[13] This captures the first difference: differences in mean healthfulness between the treatment and control groups throughout the sample. D_t is a dummy variable set to one during the campaign month in both the treated and untreated year. Similarly, P_t is a dummy variable set to one after the campaign period ends, in both the treated and untreated year. Thus the interaction $D_t \times \zeta_T$ indicated the treatment during the campaign, and $P_t \times \zeta_T$ the treatment

[12] To prevent taking the log of zero, we add 0.01 x mean of the dependent variable.
[13] $\zeta_T = 0$ from April 1, 2013 to March 31, 2014 and $\zeta_T = 1$ from April 1, 2014 to March 31, 2015.



(a) Transaction Data



(b) Cardholder Data

Figure 2: *Mean stars, control vs. treatment year*

This figure graphs the mean start per item in the control and treatment year. The horizontal axis plots the month, starting in April 2013 or April 2014 (represented by 4), going to March 2014 or March 2015. The solid vertical line at 13 denotes January 8, which was the start of the promotional campaign ran (indicated with shading), while the dashed vertical line denotes February 5, which is the end the promotional campaign ended. The Cardholder data only includes cardholders present in both years.

post-campaign. In addition, we include and day-of-week dummy variables, ξ_d, to control for differences in the foods purchased on different days of the week. γ_{jw} are either store and week-of-year fixed effects, or store $\times$ week-of-year fixed effects depending on the specification. This captures the second difference: trends which are common to the treatment and control groups (i.e. consumers purchase more unhealthy food in December). In some specifications we include dummy variables for the average number of items in a transaction in store j on day t, n_{jt}, denoted by $\delta_{n_{jtr}}$, which controls for the year-specific seasonality in holiday purhcases.[14]

Table 2: Regression Results for Mean Stars and Shares Purchased

	Model	Mean Star	0 Star Sh	1 Star Sh	2 Star Sh	3 Star Sh
During=1	**(1)**	0.013***	-0.008**	-0.025***	-0.050***	0.027***
		(0.002)	(0.003)	(0.003)	(0.005)	(0.003)
Post=1		0.010***	-0.012***	0.020***	-0.012***	0.013***
		(0.002)	(0.003)	(0.003)	(0.004)	(0.003)
During=1	**(2)**	0.013***	-0.009**	-0.026***	-0.050***	0.028***
		(0.002)	(0.003)	(0.003)	(0.005)	(0.003)
Post=1		0.010***	-0.013***	0.018***	-0.012***	0.014***
		(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
During=1	**(3)**	0.013***	-0.009**	-0.026***	-0.050***	0.028***
		(0.003)	(0.003)	(0.004)	(0.005)	(0.003)
Post=1		0.010***	-0.013***	0.018***	-0.012***	0.014***
		(0.002)	(0.003)	(0.003)	(0.004)	(0.003)

(1) 2nd year, week-in-year, day-of-week, and store FE
(2) 2nd year, week-in-year, day-of-week, number of items per transaction, and store FE
(3) 2nd year, day-of-week, number of items per transaction dummies, and store x week-in-year FE
Store-clustered standard errors in parentheses
$^{+}p < 0.1$, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

This tables presents differences-in-differences interaction coefficients for three different specifications. In each specification, the number of observations is 31121. The dependent variable is in logs. All results are quantity weighted.

[14] We also do robustness checks removing November and December and find similar results.

Table 3: Regression Results for Various Nutrients

	Calories	Dietary Fiber	Protein	Omega 3	Total Fat	Saturated Fat	Trans Fat	Sodium	Sugar
During=1	0.003	0.028***	-0.003	-0.001	0.014***	0.018***	0.042***	-0.020***	0.001
	(0.002)	(0.002)	(0.003)	(0.007)	(0.003)	(0.005)	(0.010)	(0.003)	(0.002)
Post=1	0.004*	0.012***	-0.007*	0.005	0.011***	0.011***	-0.008+	-0.021***	0.009***
	(0.001)	(0.002)	(0.003)	(0.005)	(0.002)	(0.003)	(0.005)	(0.003)	(0.002)
2nd Yr=1	0.005***	0.002	-0.009***	0.007+	0.022***	0.009***	-0.021***	0.009**	-0.015***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)
Observations	31121	31121	31121	31121	31121	31121	31121	31121	31121

Store-clustered standard errors in parentheses

$^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

This table presents differences-in-differences interaction coefficients for different nutrients. The dependent variable is in logs. The regression includes day-of-week, number of items per transaction, and store times week-in-year fixed effects. All results are quantity weighted.

Table 2 presents three specifications where mean stars and the mean share of items purchased having a certain value of stars are the dependent variables. Since the size of the shopping basket and the nutrition characteristics are related, and becasue the campaign may affect the number of items in a basket, the second specification adds indicator variables for the average number of items per day, rounded to the nearest integer, to control for potential differences in basket size. The third specification replaces the week-of-year x banner fixed effects with the week-of-year x store fixed effects which allows for different stores to have different trends. Because of the log specification, we interpret coefficients as percentage changes from the mean. For example, the model (1) during effect of 0.013 suggests that the mean stars purchased increased by 1.3 percent or increased .017 units from a mean of 1.35. A .017 star increase is difficult to interpret as stars are an ordinal measure. This is simply suggestive that consumers are purchasing higher-rated products.

Table 2 also shows that during the campaign, there is an increase in the share of 3-star items, and a decline in the shares of all other items. These are different qualitative effects than those at the initial implementation of Guiding Stars, in which one and three star items exhibited increases in shares, with much of the effect driven by 1-star items (Hobin et al. 2017).

Despite the very modest increase in the star ratings of products purchased during the campaign, almost half of these effects also disappear immediately post campaign. In the month after the campaign concludes, the mean stars are 0.9% higher than the pre-campaign period, only half of the effect persists. The composition of this effect results from the changes due to the campaign in the share of 3-star items also declining by half, the decreases in zero and two star items during the campaign reversing modestly, and an increase in the share of one star items purchased.

