The Effect of Front-Of-Package Nutrition Labelling on Product Composition

Christoph Bauner, University of Massachusetts Amherst, cbauner@umass.edu
Rajib Rahman, University of Massachusetts Amherst, mdrajiburrah@umass.edu

Selected Paper prepared for presentation at the 2023 Agricultural & Applied Economics Association Annual Meeting, Washington DC; July 23-25, 2023

The Effect of Front-Of-Package Nutrition Labelling on Product Composition

Christoph Bauner and Rajib Rahman*

May 8, 2023

Abstract

We analyze the effect of front-of-package nutrition labelling using the example of France, where Nutri-Score was adopted in 2017. Our focus is on changes in available products, i.e., on producer choices rather than consumer choices. Employing a difference-in-differences approach, we find that products introduced or altered after the change receive better Nutri-Score ratings than those introduced before the adoption, indicating a shift to items that are healthier overall. In addition, there is some evidence of bunching at the cutoffs for better Nutri-Score grades which suggests the improvements are at least in part a strategic reaction to the Nutri-Score introduction.

*We thank Colin Johnston for excellent research assistance.

1 Introduction

Poor diet is a major contributor to many serious health conditions such as obesity, heart disease, and diabetes (Centers for Disease Control and Prevention, 2021). One method by which governments and health organizations try to nudge people toward a healthier diet is by front-of-package nutrition labels (FoPLs). FoPLs are designed to prominently place easily understandable nutrition information on food products.

In this article, we investigate the impact of one FoPL, the Nutri-Score adopted by France in 2017 and since by several other countries, on new products in several food categories. We use detailed product and nutrition data from 2014 to 2021 and employ the difference-in-differences methodology, comparing changes in French foods to those in Italy and the United Kingdom where to date Nutri-Score has not been adopted. Our analysis shows a significant improvement in Nutri-Scores of new products, implying that the overall product composition became healthier after the Nutri-Score introduction. In addition, we find evidence of bunching at Nutri-Score cutoffs for two of the three product categories investigated. These results suggest that FoPLs can be effective in establishing a healthier set of choices for consumers of packaged foods. As such, they can be one tool the public health community can wield to combat nutrition-related diseases.

Our study focuses on three product categories: breakfast cereals, popcorn, and potato snacks. These categories have several advantages over many others. First, they have long been established and a significant number of products has been introduced in the French market both before and after the Nutri-Score adoption. Second, products in these categories are highly processed so that producers are able to significantly affect their nutrient composition while. At the same time, they are more homogeneous than, e.g., ready-made dinners, and are therefore more comparable. Finally, they tend to have a low fruit content which, as we will explain in more detail below, is important for our determination of the products' Nutri-scores.

Much of the prior research on FoPLs consists of studying consumer responses to FoPLs in experimental settings. Many articles in this literature have found that FoPLs can significantly affect consumer perception of and preference or willingness-to-pay for products (Van Wezemael et al., 2014; Ares et al., 2018; Tórtora et al., 2019; Franco-Arellano et al., 2020).

Research using real-world data is much scarcer. Sacks et al. (2010) and Freedman and Connors (2010) find no significant effects of FoPLs using grocery store and convenience store sales data from the United Kingdom, respectively. Similarly, a meta-study by Ikonen et al. (2020) finds only weak and decreasing effects of FoPLs on purchase behavior.

In contrast, several papers studying the effects of the introduction of mandatory warning labels in Chile in 2016 have established much more encouraging results. These labels inform consumers about high sugar, sodium, saturated fats, and calorie content and were combined with an advertising ban of labeled products in certain TV programs. Araya et al. (2022) find that this policy reduced demand for labeled breakfast cereals while no such effect is apparent for chocolates and cookies. Similarly, Barahona et al. (2020) and Barahona et al. (2022) find an overall reduction of caloric intake and sugar consumption after policy introduction. Overall, however, there is still uncertainty about the effects of FoPLs on consumer actions. In the light of this debate, analyzing producer responses to FoPLs gains increased importance.

A few studies have dealt with producer reactions to FoPLs. Using data from Dutch retailers, all of whom had just adopted the Choices logo, Vyth et al. (2010) suggest that food manufacturers reformulate products to make them healthier when confronted with FoPLs. However, their results are based on extremely small sample sizes and when applying a correction for the fact that more than 60 hypotheses are tested, many of their null-hypotheses cannot be rejected anymore. Overall, we find the evidence presented inconclusive.

Several other papers study the effect of the aforementioned Chilean warning labels on manufacturer strategies. Alé-Chilet and Moshary (2022) find that the policy lead to lower caloric intake and to bunching below the cutoffs. Pachali et al. (2022) present evidence that mandatory warning labels lead to price adjustments, in particular price decreases for most unlabeled products and increases for labeled products. Barahona et al. (2022) show that the introduction of warning labels lead to product reformulation, in particular bunching just below the relevant cutoffs. Similarly, Barahona et al. (2020) argue that the warning labels effected strong supply side responses including price adjustments and product reformulations that lead to improved nutritional intake and increased consumer surplus. It is important to remember that these studies focus on the case of warning labels.

In contrast, we analyze the effect of Nutri-Score, a reductive label that summarizes nutrition information in an easily digestible format. Ikonen et al. (2020) show that reactions to these different types of FoPL may differ and, in particular, that consumers may react more strongly to reductive labels.

Our setting also allows for cleaner identification. Studies on the Chilean warning labels are typically set up as an event study comparing the same market before and after label introduction. We employ the classic difference-in-differences methodology comparing France with other countries that did not implement Nutri-Score. In addition, no advertising ban coincided with the introduction of Nutri-Score in France reducing the risk of mis-attributing effects. Despite these differences, our results are largely consistent with those of prior authors: We find improvements in the nutritional profile of products.

Our results suggest that FoPL policies can be effective at furthering public health goals. While one widely studied mechanism for such improvements is to enable consumers to make more informed choices between existing products, we are pointing to evidence of an additional mechanism: FoPLs provide an incentive for manufacturers to reformulate their products or develop new, healthier alternatives. Thus, they can help promote the broader availability of products with positive nutritional attributes. This can be particularly important for populations, e.g. rural and low-income demographics, that may not have easy access to full-range grocery stores and, therefore, rely less on fresh produce and more on consumer packaged goods for their meals.

We also find some evidence for bunching beneath thresholds, but this varies by product category. We thus contribute to an emerging consensus that manufacturers react significantly and strategically to FoPLs, although these responses may differ across product categories. These findings indicate that policy makers interested in adopting FoPLs need to carefully consider the features of the specific label under debate. In particular, a system with more nuanced grading may be more difficult to game for manufacturers. This has to be weighed against the cognitive burden for consumers as complicated labels may face the risk of being used sparingly by consumers, similar to the current back-of-package nutrition labels.

