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. We employ a difference-in-differences approach with Italy and the United Kingdom serving as controls. We find that products introduced or altered after the change receive a better Nutri-Score rating 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.
1 Introduction

Poor diet is a major contributor to many serious health conditions such as obesity, heart disease, and diabetes. 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, 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. 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.

Much of the prior research on FoPLs consists of studying consumer responses to FoPLs in experimental settings. This literature has 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 focuses mostly on the effects of the introduction of mandatory warning labels in Chile in 2016. These

---

1 Centers for Disease Control and Prevention, https://www.cdc.gov/nutrition/about-nutrition/why-it-matters.html
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.

A few studies have dealt with producer reactions to FoPLs. 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, working with their reported numbers we are often unable to replicate their findings of significance, 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 in this contribution 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.

We add to this literature by analyzing experiences from a different country and with a different label – Nutri-Score in France – to those from warning labels in Chile. Not only could reactions vary due to the different cultural and economic situations of the countries and due to the different design of the labels; identification of the effect of FoPLs is also cleaner in our study because no advertising ban coincided with the introduction of Nutri-Score in France. Our results are largely consistent with those of prior authors: We find improvements in the nutritional profile of products. 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.

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.

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.

FoPLs were first proposed in France in 2014 and, after a consultation process with multiple stakeholders, Nutri-Score was adopted on a voluntary basis on October 31st, 2017 (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.²

The initial rollout of Nutri-Score included only six manufactur-

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

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. The Mintel Global New Products Database provides photos of the product packages, so that we are often able to manually correct the erroneous information we encounter.

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 af-

---

3 Auchan, Intermarché, Leclerc, Fleury Michon, Danone, and McCain
ter 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. However, 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.\(^4\)

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

\(^4\) Guidance published by Santé Publique France specifically excludes potatoes and popcorn from consideration as vegetables. See https://www.santepubliquefrance.fr/media/files/02-determinants-de-sante/nutrition-et-activite-physique/nutri-score/qr-scientifique-technique-en
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Breakfast Cereals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>2.89</td>
<td>4.61</td>
<td>-9.00</td>
<td>20.00</td>
<td>1,168</td>
</tr>
<tr>
<td>Italy</td>
<td>1.87</td>
<td>4.54</td>
<td>-7.00</td>
<td>16.00</td>
<td>447</td>
</tr>
<tr>
<td>UK</td>
<td>1.02</td>
<td>4.53</td>
<td>-7.00</td>
<td>19.00</td>
<td>1,498</td>
</tr>
<tr>
<td><strong>Popcorn</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>8.76</td>
<td>8.24</td>
<td>-6.00</td>
<td>32.00</td>
<td>152</td>
</tr>
<tr>
<td>Italy</td>
<td>4.18</td>
<td>7.22</td>
<td>-7.00</td>
<td>16.00</td>
<td>60</td>
</tr>
<tr>
<td>UK</td>
<td>8.29</td>
<td>7.27</td>
<td>-7.00</td>
<td>29.00</td>
<td>390</td>
</tr>
<tr>
<td><strong>Potato Snacks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>8.52</td>
<td>3.65</td>
<td>-1.00</td>
<td>20.00</td>
<td>853</td>
</tr>
<tr>
<td>Italy</td>
<td>7.95</td>
<td>4.94</td>
<td>-2.00</td>
<td>24.00</td>
<td>393</td>
</tr>
<tr>
<td>UK</td>
<td>6.51</td>
<td>3.50</td>
<td>-3.00</td>
<td>21.00</td>
<td>874</td>
</tr>
</tbody>
</table>

*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 $\leq -1$ for A, $0 \leq$ score $\leq 2$ for B, $3 \leq$ score $\leq 10$ for C, $11 \leq$ score $\leq 18$ for D, and $19 \leq$ score for E.
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 causality has been established. In the next section, we present regression analysis to overcome this limitation.
Figure 1: Empirical CDFs of Nutri-Score in France
Figure 2: Empirical CDFs of Nutrients in Breakfast Cereal in France
Figure 3: Empirical CDFs of Nutrients in Popcorn in France
Empirical CDFs of Different Nutrients
Potato Snacks in France (2014-2021)

Figure 4: Empirical CDFs of Nutrients in Potato Snacks 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 snacks). We estimate the model:

\[ n_{itc} = \gamma_c + \delta_t + D_{tc}\beta + \epsilon_{itc} \]  

where \( n_{itc} \) is the Nutri-Score for product \( i \) in country \( c \) and year \( t \); \( \gamma_c \) is the country fixed effect; \( \delta_t \) 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 \( \epsilon_{itc} \) captures the identically and independently distributed error terms. \( \beta \), 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
Figure 5: Trends of Calculated Nutri-Scores across Treatment and Control Markets
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 category. For popcorn and potato snacks, there is more movement. However, the long-term trends seem roughly parallel.

We present the results of estimating of (1) in table 2. For all product categories our estimates of $\beta$ are significantly negative, indicating an improvement in the nutrition profile of the average new product in the market.\footnote{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.} 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 box covers the 25th to 75th percentile of the distribution, the hori-
Table 2: Regression Results for Nutri-Score by Product Category

<table>
<thead>
<tr>
<th>Nutri-Score</th>
<th>(1) Breakfast Cereal</th>
<th>(2) Popcorn</th>
<th>(3) Potato Snacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post Policy France(β)</td>
<td>-1.031*** (0.109)</td>
<td>-3.242*** (0.096)</td>
<td>-0.385* (0.213)</td>
</tr>
<tr>
<td>Year FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Market FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>3,113</td>
<td>602</td>
<td>2,120</td>
</tr>
</tbody>
</table>

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.

Figure 6: Changes to Different Nutrients in Breakfast Cereal
zontal 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 the 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. 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:

\[ \text{below}_{itc} = \gamma_c + \delta_t + D_{tc} \beta + \epsilon_{itc} \]  

Here, \( \text{below}_{itc} \) is a dummy equal to 1 if the calculated Nutri-Score is just below the letter grade cutoff. In other words, \( \text{below}_{itc} = 1 \) 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 increased the share of products with Nutri-Scores just below the letter grade cutoff by 3.1 or 7.4 percentage points, respectively.\(^7\) In contrast,\(^6\)

---

\(^6\) This assumption seems reasonable. Otherwise, devices such as FoPL would not be necessary as manufacturers would be producing healthy foods already.

\(^7\) The significance level remains unchanged if we apply sharpened q-values (Anderson, 2008) to adjust for multiple hypothesis testing.
Figure 9: Histogram of Calculated Nutri-Scores by Product Category in France
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 9 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 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

<table>
<thead>
<tr>
<th>Dummy for Nutri-Score below Cutoff</th>
<th>(1) Breakfast Cereal</th>
<th>(2) Popcorn</th>
<th>(3) Potato Snacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post Policy France(β)</td>
<td>0.031***</td>
<td>0.074***</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.015)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>Year FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Market FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>3,113</td>
<td>602</td>
<td>2,120</td>
</tr>
<tr>
<td>Falsification Percentile</td>
<td>0.844</td>
<td>0.954</td>
<td>0.197</td>
</tr>
</tbody>
</table>

*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 the 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. Importantly, producer incentives need to be taken into account when designing FoPL systems. Our evidence for bunching suggests that systems like Nutri-Score, which uses multiple letter grades, or systems providing a continuous score may be more successful at improving nutrient contents than binary systems.

However, policy makers would 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. This needs to be traded off against simplicity for consumers.
References