Table 4: Regression Results by Category

	Mean Stars	Mean Star	0 Star Sh	1 Star Sh	2 Star Sh	3 Star Sh
Fruits & vegetables=1	2.799	0.002**	-0.028**	0.140***	-0.055***	0.004***
		(0.001)	(0.010)	(0.020)	(0.012)	(0.001)
Baby foods=1	2.236	0.002	-0.012	-0.025	-0.103+	0.005
		(0.003)	(0.013)	(0.029)	(0.062)	(0.004)
Health foods=1	1.037	0.031***	-0.030***	-0.018**	0.095***	0.018*
		(0.006)	(0.007)	(0.007)	(0.012)	(0.008)
Breads, grains, and breakfast cereals=1	0.982	-0.002	-0.029***	0.084***	-0.056***	0.002
		(0.004)	(0.007)	(0.008)	(0.007)	(0.011)
Meats, fish, legumes, & meat alternatives=1	0.971	-0.167***	0.098***	-0.008	-0.203***	-0.174***
		(0.012)	(0.007)	(0.018)	(0.018)	(0.014)
Dairy, dairy alternatives, & eggs=1	0.888	-0.056***	0.027***	-0.029***	-0.079***	-0.047***
		(0.005)	(0.003)	(0.007)	(0.010)	(0.005)
Beverages=1	0.703	-0.053***	0.029***	-0.021	-0.065***	0.209***
		(0.008)	(0.005)	(0.013)	(0.008)	(0.028)
Condiments, sauces, spreads, & salad fixing=1	0.527	0.080***	-0.019***	0.023+	-0.024	0.139***
		(0.010)	(0.003)	(0.013)	(0.019)	(0.013)
Baking & spices=1	0.416	-0.056***	0.017***	-0.083**	-0.114***	-0.031+
		(0.010)	(0.003)	(0.027)	(0.013)	(0.016)
Mixed dishes, soups, & side dishes=1	0.256	-0.046***	0.016***	-0.137***	0.077***	0.012
		(0.011)	(0.002)	(0.013)	(0.017)	(0.030)
Desserts and snacks=1	0.168	0.000	0.004**	-0.075***	-0.028	0.183***
		(0.011)	(0.002)	(0.011)	(0.019)	(0.017)

Store-clustered standard errors in parentheses
$^{+}p < 0.1$, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

This tables presents differences-in-differences interaction coefficients for different categories. Each row and column correspond to a different regression. The dependent variable is in logs. The regression includes day-of-week, number of items per transaction, and store times week-in-year fixed effects. All results are quantity weighted.

Because stars are an ordinal measure, and therefore difficult to interpret, we also present results by nutritional characteristics. Results are shown in Table 3. The campaign leads to an increase in fiber and fat (total, as well as saturated and trans fat), and a decrease in sodium. The coefficient of 0.028 for the during period of dietary fiber suggests that dietary fiber increased during the campaign by 2.8 percent per item or an extra 0.043 grams per item from the mean of 1.53 grams per item.

To assess what is driving the results, we run the same analyses shown above by category. The 11 food categories we assess were adapted from those used in the What We Eat in America dietary intake component of the National Health and Nutrition Examination Survey[15] and the Canadian Nutrient File[16] In Table 4 we report the estimated treatment effect for 11 categories, running a separate regression for each category with mean stars as the dependent variable. The categories which increase in mean stars the most are baby foods, health foods, condiments and salad fixings all increase in stars. In each of these categories there is an increase in the share of 3 star items for that category. [The following is new:] While not shown, there is a considerable decline in sodium for health foods and condiments and salad fixings. Baby foods show a sizable increase in both dietary fiber and protein.

4.2 Robustness Checks

We run a number of other specifications to test the robustness of these results. Table A.1 weights each items by the number of servings rather than the number of units (or by weight for weighed items). The serving weighted results are larger in magnitude during the campaign than those weighted by UPC volume, suggesting that in addition to purchasing a larger share of higher star rated products, consumers are either purchasing larger sizes of higher star items or smaller sizes of lower star items. In other words, the ef-

[15]United States Department of Agriculture. What We Eat in America food categories. http://www.ars.usda.gov/SP2UserFiles/Place/80400530/pdf/1112/food_category_list.pdf. Published 2012.

[16]Canadian Nutrient File. Government of Canada website. https://food-nutrition.canada.ca/cnf-fce/index-eng.jsp. Modified July 14, 2016.

fects are caused by a change in the volume of specific products in addition to substitution to and from products of different star levels. These effects are more driven by purchases of three star items, as those coefficients gain the most, while one star items see little change in their share with these results. Table A.2 weights purchases across store-days by the number of items purchased that store-day, rather than by the number of transactions. These results apply more weight to larger purchases, or store-days where more is sold per item, in the regressions. These results are very similar to the main specifications.

To test robustness to specification choices, in Table A.3, we use levels for the dependent variables, i.e. we do not log transform the dependent variable. The coefficient of 0.020 is consistent with the 0.017 level difference predicted by Table 2.

All of our results are consistent across these robustness checks. That said, the main assumption we made was that the previous year is a suitable control group. Although the seasonal trends are mirrored across years, the difference in the extent of the holiday effects lead us to estimate an alternative specification in which we only use the January-March period after the campaign starts, so that the campaign effect is estimated relative to the post-campaign levels. This would be the campaign effect if we assume that the entirety of the campaign effect disappeared after the campaign concluded, and thus can be considered a lower bound. Results are shown in Table 5. In this case, we find no significant effect on mean stars. We do, however, see a decline in both one and two star shares, and an increase in zero and three star shares. This highlights one of the main issues with the Guiding Stars program. Zero stars can indicate either no information, or unhealthy items since items that earn zero stars are simply not given a label, instead of being labelled with zero stars. This ambiguity has ramifications; consumers purchase less one and two star items, but they are not necessarily substituting to higher star products. We interpret this as a lower bound on the effect of the campaign.