The remainder of this paper is organized as follows: Section 2 provides some background on the Nutri-Score. Section 3 gives an overview of our data and shows some preliminary evidence of changes in product formulations. Section 4 presents our main analysis applying the difference-in-differences approach. Section 5 concludes with a discussion of the policy relevance of our findings.

2 Nutri-Score Design and Adoption

Nutri-Score is a simple, easily understandable rating system for the overall nutritional value of foods. Nutri-Score is originally calculated as a numerical value which is then converted to a letter grade from A (healthiest) to E (unhealthiest) and is generally provided as a color-coded FoPL. Proponents believe it helps consumers make healthier food choices by eliminating the necessity of reading and understanding the more detailed and more complicated nutrition information already provided.

The first proposal for introducing FoPLs in France was made in 2014. After the proposal went through a consultation process with multiple stakeholders, the Nutri-Score formula and label design were finalized based on numerous peer-reviewed studies (Hercberg et al., 2021). On October 31st, 2017, Nutri-Score was adopted in France on a voluntary basis (Julia et al., 2018).

Nutri-Score ratings are based on a comparison of healthy (e.g., protein) and unhealthy (e.g., sodium) nutrients as well as fruit and vegetable content. For each nutrient, a discrete number of points from 0 to 10 are assigned based on the amount contained in the product. The points for healthy nutrients and fruit and vegetables are then subtracted from the sum of points for unhealthy nutrients before the final score is converted to a letter grade between A and E. Julia and Hercberg (2017) provide a more detailed explanation of the calculation.

Several European countries, including Spain in 2018, Belgium in 2019, and Germany in 2020, have followed France's lead and adopted Nutri-Score labelling themselves. Currently, Nutri-Score is voluntarily applied by producers. However, the Nutri-Score label is registered as trademark across the European Union and Santé Publice France as the rightsholder requires each food producer intending to use Nutri-Score on their products to register the usage and to provide Nutri-Score on all of their products within two years.[1]

The initial rollout of Nutri-Score included only six manufacturers.[2] However, by July 2018, about 70 firms had adopted Nutri-Score voluntarily (Mialon et al., 2018), and in early 2022 the number had risen to more than 500 participating firms, covering about 50% of food products in France (Nutri-Score Assessment Report, 2021). Consumer feedback on Nutri-Score to assess the nutritional quality of products is improving over the years since its adoption in 2017. According to Public Health France's findings, about 1 percent consumers had positive feedback about Nutri-Score label to assess the

1 Santé Publice France, Conditions of the Use of the « Nutri-Score » Logo, January 2022, https://www.santepubliquefrance.fr/media/files/02-determinants-de-sante/nutrition-et-activite-physique/nutri-score/reglement-usage-en

2 Auchan, Intermarché, Leclerc, Fleury Michon, Danone, and McCain

nutritional quality of products in April 2018. This increased to 18 percent in 2020 (Nutri-Score Assessment Report, 2021; Southey, 2021).

Despite the increased uptake and acceptance, still only 40% of food products in France bear the Nutri-Score label. This has lead to campaigns urging producers to voluntarily display Nutri-Scores, and to calls for mandatory adoption throughout the European Union (Southey, 2022; FoodWatch, 2021; Das, 2020). At the same time, Nutri-Score has come under some criticism for reasons ranging from posing a perceived threat to traditional foods in some cultures to being too simplistic, not taking into account personalized diets (Julia et al., 2022).

3 Data and Preliminary Analysis

Our data come from the Mintel Global New Products Database and contain information on new products found in grocery stores in numerous countries. The most important variables for our purposes are nutrition facts. The database also contain information on package size, prices, and ingredients, among many other variables. While the provided data is generally of high quality, some errors do occur. We run several tests to identify such errors, e.g., comparing salt and sodium content when both are provided, comparing total reported nutrient content to the serving size, scanning for implausibly large values, and spot checking. The Mintel Global New Products

Database provides photos of the product packages, including the nutrition facts label, so that we are often able to manually correct the erroneous information.

Geographically, our focus is on France as the country that first adopted Nutri-Score, as well as Italy and the United Kingdom as controls. Other countries such as Germany and Belgium are harder to use as viable controls as they introduced Nutri-Score shortly after France did. We use data from 2014 to 2021 to have a substantial time frame before and after 2017, the year of the policy introduction in France.

We analyze three distinct product categories: breakfast cereals, popcorn, and potato snacks. Table 1 provides an overview of Nutri-Scores by country and product category. We calculate these scores according to the published Nutri-Score information. However, our values may deviate somewhat from actual Nutri-Scores because the Nutri-Score calculation takes into account the percentage of fruit or vegetables in the product. While we observe ingredients, we do not see the percentages of each ingredient, so we are forced to ignore this component. Nonetheless, we believe that our partial Nutri-Scores are typically fairly accurate as, for our chosen categories, fruits and vegetables are unlikely to constitute more than 40% of the product, the lowest threshold for improving the Nutri-Score. This certainly seems an unlikely bar to clear for the popcorn and potato snack categories.[3]

[3] Guidance published by Santé Publice France specifically excludes potatoes and popcorn from consideration as vegetables. See

In addition, at least for the analysis of overall composition of products, this missing piece of information would only affect our results if, beyond the sheer impact of ingredients on nutrition, manufacturers fruit and vegetable choices were directly correlated with nutrition facts, e.g., if manufacturers were using fruit content to make up for a high sodium content.

Table 1: Descriptive Statistics for Nutri-Scores

	Mean	Std. Dev.	Min	Max	N
Breakfast Cereals					
France	2.89	4.61	-9.00	20.00	1,168
Italy	1.87	4.54	-7.00	16.00	447
UK	1.02	4.53	-7.00	19.00	1,498
Popcorn					
France	8.76	8.24	-6.00	32.00	152
Italy	4.18	7.22	-7.00	16.00	60
UK	8.29	7.27	-7.00	29.00	390
Potato Snacks					
France	8.52	3.65	-1.00	20.00	853
Italy	7.95	4.94	-2.00	24.00	393
UK	6.51	3.50	-3.00	21.00	874

Note: Nutri-Score based on own calculation using Mintel GNPD data and ignoring fruit & vegetable content. A lower score indicates a healthier product. Cutoffs for Nutri-Score letter grades are: score ≤ -1 for A, $0 \leq$ score ≤ 2 for B, $3 \leq$ score ≤ 10 for C, $11 \leq$ score ≤ 18 for D, and $19 \leq$ score for E.

Table 1 shows that, on average, French products in our sample are less healthy than those in Italy and the United Kingdom, i.e., have a

https://www.santepubliquefrance.fr/media/files/02-determinants-de-sante/nutrition-et-activite-physique/nutri-score/qr-scientifique-technique-en higher Nutri-Score. This may induce skepticism of the effectiveness of Nutri-Score in guiding manufacturer choices. However, as the presented data provide include products introduced before and after the Nutri-Score adoption, the FoPL may still have a beneficial effect.