Table 5: Regression Results for Mean Stars and Shares Purchased: January–March

	Model	Mean Star	0 Star Sh	1 Star Sh	2 Star Sh	3 Star Sh
During=1	**(1)**	-0.002	0.009***	-0.041***	-0.042***	0.008***
		(0.002)	(0.002)	(0.003)	(0.004)	(0.002)
During=1		-0.002	0.009***	-0.041***	-0.042***	0.008***
		(0.002)	(0.002)	(0.003)	(0.004)	(0.002)
During=1	**(2)**	-0.002	0.009***	-0.041***	-0.042***	0.008***
		(0.002)	(0.002)	(0.003)	(0.004)	(0.002)

(1) 2nd year, week-in-year, day-of-week, and store FE
(2) 2nd year, week-in-year, day-of-week, number of items per transaction, and store FE
(3) 2nd year, day-of-week, number of items per transaction dummies, and store x week-in-year FE
Store-clustered standard errors in parentheses
$^{+}p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This tables presents differences-in-differences interaction coefficients for three different specifications. In each specification, the number of observations is 31121. The dependent variable is in logs. All results are quantity weighted.

5 Household Analyses

While the main effect of the campaign is important, understanding which households are affected is also critical if the goal of the campaign is to increase the nutritional quality of food purchased (and consumed) by households who are not already doing so. Using each household's previous year as its control also alleviates concerns about the prior results being driven by a changing consumer composition over time. The household approach allows us to identify the average effects of interest exclusively from time-series variation, completely controlling for household heterogeneity. Furthermore, it allows us to examine the effects of the campaign more granularly. We check that the store level results are not due to changing consumer composition over time. Then we explore whether the campaign affects those who already eat the healthiest (or not) and how quickly shopping habits revert to pre-campaign levels. For these analysis, the unit of observation is a household's shopping trip, rather than an average of shopping trips at the store-day level. The data sample include 13,803,074 transactions by 834,134 households over the two year period.

In order to see whether the result is due to current customers changing their shopping patterns or new customers as a result of labelling, we ran similar regressions as before using household-level shopping data. We use the same regression equation as before in which the t subscript now indicates a transaction, and we replace store x week-of-year fixed effects with household x week-of-year fixed effects as shown below:

$$log(y_{it}) = \alpha(D_t \times \zeta_T) + \beta(P_t \times \zeta_T) + \zeta_T + \gamma_{iw} + \xi_d + \delta_{n_{it}} + \epsilon_{it}. \quad (2)$$

The number of item dummies are now for the number of items in that specific transaction (with dummies for each group of five). The results from the third specification (using household x week FE) are shown in Table 6.[17] We see the same pattern of results as we did with the aggregate data, with the campaign leading to an increase in the number of three star rated products purchased during the campaign and a smaller effect in the post-campaign period.

Table 6: Household Regression Results

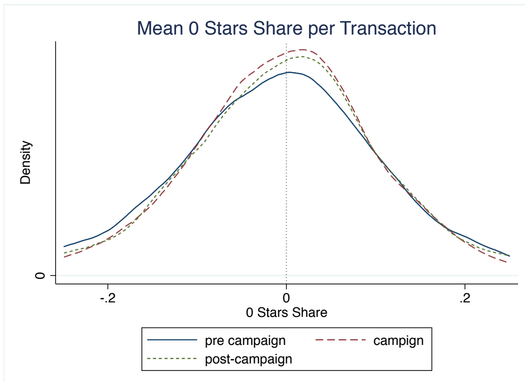
	Mean Star	0 Star Sh	1 Star Sh	2 Star Sh	3 Star Sh
camp	0.00247	-0.00997	-0.0539**	-0.106***	0.0275**
	(0.00639)	(0.00769)	(0.0193)	(0.0185)	(0.00959)
endcamp	0.00897+	-0.0147*	0.0631***	-0.0269+	0.0123
	(0.00528)	(0.00645)	(0.0163)	(0.0154)	(0.00801)
Observations	9686079	9686079	9686079	9686079	9686079

Household-clustered standard errors in parentheses
$^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

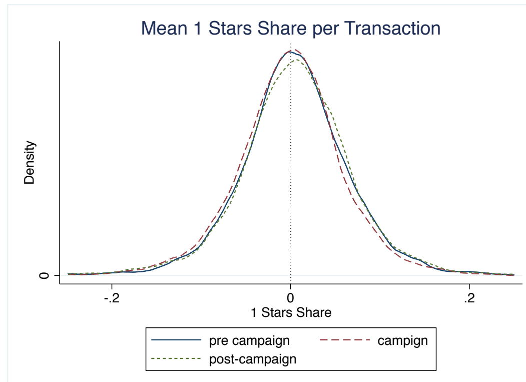
This table presents differences-in-differences interaction coefficients for household level results. Each column corresponds to a different regression. The dependent variable is in logs. The regression includes day-of-week, number of items per transaction, and store times week-in-year fixed effects.

To assess *household-specific* changes in shopping basket composition, we calculate the difference in the shares of the different star ratings of items purchased in the one month prior to the campaign, during the one month campaign, and one month post-campaign

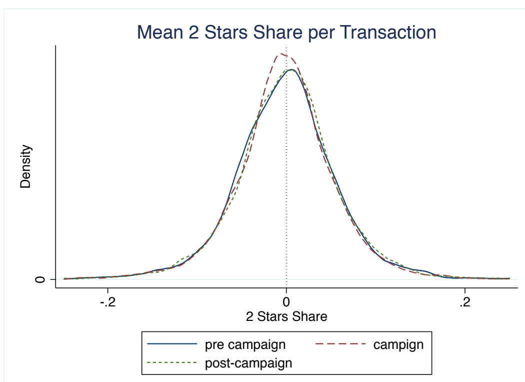
[17]We ran all of the same robustness checks as before, substituting household for store, and as with the store-day regressions, the results are virtually unchanged across specifications.