To look further into this question, we compare products introduced to the French market before and after the adoption of the Nutri-Score in France. For each of our product categories, figure 1 shows the empirical CDFs of calculated Nutri-Scores for the two time periods. For each product category, the CDF after policy introduction is overwhelmingly to the top-left compared to before. This indicates an improvement in terms of nutritional value of new products sold in France.

Figure 2 presents the changes of breakfast cereals at the nutrient level. While there are no substantial changes for energy and saturated fat, products contain less sugar and sodium and more fibre and protein after the adoption of Nutri-Score. This confirms the impression from the analysis of Nutri-Scores that the FoPL policy coincided with the introduction of healthier products.

The picture is similar for popcorn with reductions of energy, sugar, and sodium, and increases in fibre and protein (see figure 3). For potato snacks no significant changes are visible, except a slight increase in fibre content (see figure 4). Across the three product categories, we see an improvement nutritional content. However, the evidence presented so far is preliminary in the sense that no causal-

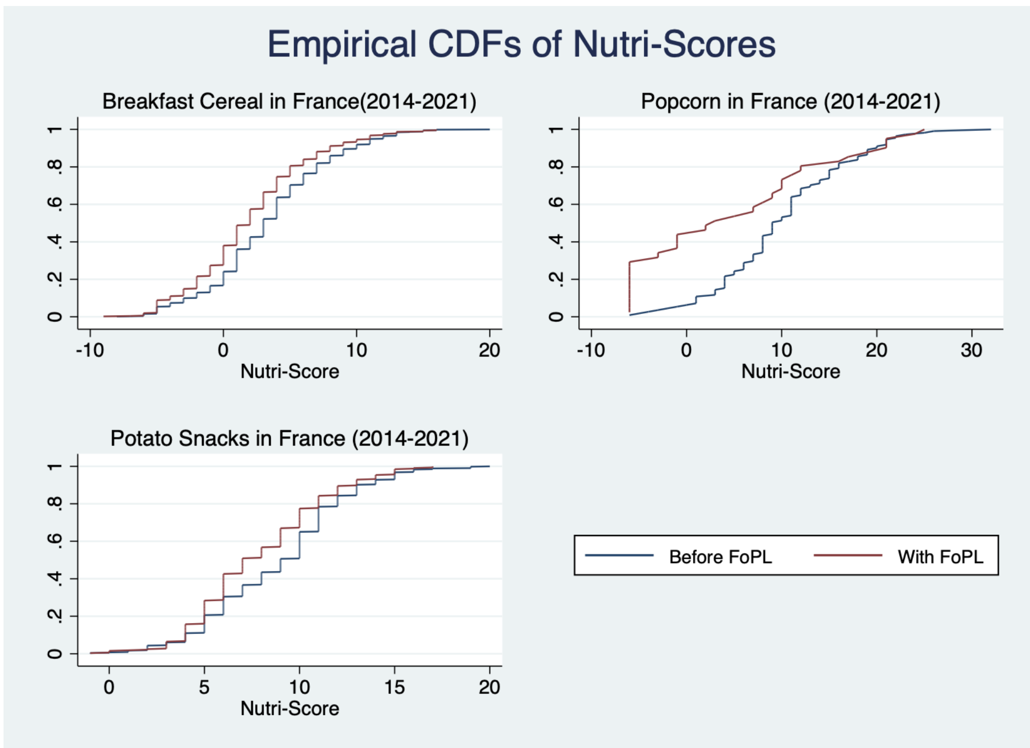

Figure 1: Empirical CDFs of Nutri-Score in France

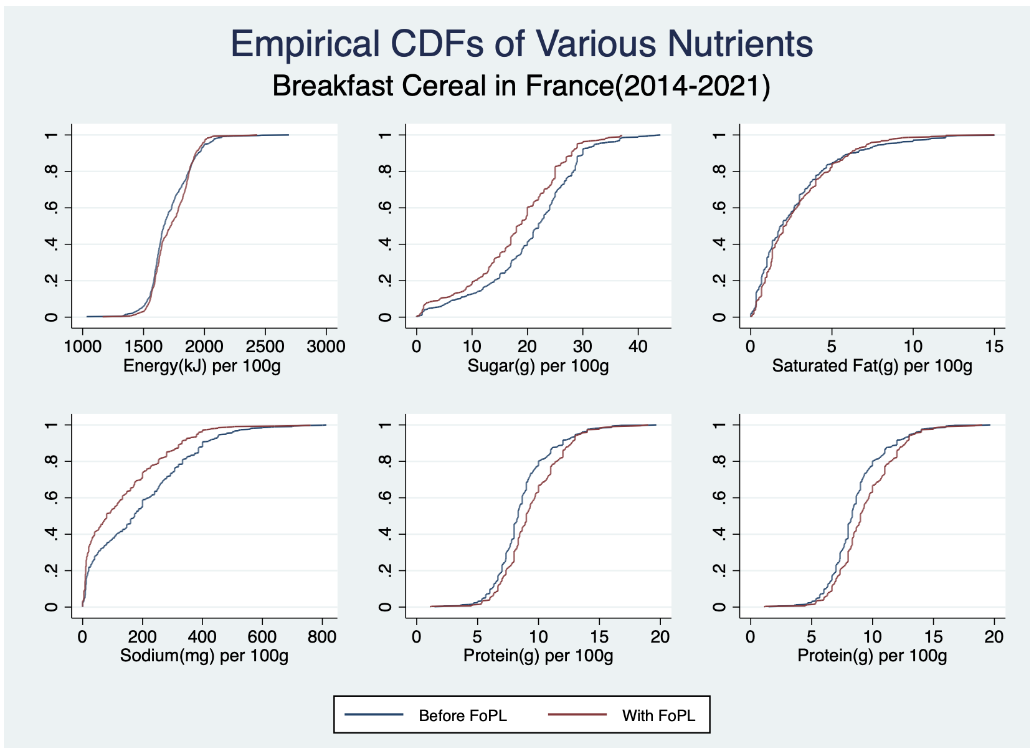

Figure 2: Empirical CDFs of Nutrients in Breakfast Cereal in France

ity has been established. In the next section, we present regression analysis to overcome this limitation.