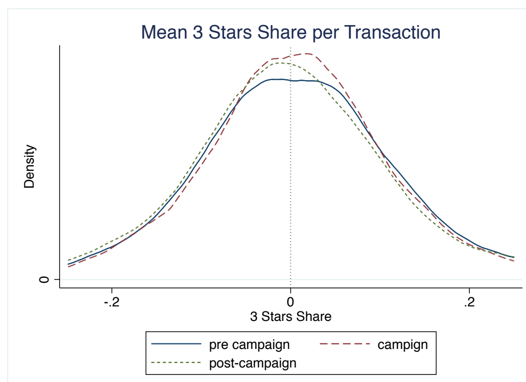
(a) Zero star share



(b) One star share



(c) Two star share



(d) Three star share

Figure 3: *HH Density of Changes in Shares*

This figure presents kernel density plots of the change in star level shares by household in comparison to the changes by the household in the previous year. The plots only include households who bought more than 50 items in each of the three periods. The 0 center line represents households who do not change the share of that amount of stars they purchase.

for households with at least 50 items purchased in all three periods, relative to the previous control year. Unlike with the regression results, this analysis focuses inference on just the month before the campaign, the month of the campaign, and the month after the campaign in order to remove the effects of macro trends. On average, there is an increase in the share of zero and three star shares (although the three-star share increase is insignificant on average) and a decrease in one and two star shares. Figure 3 shows the kernel densities of the household-level shifts in the shares of the different star levels. The most discernible shifts are the share of two star items decreasing and zero and three star items increasing during the campaign.[18] These results are consistent with the mean effects estimated using our difference-in-difference analysis in which consumers shifted away from one and two star items to zero and three star items during the campaign.

Another striking result is the reduction in the variance due to the campaign, in particular for the zero and three star items. This is clear from how much lower the peaks of the kernel densities are in the pre-period. This indicates that the campaign actually leads to more consistency in the star ratings of products purchased with the previous control year, providing the first evidence that the national promotional campaign may reinforce current habits rather than leading households to consume differently. One of the goals of nutrition labelling education is to increase the nutritional quality of food purchased by consumers who do not already purchase healthy foods, and these densities indicate that this may not be occurring.

To check whether those who purchase products with more stars revert quickly to their pre-campaign level shopping habits, we examine correlations between changes during the campaign to after the campaign. The first panel of table 7 shows correlations between during-campaign relative to the pre-campaign with campaign, and post-campaign relative to pre-campaign. There is large, positive correlation in the during and post campaign star shares such that 60 percent of the effects are maintained for all star rating levels. This suggests some persistence in the effect at the household level. However, the effect

[18]The results are robust to including the previous September as the ”pre” month which removes any differences in the holiday effect across years.

is less than perfectly correlated and roughly 40 percent of the effect disappears after the campaign ends.

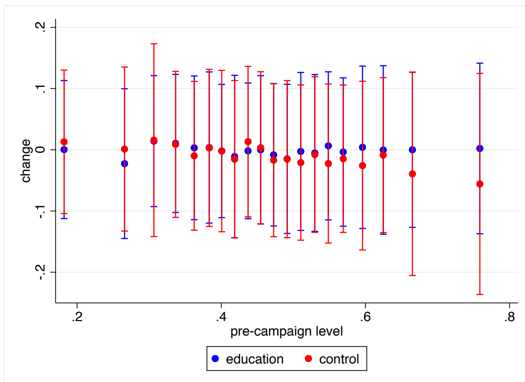
Table 7: Correlation Between Campaign and Post-Campaign Response

		post-campaign relative to before			
	Star	0 Star	1 Star	2 Star	3 Star
campaign	0	0.5929			
relative	1	-0.1313	0.5319		
to before	2	-0.1633	-0.0753	0.5813	
	3	-0.4588	-0.1214	-0.1059	0.59951

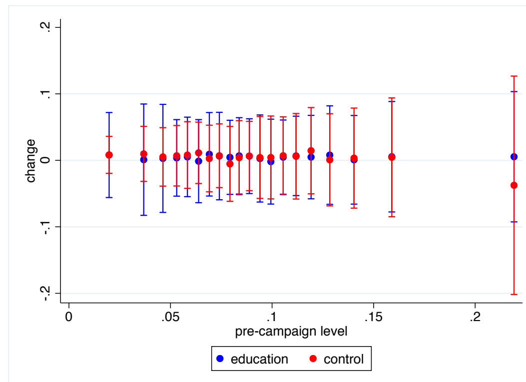
This table presents household level correlations between during-campaign relative to the pre-campaign with campaign, and post-campaign relative to pre-campaign.

Next, we explore whether consumers who already purchase relatively more nutritious baskets are the ones most impacted by the campaign. Because of the number of consumers, we use a bin-scatter plot, which bins consumers into 20 groups by their pre-campaign share, then takes the mean within each bin. For each discrete star rating (0-3), we plot the change in the shares of items purchased with that rating against that consumers' pre-campaign share for that rating.The mean shifts and the standard deviation across households are shown in Figure 4 for both the control and treatment year. In the control year (shown in red), there is regression to the mean as expected, but in the treatment year there is no such pattern. The effect of the campaign is to reinforce current purchasing habits for those in the upper-most ventile. This means those who purchase mostly zero star items purchase even more during the campaign, and those who purchase more three star items also purchase more of them during the campaign.