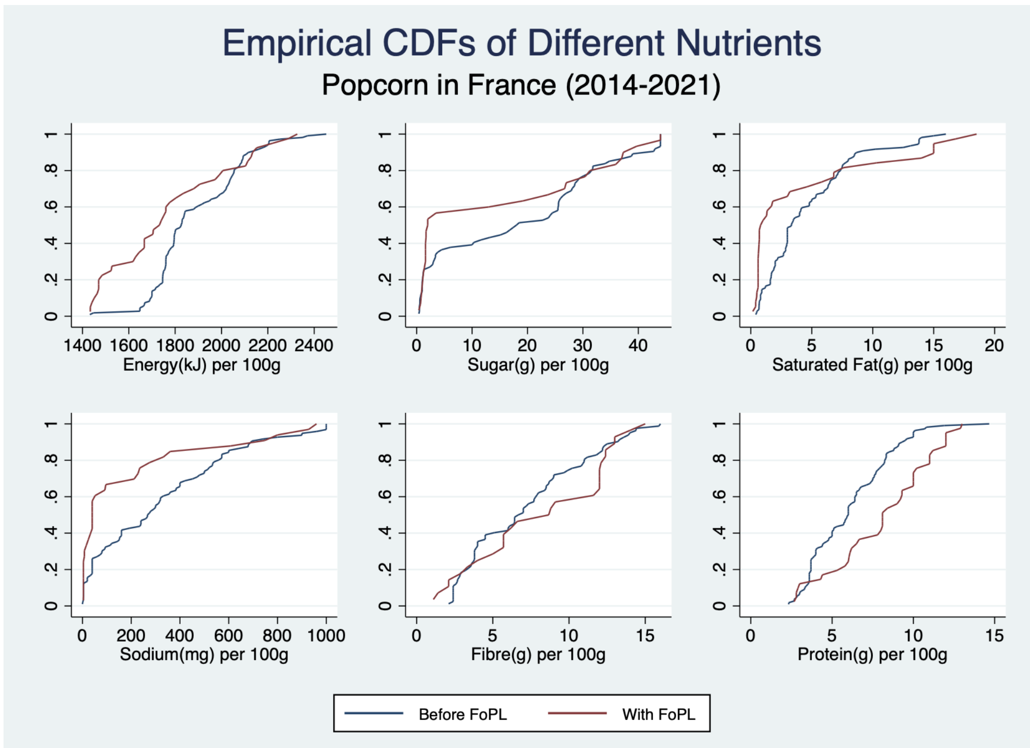

Figure 3: Empirical CDFs of Nutrients in Popcorn in France

4 Empirical Analysis

4.1 Nutri-Scores

We employ the difference-in-differences approach to identify causal effects of the Nutri-Score introduction. We run several regressions separately by product category (breakfast cereals, popcorn, potato

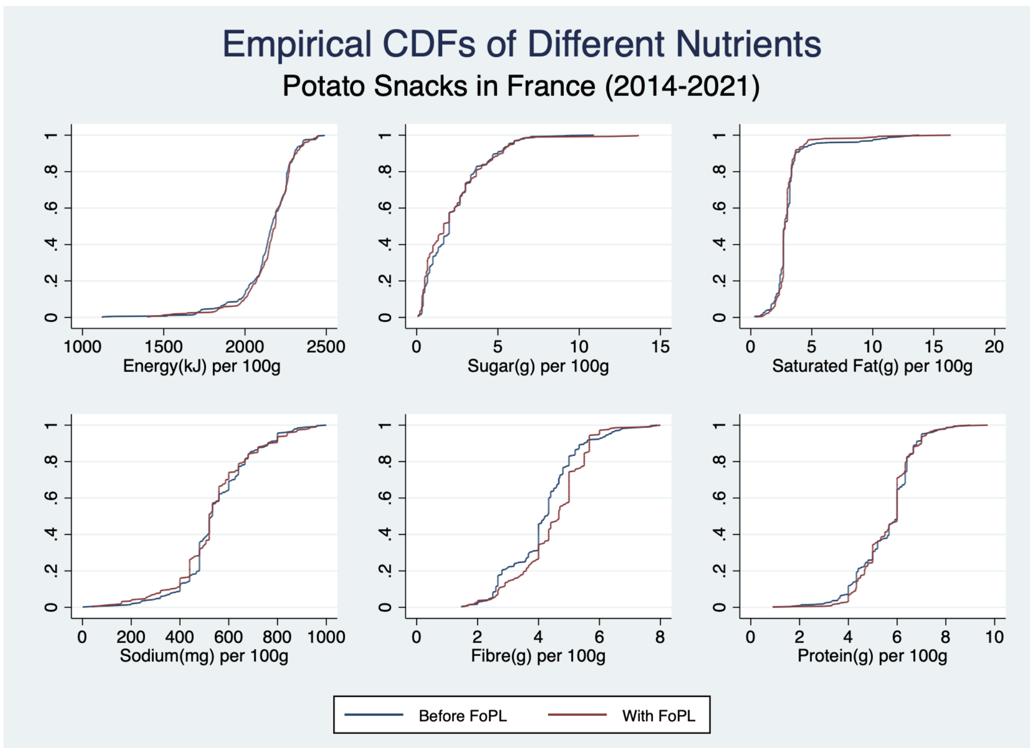

Figure 4: Empirical CDFs of Nutrients in Potato Snacks in France

snacks). We estimate the model:

$$n_{itc} = \gamma_c + \delta_t + D_{tc}\beta + \varepsilon_{itc} \quad (1)$$

where n_{itc} is the Nutri-Score for product i in country c and year t; γ_c is the country fixed effect; δ_t is the year fixed effect; D_{tc} is a dummy equal to 1 for treated observations, i.e., products brought to market in France after the introduction of Nutri-Score; and ε_{itc} captures the identically and independently distributed error terms. β, the coefficient of interest, indicates the effect of the Nutri-Score introduction.

Regression of (1) provides an estimate of the causal effect of Nutri-Score under the standard assumptions of the difference-in-differences approach. In particular, this identification strategy is only valid if treatment and control groups exhibit parallel trends prior to the policy introduction. We analyze this graphically in figure 5.

The figure shows average Nutri-Scores of new products between 2014 and 2021. Nutri-Score introduction started in late 2017. It is of course possible that food producers started bringing healthier products to the market before the introduction in preparation. If this were the case, then we may be underestimating the true effect of the Nutri-Score introduction. However, the health law adopting FoPL was enacted only in January of 2017 and Nutri-Score was announced in March 2017. Given the time required to reformulate products we do not think this effect is significant.