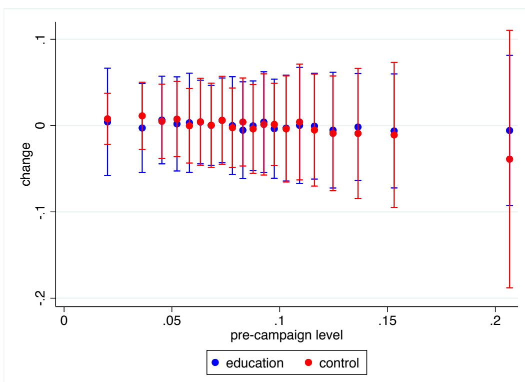
Significance for each ventile can be determined with a t-test by dividing the difference across years in the ventile means by the square root of the sum of squared standard deviations divided by the number of households in that ventile. The only significant differences across the years are for the top two gentiles for zero shares, and the top decile for one, two, and three star shares. Thus the only change during the campaign relative to the previous year is from a small fraction of households who are more likely to reinforce their current purchase patterns. The effect we see on zero star items can partially be



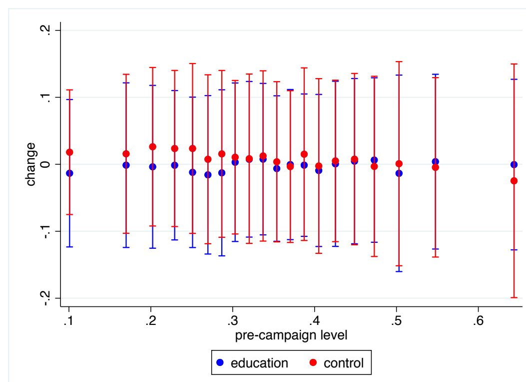
(a) Zero star share



(b) One star share



(c) Two star share



(d) Three star share

Figure 4: *Heterogeneity in Campaign Effects, Pre/Post*

This figure presents a "binscatter" plot of pre-campaign levels of shares of stars against their changes. The horizontal axis is grouped into 20 ventiles of pre-campaign levels of shares of stars. The vertical axis captures pre-campaign to during campaign difference in shares for both the year of the campaign and the previous year, showing both means and standard deviations.

explained by the ambiguity between items with zero stars because they are the least nutritious and items which are not rated at all. No distinction is made in the labelling, which was the sole departure from the National Academies' recommendations for effective FOP labelling systems.

5.1 Robustness Check

As with the transaction data, we also estimate a specification of the main household model using just the period after the campaign starts in early January. Results are in Table 8. We still see the decline in the share of one and two-star items, as we did with the transaction data. The effect on mean stars is again not significant. We again interpret this as a lower bound on the effect of the campaign.

Table 8: Household Regression Results: January–March

	Mean Star	0 Star Sh	1 Star Sh	2 Star Sh	3 Star Sh
camp	-0.00560	0.00598	-0.128***	-0.0759***	0.0173
	(0.00741)	(0.00889)	(0.0224)	(0.0218)	(0.0111)
Observations	2146005	2146005	2146005	2146005	2146005

Household-clustered standard errors in parentheses
$^{+}p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
This table presents differences-in-differences interaction coefficients for household level results. Each column corresponds to a different regression. The dependent variable is in logs. The regression includes day-of-week, number of items per transaction, and store times week-in-year fixed effects.

6 Mechanism

6.1 Design

Finally, we examine the impact of the national promotional campaign on consumer awareness, understanding, self-reported use, and trust of the Guiding Stars system in two separate supermarket banners in Ontario owned by the retailer. This allows us to explore some possible mechanisms for our results. Mainly, are more consumers aware of and using the labelling system after the launch of the national promotional campaign? To do this, we conducted a 10-minute survey with approximately 100 randomly selected shoppers outside each of 4 supermarket stores for banner 1 in the Greater Toronto Area and 4 supermarket stores for banner 2 in the Region of Waterloo, both of which are in Ontario. Eligible survey respondents included adults over the age of 18 years who reported purchasing food and/or beverage products at the supermarket on the day of the survey. In total, our survey sample includes 883 participants before, and 836 participants after the national promotional media campaign was implemented across Canada.

The exit survey included validated survey items assessing individual-level factors including age, gender, household size, ethnicity, self-reported health status, education levels, self-reported height and weight, perceived weight status, shopping and label use habits, self-reported diet quality, objective and perceived nutrition knowledge, and interest in nutrition. Consistent with measures applied in previous evaluations of nutrition interventions in supermarkets, the primary outcome measures were: unprompted consumer awareness (Did you notice a symbol on the shelf tag located underneath the product beside the price and if yes, Can you describe what you saw? a correct response included a reference to stars), understanding (Can you please tell me what the symbol means?- a correct response included a reference to health or nutrition or diet), self-reported use (Did you use the symbol to help you decide on food purchases today?), and trust (On a scale of 1 to 5, where 1 is not trustworthy and 5 is extremely trustworthy, please tell me the extent to which you trust the Guiding Stars symbols on shelf tags for providing food and nutrition information responses of 4 or 5 were categorized as trusting the Guiding Stars system) of the Guiding Stars system.

Using a within-groups design, we compared differences in consumer awareness, understanding, self-reported use, and trust among shoppers in both banners' supermarkets across 2 survey waves, representing time before and after the launch of the national promotional campaign. Using 4 separate logistic regression models, using awareness, understanding, self-reported use, and trust as dependent variables, we examined wave as a main effect, in unadjusted, and adjusted models, wherein supermarket and socio-demographic factors, shopping habits, and food-related knowledge and behaviors were included as covariates.