The pre-trends are nearly perfectly parallel in the breakfast cereal

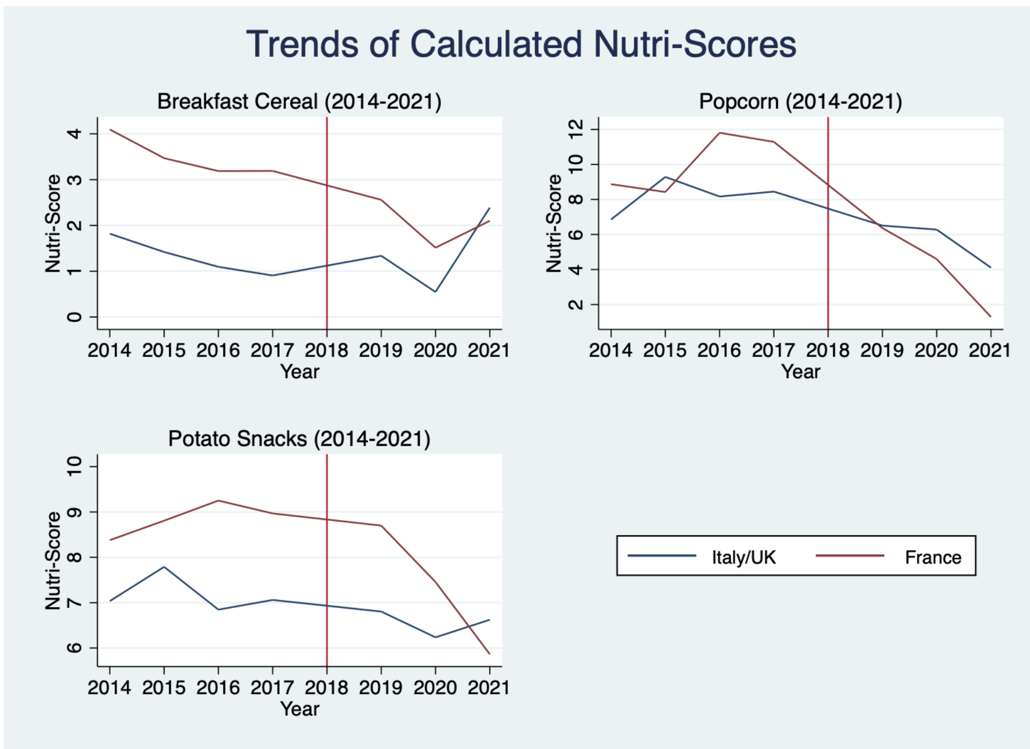

Figure 5: Trends of Calculated Nutri-Scores across Treatment and Control Markets

category. For popcorn and potato snacks, there is more movement. However, the long-term trends seem roughly parallel. Importantly, our results still hold if we were to focus exclusively on breakfast cereals, the category for which the parallel trends assumption is most credible.

We present the results of estimating of (1) in table 2. For all product categories our estimates of β are significantly negative, indicating an improvement in the nutrition profile of the average new product in the market.[4] This effect is smallest and least significant for potato snacks. It is highly significant (at the 99% confidence level) and large for the other two product categories. To put the estimates in context quantitatively, a Nutri-Score of 3 (-1) corresponds to a letter grade of C (A). In other words, the nutrition improvement in the popcorn category is nearly large enough to jump from a mediocre grade to the best grade possible. The effect is smaller, but still sizeable for breakfast cereals.

4.2 Individual Nutrients

To gain further insight into how the Nutri-Score improvement is achieved, we analyze graphically the changes to the most important nutritional variables (energy, sugar, saturated fat, sodium, fibre, and protein).

Figure 6 shows the changes for breakfast cereals. In this plot, the

[4] The significance level remains unchanged for the breakfast cereals and popcorn categories and it increases to significance at the 5% level for potato snacks if we apply sharpened q-values (Anderson, 2008) to adjust for multiple hypothesis testing.

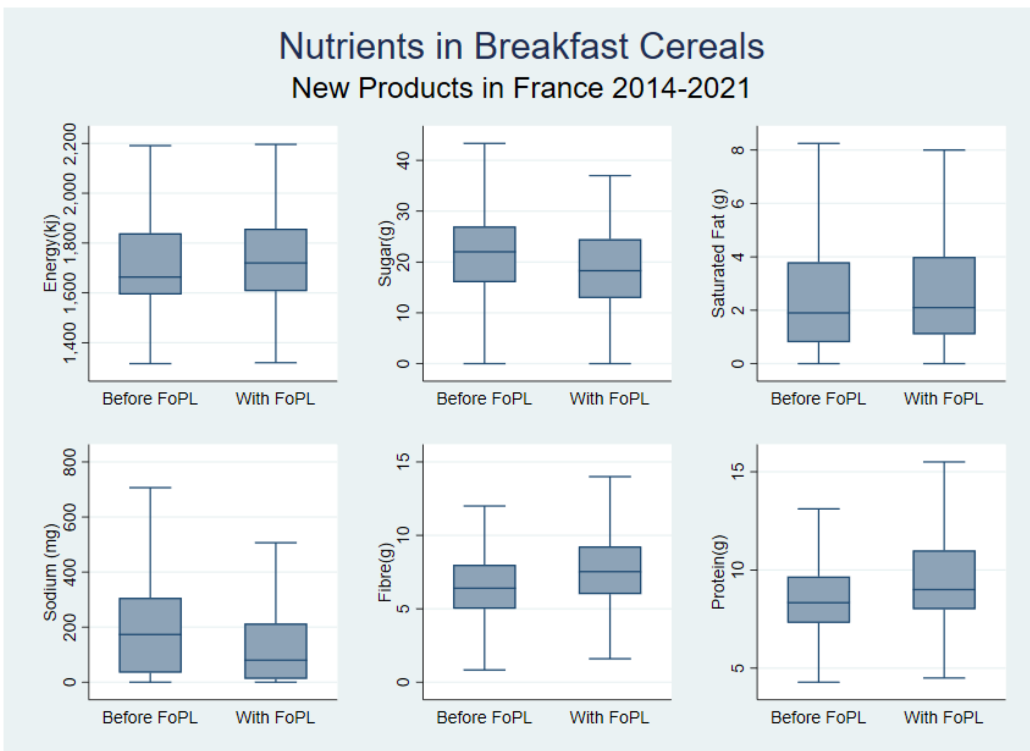

Figure 6: Changes to Different Nutrients in Breakfast Cereal

Table 2: Regression Results for Nutri-Score by Product Category

Nutri-Score	(1) Breakfast Cereal	(2) Popcorn	(3) Potato Snacks
Post Policy France(β)	-1.031***	-3.242***	-0.385*
	(0.109)	(0.096)	(0.213)
Year FE	Yes	Yes	Yes
Market FE	Yes	Yes	Yes
F	90.02	1133.26	3.28
R-squared	0.051	0.073	0.071
N	3,113	602	2,120

Note: Nutri-Score based on own calculation using Mintel GNPD data and ignoring fruit & vegetable content. Standard errors clustered at the country level and reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

box covers the 25th to 75th percentile of the distribution, the horizontal line in the box shows the median, and the vertical line represents the range between the adjacent values (Tukey, 1977). We see decreases in sugar and sodium content and increases in fibre and protein content of breakfast cereals after adoption of the Nutri-Score. These movements explain the improvement in Nutri-Scores.

The equivalent plots for popcorn and potato snacks are presented in figures 7 and 8, respectively. For popcorn, changes are most drastic as the medians of energy, sugar, saturated fat, and sodium content all fall below the previous 25th percentile and the medians of fibre and protein content increase, the latter beyond the previous 75th percentile. Given these changes, it is unsurprising that our regression found the largest effect in the popcorn category.