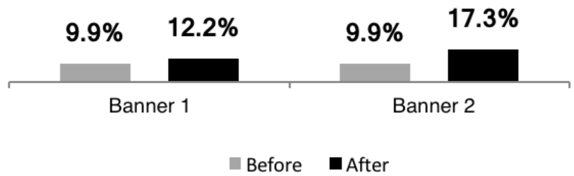
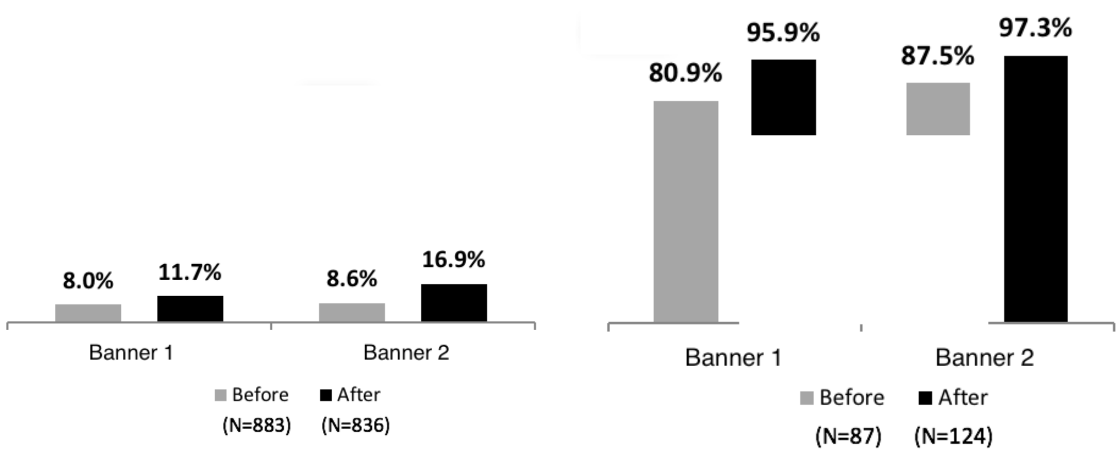


Figure 5: *Self-Reported Awareness of Star Labels*

Survey results averaged across respondents.

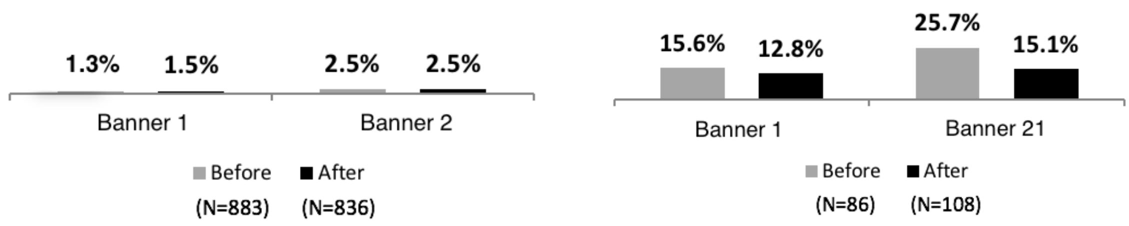


(a) Unconditional

(b) Conditional on Awareness

Figure 6: *Self-Reported Understanding of Star Labels*

Survey results averaged across respondents.



(a) Unconditional (b) Conditional on Awareness & Understanding

Figure 7: *Self-Reported Use of Star Labels*

Survey results averaged across respondents.

6.2 Results

Figure 5 shows the percentage of respondents who were aware of the Guiding Stars system before and after the launch of the national promotional campaign in the supermarkets. Figure 6 shows the percentage of respondents who reported that they understood the labels (overall and among those who are aware), and Figure 6 shows the percentage who reported using the labels (overall and among those who are aware and understood). In total, 25.4% and 28.8% of respondents indicated they trust the Guiding Stars system before and after the national promotional campaign, respectively.

Results from the adjusted logistic regression models indicate that only awareness and understanding were significantly higher after relative to before the launch of the national campaign (OR=1.6, 95% CI:1.2, 2.2, p=0.004; OR=1.9, 95CI:1.3, 2.6, p=0.0002, respectively). Consumer factors including shopping at their main supermarket (OR=1.5, 95%CI:1.0-2.1, p=0.04), and those who reported using the Nutrition Facts panel (OR=1.6, 95%CI:0.2-2.3, p=0.02) were significantly associated with awareness of the Guiding Stars labels. Whereas, consumers ages 25-44 years were more likely to report understanding the Guiding Stars system (OR=1.5, 95%CI:1.0 2.1, p=0.04) compared to those 45-64 years, and those who reported using the Nutrition Facts panel (OR=1.6, 95%CI:1.1-2.2, p=0.02). Consumers 65+ years were less likely to understand the system (OR=0.5, 95%CI:0.3-0.9, p=0.03).

There were no significant differences in self-reported use before and after the launch of the national campaign (Adjusted OR=0.8, 95%CI:0.4-1.7, p=0.59). However, results indicate those who reported trusting the Guiding Stars system were more likely to report using this system when shopping in the supermarket (OR=2.7, 95% CI: 1.9-4.0, $p<.0001$). Finally, consumers ages 18-24 years (OR=0.5, 95% CI: 0.2-1.0, p=0.04 and 25-44y, OR=0.7, 95% CI: 0.5-0.9, p=0.02) were less likely to trust the Guiding Stars system than those 45-64 years. Those that were shopping for someone with a health condition, and those reporting that they use the Nutrition Facts Panel were also less likely to report trusting Guiding Stars (OR=0.7, 95% CI: 0.6-0.9, p=0.01); OR=0.8, 95% CI: 0.6-1.0, p=0.04, respectively). In contrast, those shopping at their main supermarket, and those reporting that they understood the Guiding Stars system, were more likely to report trusting the Guiding Stars system (OR=1.4, 95% CI:1.1-1.9, p=0.01; OR=4.2, 95% CI:3.0-5.9, $p<.0001$, respectively).

7 Discussion

7.1 Limitations

We do find a positive effect of the campaign on the healthiness of consumer purchases (as measured by their star ratings). The main limitation of the study is the lack of a contemporaneous control group, and it is thus not clear that the entirety of the effect can be attributed to the campaign. The use of household level data helps to remedy this concern, since changes across years from compositional effects are ruled out as an alternative explanation. Even if one does attribute the entire estimated effect to the campaign, the behavioral responses are still small in magnitude, indicating that such campaigns may not be sufficient to nudge consumers to change their behaviors.

One issue could be in the campaign itself. The campaign was designed to increase the awareness, understanding, and use of the Guiding Stars program, but it could be that other messaging might have been more effective. The campaign was also fairly short, with a duration of one month, with no meaningful follow-up activities to encourage consumers to continue to purchase healthier items. In addition, one of the main critiques of the Guiding Stars program is the ambiguity associated with zero-stars, but the campaign did nothing to help remedy any confusion for consumers. The in-store materials directed consumers to look for one, two, and three star products but did not explain 0-star items or the absence of 0-star labels.