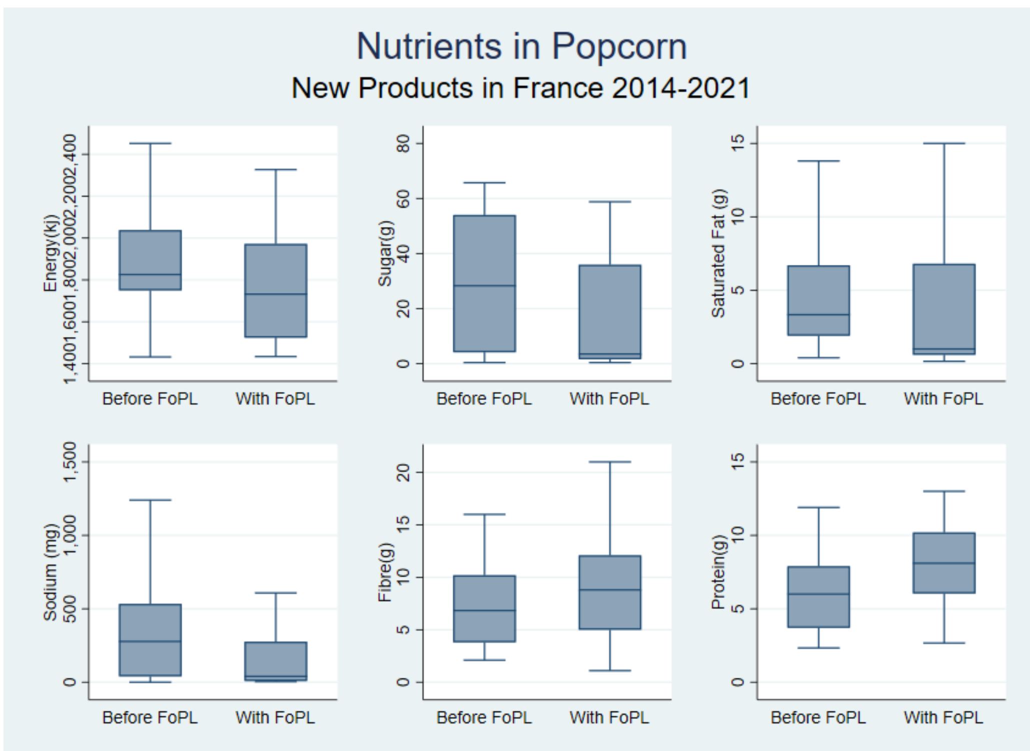

Figure 7: Changes to Different Nutrients in Popcorn

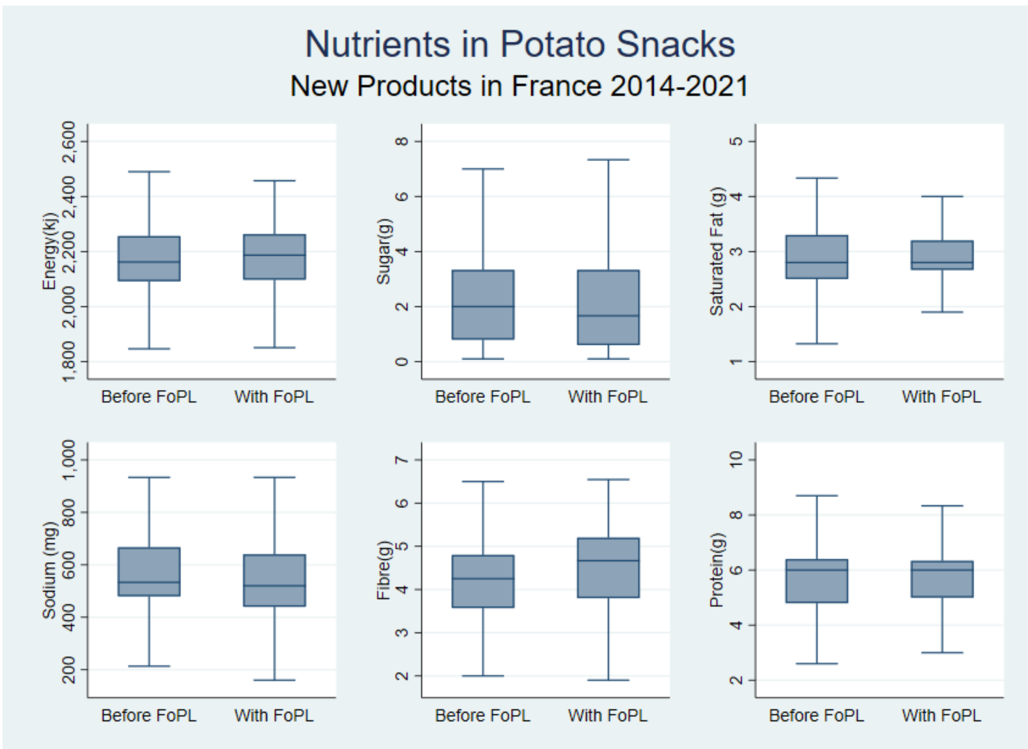

Figure 8: Changes to Different Nutrients in Potato Snacks

In comparison, changes in potato snacks are more modest. The largest one, by far, is for fibre, where 25th percentile, median, and 75th percentile all shift up substantially. However, contents of most other nutrients change little, or in a way that reduces nutrition quality. For instance, both the lower adjacent value and the 75th percentile of saturated fat content increase clearly. Overall, the graphical analysis explains well why we found only a small improvement in Nutri-Scores for potato snacks.

We refrain from a more formal analysis of individual nutrients as, contrary to the Nutri-Scores themselves, we are not comfortable with the parallel trends assumption. We suspect this difference is due to the Nutri-Score's methodology smoothing over some outliers that, given a relatively small number of observations on a per-annum-per-country basis, can be problematic when studying nutrients separately.

4.3 Bunching

A final question we are interested in is whether there is any evidence of bunching. Given the discrete nature of the Nutri-Score formula, it is plausible that manufacturers design their products to land just under the cutoff for an improved letter grade. For example, a letter score of C is achieved with a Nutri-Score between 3 and 10. Hence, if reaching a Nutri-Score of 2 proves too costly or detrimental to other product characteristics, it would seem rational to choose ingredients such that the Nutri-Score will equal 10, assuming that manufacturers find it undesirable to make the product healthier than necessary.[5] In this section we analyze whether manufacturers formulate their products to land just below the cutoff for a better letter grade.

Figure 9 shows for each product category the histograms of calculated Nutri-Scores before and after the introduction of FoPL in France. Visually, the evidence is not obvious one way or the other, with the exception of the potato snacks category which does not show any increases in Nutri-Scores below the cut-offs. However, increases in the number of products with Nutri-Scores below cutoffs are visible for breakfast cereals (at the cutoffs between A and B, and B and C, respectively) and popcorn (at the cutoffs between A and B, B and C, and C and D, respectively).

To investigate this issue more closely, we again employ regressions in the difference-in-differences design:

$$below_{itc} = \gamma_c + \delta_t + D_{tc}\beta + \varepsilon_{itc} \tag{2}$$

Here, $below_{itc}$ is a dummy equal to 1 if the calculated Nutri-Score is just below the letter grade cutoff. In other words, $below_{itc} = 1$ iff $n_{itc} \in \{-1, 2, 10, 18\}$. Table 3 shows the regression results. For the breakfast cereals and popcorn categories, we find positive results significant at the 1% level indicating that the introduction of FoPL in-

[5] This assumption seems reasonable. Otherwise, devices such as FoPL would not be necessary as manufacturers would be producing healthy foods already.

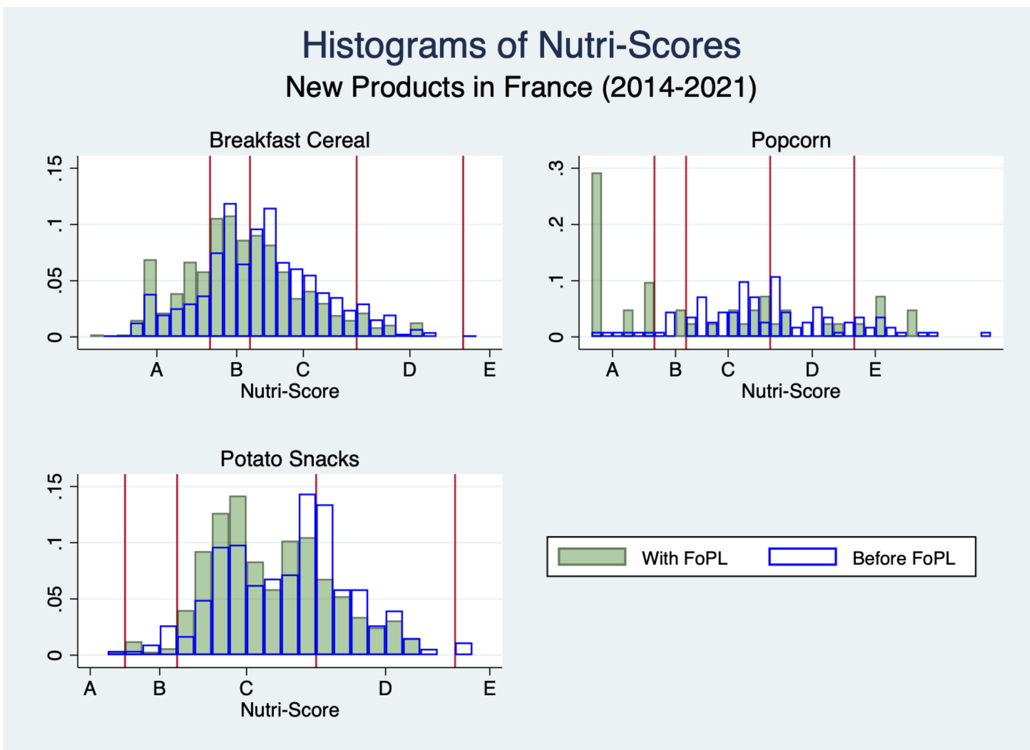

Figure 9: Histogram of Calculated Nutri-Scores by Product Category in France

creased the share of products with Nutri-Scores just below the letter grade cutoff by 3.1 or 7.4 percentage points, respectively.[6] In contrast, the estimated coefficient for potato snacks is negative and insignificant. This is not very surprising given we only found weak evidence of decreases in Nutri-Scores for this product category.

A potential concern with the analysis of bunching is that we pick up false positive based on the overall negative trend of Nutri-Scores. If the average Nutri-Score shifts down, this could possibly move enough products across the cutoff to let our regression coefficient be significant even if producers do not actively try to engineer Nutri-Scores just below the threshold. To gauge whether this is a problem here, we run falsification tests with counterfactual cutoffs. We randomly draw four cutoffs between -5 and 25 (roughly the range of Nutri-Scores we observe in our data) and repeat regression (2) with those new cutoffs. After repeating this process 1,000 times for each product category we then compare the estimates based on true cutoffs with those based on hypothetical cutoffs. If there are genuine bunching effects, we should expect the former to be larger than most of the latter.

The last row of table 3 shows the percentage of falsification tests producing a smaller coefficient estimate than the regression using the true cutoffs. For breakfast cereals, using the true cutoffs produces a coefficient larger than 84.4% of falsification trials. For popcorn the number is 95.4%, indicating that at least for this product category the

[6] The significance level remains unchanged if we apply sharpened q-values (Anderson, 2008) to adjust for multiple hypothesis testing.

introduction of FoPL did indeed lead to bunching below the Nutri-Score cutoffs. For potato snacks, we had already found no evidence of bunching, so it is unsurprising that the coefficient based on the true cutoffs is larger than only a small percentage (19.7%) of falsification runs.

Table 3: Regression Results for Bunching of Nutri-Scores by Product Category

	(1)	(2)	(3)
Dummy for Nutri-Score below Cutoff	Breakfast Cereal	Popcorn	Potato Snacks
Post Policy France(β)	0.031***	0.074***	-0.024
	(0.008)	(0.015)	(0.056)
Year FE	Yes	Yes	Yes
Market FE	Yes	Yes	Yes
R-squared	0.01		
N	3,113	602	2,120
Falsification Percentile	0.844	0.954	0.197

Note: Nutri-Score based on own calculation using Mintel GNPD data and ignoring fruit & vegetable content. Standard errors clustered at the country level and reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Falsification Percentile is the share of regressions in a randomized falsification test that produce a smaller coefficient than the one using true cutoffs.

5 Conclusion

While the effect of FoPL on consumer choices has been widely studied, its impact on manufacturer behavior is not well understood yet. Analyzing the introduction of Nutri-Score in France as an example, we investigate how characteristics of newly introduced products are altered by FoPL. We find significant reductions of calculated Nutri-Scores (improvements of overall nutrition), although the magnitude of the effect varies drastically across product categories. The way these improvements are achieved also differs across product categories and can include reductions of "bad" nutrients, e.g. sodium, or increases of "good" nutrients, e.g. fibre.

Finally, we find some evidence of bunching below the cutoffs for better Nutri-Score letter grades. This implies that our findings are at least partially driven by producer strategies intended to improve their products' competitiveness through a lower Nutri-Score, rather than by a genuine preference for producing healthier foods.