That said, the fact that we did observe a (small) increase in the awareness and understanding of the campaign indicates that we cannot simply blame the campaign specifics for the small effect on purchases. The fact that awareness and understanding of the labels increased but usage of the labels did not is much more indicative that the campaign was effectively designed, but that many consumers are still not interested in using the labels. This is what leads us to question whether other campaigns can overcome this last hurdle, getting consumers who understand the labels and are not using them to begin using them.

7.2 Concluding Remarks

Moorman et al. (2012) and others who have examined the effect of the NLEA on nutritional quality of existing products have also advocated for the use of FOP labels as a way to increase consideration of nutritional content at the point of sale, and there is evidence that such labels may increase understanding and purchase intent (Newman et al. 2015, 2018). However, only recently have we seen such efforts in practice. However, the emerging literature examining the effectiveness of simplified nutritional labelling systems for consumer packaged goods in grocery markets on increasing the nutritional quality of food purchases has shown a significant but very modest effect. Previous research has found a positive effect of state-level nutrition education on obesity McGeary (2009), and consumer education through informative promotional campaigns in conjunction with labeling has been advocated by the National Academies, with the hope that these campaigns can increase awareness, understanding, and use of the labels.

After the national campaign, our results indicate consumers purchased foods with moderately higher star ratings relative to before the campaign. The pattern in the star ratings of consumer food purchases after the national campaign reflects consumers purchasing patterns before the campaign in that, consumers who bought foods with higher star ratings at baseline were even more likely to purchase foods with higher star ratings at follow-up. I other words, we find that those consumers increasing the healthiness of their purchases are those that are already consuming healthy food, and that consumers purchasing the least healthy foods actually purchase less healthy products after the campaign. The campaign thus reinforces current behavior. If the consumers who are consuming less healthy food have lower income, then our results are consistent with both McGeary (2009) and Bollinger et al. (2011) who find that higher income, more educated consumers are more affected by information. More research is needed to identify the distributional impacts of such policies.

These small or insignificant effects can be explained by our exit survey data, which indicate that consumer awareness and understanding of the Guiding Stars system slightly increased after the campaign, but self-reported use of the labelling system remained extremely low. One potential explanation that the use of the Guiding Stars system is low overall, and remains low even after the national campaign, is the majority of products in the supermarkets earn a 0-star rating, and, under the Guiding Stars system, no star symbol is displayed on shelf tags to communicate to consumers that these products have been rated as relatively less nutritious. This labelling limitation, in conjunction with the small response to the campaign and the fact that the campaign is not leading to higher star purchases by those consumers whom policymakers would most want to target, suggest that other policy instruments would be necessary if the goal is to affect change for more than a small fraction of consumers.

References

Barragan, N. C., Noller, A. J., Robles, B., Gase, L. N., Leighs, M. S., Bogert, S., Simon, P. A. and Kuo, T. (2014). The sugar pack health marketing campaign in los angeles county, 2011-2012, *Health Promot Pract March* **15**(2): 208–216.

Bollinger, B., Leslie, P. and Sorensen, A. (2011). Calorie posting in chain restaurants, *American Economic Journal: Economic Policy* pp. 91–128.

Cawley, J., Sweeney, M. J. and Sobal, J. e. a. (2014). The impact of a supermarket nutrition rating system on purchases of nutritious and less nutritious foods, *Public Health Nutr* .

Davis, R. M., Gilpin, E. A., Loken, B., Viswanath, K. and Wakefield, M. A. (2008). The role of the media in promoting and reducing tobacco use, *Technical report*, U.S. Department of Health and Human Services.

Hawley, K. L., Roberto, C. A., Bragg, M. A., Liu, P. J., Schwartz, M. B. and Brownell, K. D. (2013). The science on front-of-package food labels, *Public Health Nutrition* **16**(3): 430–439.

Hobin, E., Bollinger, B., Sacco, J., Liebman, E., Vanderlee, L., Rosella, L., LAbbe, M., Manson, H. and Hammond, D. (2017). Shoppers response to an on-shelf nutrition labelling system in supermarkets: evidence to inform policy and practice., *Milbank Quarterly* **95**(3): 494–534.

Ippolito, P. M. and Mathios, A. D. (1990). The regulation of science-based claims in advertising, *Journal of Consumer Policy* **13**(4): 413–445.

Ippolito, P. M. and Mathios, A. D. (1995). Information and advertising: The case of fat consumption in the united states, *The American Economic Review* **85**(2): 91–95.

Katz, D. L., Njike, V. Y., Rhee, L. Q., Reingold, A. and Ayoob, K. T. (2010). Performance characteristics of nuval and the overall nutritional quality index, *Am J Clin Nutr* **91**(4): 1102S–1108S.

Kiesel, K. and Villas-Boas, S. (2008). Another nutritional label: experiementing with grocery store shelf labels and consumer choice, *CUDARE Working Paper Series 1060* .

Lang, J. E., Mercer, N., Tran, D. and Mosca, L. . (2000). Use of a supermarket shelf-labeling program to educate a predominately minority community about foods that promote heart health, *Journal of the American Dietetic Association* **100**: 804–809.

Levy, A. S., Mathews, O., Stephenson, M., Tenney, J. E. and Schucker, R. E. (1985). The impact of a nutrition information program on food purchases, *Journal of Public Policy and Marketing* **4**(113.).

Liu, P. J., Roberto, C. A., Liu, L. J. and Brownell, K. D. (2012). A test of different menu labeling presentations, *Appetite* **59**: 770–777.

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): 651–678.

McGeary, K. A. (2009). The impact of state-level nutrition-education program funding on bmi: Evidence from the behavioral risk factor surveillance system. NBER Working Paper No. 15001.