Overall, our results suggest that FoPL can be effective not only at nudging consumers towards healthier choices, but also at positively affecting producer decisions particularly on the types of products brought to market. The resulting healthier food options available to consumers could be a significant step toward important health goals related to nutrient intake.

Importantly, producer incentives need to be taken into account when designing FoPL systems. Our evidence for bunching suggests that at least some producers react strategically to FoPL. Their actions seem to indicate tension between the goals of improving Nutri-Scores and of keeping the product tasty without increasing costs too much. Thus, some producers may choose not to improve the nutrient profile of their packaged foods more than necessary for achieving the next better grade. Policy makers should take this into account when adopting an FoPL system. For example, more narrow grade levels instead of few wide ones may make strategic targeting of certain scores less appealing. The key tradeoff here is simplicity for consumers, in particular as all or most of the information going into FoPLs is already available in the back-of-package nutrition facts label and the list of ingredients. Hence, FoPLs' added value seems to stem from visually emphasizing this information and reducing the cognitive cost of processing it. A complicated FoPL system, even if it theoretically helps address concerns about producer incentives, could possibly counteract these advantages.

Policy makers would also be well-advised to remember that our findings vary by product category. Hence, the mere fact that an FoPL works well for one product category does not necessarily imply it does so for others. Presumably, there is benefit in using the same system across multiple product categories in order to avoid confusion among consumers and limit compliance cost among producers. Policy makers should strive to balance these effects against the desire to find the optimal system for each product category.

Despite these questions, our work demonstrates that FoPLs can be useful tools in the fight against nutrition-related diseases. This result is particularly welcome as one would expect them to be relatively easy to implement. Most of the information necessary for Nutri-Score, for instance, is already collected and displayed on the package. In addition, in contrast to, e.g., sin taxes, FoPLs are unlikely to cause significant deadweight loss - in fact, they may help reduce the deadweight loss associated with informational asymmetries present in the market for packaged foods.

References

Alé-Chilet, J. and Moshary, S. (2022). Beyond consumer switching: Supply responses to food packaging and advertising regulations. *Marketing Science*, 41(2):243–270.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484):1481–1495.

Araya, S., Elberg, A., Noton, C., and Schwartz, D. (2022). Identifying food labeling effects on consumer behavior. *Marketing Science*.

Ares, G., Aschemann-Witzel, J., Curutchet, M. R., Antúnez, L., Machín, L., Vidal, L., and Giménez, A. (2018). Product reformulation in the context of nutritional warning labels: Exploration of consumer preferences towards food concepts in three food categories. *Food Research International*, 107:669–674.

Barahona, N., Otero, C., and Otero, S. (2020). Equilibrium effects of food labeling policies. *Working Paper*.

Barahona, N., Otero, C., Otero, S., and Kim, J. (2022). On the design of food labeling policies. *Working Paper*.

Centers for Disease Control and Prevention (2021). Why it matters. Retrieved on 4/28/2023 from: https://www.cdc.gov/nutrition/about-nutrition/why-it-matters.html.

Das, D. (2020). Nestlé calls for mandatory nutri-score labelling across the eu. *European Supermarket Magazine*. Retrieved from: https://www.esmmagazine.com/a-brands/nestle-calls-mandatory-nutri-score-labelling-across-eu-97060.

FoodWatch (2021). WHO: EU-wide introduction of nutri-score necessary. Retrieved on 4/20/2023 from: https://www.foodwatch.org/en/news/2021/who-eu-wide-introduction-of-nutri-score-necessary/.

Franco-Arellano, B., Vanderlee, L., Ahmed, M., Oh, A., and L'Abbé, M. (2020). Influence of front-of-pack labelling and regulated nutrition claims on consumers' perceptions of product healthfulness and purchase intentions: A randomized controlled trial. *Appetite*, 149:104629.

Freedman, M. R. and Connors, R. (2010). Point-of-purchase nutrition information influences food-purchasing behaviors of college students: A pilot study. *Journal of the American Dietetic Association*, 110:1222–1226.

Hercberg, S., Touvier, M., and Salas-Salvado, J. (2021). The nutri-score nutrition label. *International Journal for Vitamin and Nutrition Research*, 92:147–157.

Ikonen, I., Sotgiu, F., Aydinli, A., and Verlegh, P. W. J. (2020). Consumer effects of front-of-package nutrition labeling: an interdisciplinary meta-analysis. *Journal of the Academy of Marketing Science*, 48:360–383.

Julia, C., Etilé, F., Hercberg, S., et al. (2018). Front-of-pack nutri-score labelling in france: an evidence-based policy. *Lancet Public Health*, 3(4):e164.

Julia, C. and Hercberg, S. (2017). Development of a new front-of-pack nutrition label in france: the five-colour nutri-score. *PANORAMA*, page 712.

Julia, C., Touvier, M., Kesse-Guyot, E., Galan, P., and Hercberg, S. (2022). Nutri-score in tug-of-war between public health and economic interests in the european union. *Nature Foods*, 3:181.

Mialon, M., Julia, C., and Hercberg, S. (2018). The policy dystopia model adapted to the food industry: the example of the nutri-score saga in france. *World Nutrition*, 9(2):109–120.

Nutri-Score Assessment Report (2021). *Assessment Report after three-year of Nutri-Score Implementation*. Santé publique France, Ministère de la Santé et de la Prévention.

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

Sacks, G., Rayner, M., and Swinburn, B. (2010). Impact of front-of-pack 'traffic-light' nutrition labelling on consumer food purchases in the uk. *Health Promotion International*, 24:1222–1226.

Southey, F. (2021). Is nutri-score working in france? the results are in... *FoodNavigator.com*. Retrieved from: https://www.foodnavigator.com/Article/2021/03/05/Is-Nutri-Score-working-in-France-The-results-are-in.

Southey, F. (2022). Labelling debate heats up as deadline looms: Does nutri-score offer 'transparency' or is it 'obsolete'? *FoodNavigator.com*. Retrieved from: https://www.foodnavigator.com/Article/2022/10/14/labelling-debate-heats-up-as-deadline-looms-does-nutri-score-offer-transparency-or-is-it-obsolete.

Tórtora, G., Machín, L., and Ares, G. (2019). Influence of nutritional warnings and other label features on consumers' choice: Results from an eye-tracking study. *Food research international*, 119:605–611.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, Mass.

Van Wezemael, L., Caputo, V., Nayga Jr, R. M., Chryssochoidis, G., and Verbeke, W. (2014). European consumer preferences for beef with nutrition and health claims: A multi-country investigation using discrete choice experiments. *Food Policy*, 44:167–176.

Vyth, E. L., Steenhuis, I. H., Roodenburg, A. J., Brug, J., and Seidell, J. C. (2010). Front-of-pack nutrition label stimulates healthier product development: a quantitative analysis. *International Journal of Behavioral Nutrition and Physical Activity*, 7(1):1–7.