Moorman, C., Ferraro, R. and Huber, J. (2012). Unintended nutrition consequences: Firm responses to the nutrition labeling and education act, *Marketing Science* **31**(5): 717–737.

Newman, C. L., B., S., Andrews, J. C., Netemeyer, R. G. and Kees, J. (2018). Marketersuse of alternative front-of-package nutrition symbols:an examination of effects on product evaluations, *Journal of the Academy of Marketing Science* **46**(3): 453–476.

Newman, C. L., Howlett, E. and Burton, S. (2015). Effects of objective and evaluative front-of-package cues on food evaluation and choice: The moderating influence of comparative and noncomparative processing contexts, *Journal of Consumer Research* **42**(5): 749–766.

Noar, S. (2006). A 10-year retrospective of research in health mass media campaigns: Where do we go from here?, *Journal of Health Communication: International Perspectives* **11**: 21–42.

Rahkovsky, I., Lin, B.-H., Lin, C.-T. J. and 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.

Ross, C., Shaw, S., Marshall, S., Stephen, S., Bailey, K. and et al. (2016). Impact of a social media campaign targeting men who have sex with men during an outbreak of syphilis in winnipeg, canada, *Canada Communicable Disease Report* **42**(2): 45–49.

Sacks, G., Rayner, M. and Swinburn, B. (2009). Impact of front-of-pack traffic-light nutrition labelling on consumer food purchases in the UK., *Health promotion international* **4**(24): 344–352.

Schmeiser, S. (2014). Consumer inference and the regulation of consumer information, *International Journal of Industrial Organization* **37**: 192–200.

Sonnenberg, L., Gelsomin, E., Levy, D. E., Riis, J., Barraclough, S. and 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**(4): 253–257.

Stekler, J. D., Baldwin, H. D., Louella, M. W., Katz, D. A. and Golden, M. R. (2013). ru2hot?: A public health education campaign for men who have sex with men to increase awareness of symptoms of acute hiv infection, *Sex Transm Infect* **89**: 409–414.

Sutherland, L. A., Kaley, L. A. and Fischer, L. (2010). Guiding stars: The effect of a nutrition navigation program on consumer purchases at the supermarket, *Am J Clin Nutr* **91**(4): 1090S–1094S.

Walls, H. L., Peeters, A., Loff, B. and Crammond, B. R. (2009). Why education and choice won't solve the obesity problem, *American Journal of Public Healt* **99**(4): 590–592.

Appendix

Table A.1: Serving Weighted Differences-in-Differences Results

	Model	Mean Star	0 Star Sh	1 Star Sh	2 Star Sh	3 Star Sh
During=1	**(1)**	0.027***	-0.008***	-0.005+	-0.074***	0.069***
		(0.002)	(0.001)	(0.003)	(0.006)	(0.003)
Post=1		0.009***	-0.005**	0.012***	-0.024***	0.020***
		(0.003)	(0.002)	(0.003)	(0.005)	(0.003)
During=1	**(2)**	0.027***	-0.008***	-0.008**	-0.075***	0.070***
		(0.002)	(0.001)	(0.003)	(0.006)	(0.003)
Post=1		0.009**	-0.004**	0.010**	-0.025***	0.021***
		(0.003)	(0.002)	(0.003)	(0.005)	(0.004)
During=1	**(3)**	0.028***	-0.008***	-0.008*	-0.074***	0.070***
		(0.002)	(0.001)	(0.003)	(0.006)	(0.003)
Post=1		0.009**	-0.005**	0.010**	-0.025***	0.021***
		(0.003)	(0.002)	(0.003)	(0.006)	(0.004)

(1) 2nd year, week-in-year, day-of-week, and store FE
(2) 2nd year, week-in-year, day-of-week, number of items per transaction, and store FE
(3) 2nd year, day-of-week, number of items per transaction dummies, and store x week-in-year FE
Store-clustered standard errors in parentheses
$^{+}p < 0.1$, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

This tables presents differences-in-differences interaction coefficients for three different specifications. In each specification, the number of observations is 31121. The dependent variable is in logs. The first specification denoted by (1) has second year, week-in-year, day of week and store fixed effects. Unlike the results in the draft, this table uses serving weights rather than quantity weights.

Table A.2: Transactions are Weighted Across Stores by Number of Transactions and Items

	Mean Star	0 Star Sh	1 Star Sh	2 Star Sh	3 Star Sh
During=1	0.013***	-0.009**	-0.028***	-0.049***	0.028***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
Post=1	0.011***	-0.014***	0.018***	-0.011**	0.014***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
2nd Yr=1	-0.014***	0.013***	0.003+	0.007*	-0.020***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Observations	31121	31121	31121	31121	31121

Standard errors in parentheses
$^{+}p < 0.1$, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

This tables presents differences-in-differences interaction coefficients. The dependent variable is in logs. The regression includes day-of-week, number of items per transaction, and store times week-in-year fixed effects. Unlike the results in the draft, this table weighs by number of transactions and number of items across stores. Each item is quantity weighted.

Table A.3: Differences-in-Differences Results In Levels

	Mean Star	0 Star Sh	1 Star Sh	2 Star Sh	3 Star Sh
During=1	0.020***	-0.003**	-0.003***	-0.005***	0.010***
	(0.003)	(0.001)	(0.000)	(0.000)	(0.001)
Post=1	0.014***	-0.006***	0.002***	-0.001**	0.005***
	(0.003)	(0.001)	(0.000)	(0.000)	(0.001)
2nd Yr=1	-0.020***	0.006***	0.000+	0.001+	-0.007***
	(0.003)	(0.001)	(0.000)	(0.000)	(0.001)
Observations	31121	31121	31121	31121	31121

Standard errors in parentheses
$^{+}p < 0.1$, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

This tables presents differences-in-differences interaction coefficients. Unlike the results in the draft, the dependent variable is in levels, rather than logs. The regression includes day-of-week, number of items per transaction, and store times week-in-year fixed effects. Results are quantity weighted.