Nutritional gain and economic cost of 5 front-of-pack nutritional labels

Paolo Crosetto[a], Anne Lacroix[a], Laurent Muller[a], Bernard Ruffieux[a]

[a]INRA,GAEL, 3800 Grenoble, France

Abstract

We study in a laboratory framed field experiment the impact of five Front of Pack (FOP) labels on the nutritional value and cost of a daily consumption basket. We employ a difference in difference experimental design, between subjects, to cleanly identify the impact of FOP labels. 691 subjects issued of the general population shop within a catalog of 290 products twice: once, without and a second unannounced time with labels. Purchases are real. We test five different labels and compare result against a benchmark treatment in which subjects shop twice with no labels. Labels include the existing Multiple Traffic Lights, Reference Intakes and Health Star Rating, and two newly proposed desings: NutriScore, a 5-color mono-dimensional synthetic label, and SENS, a frequency-based recommendation label. We measure nutritional quality in terms of the FSA score. All labels but Reference Intakes significantly improve nutritional quality. NutriScore is significantly more effective than all other labels, followed by the Australian Health Star and Multiple Traffic Lights. Nutritional improvement come at an economic cost, as the average cost of 2000Kcal increases for all labels. Nonetheless, we show that the extra cost for a unit nutritional improvement is borne mainly by richer households. Behaviorally, change is concentrated in the extremal categories of each label, and easier to understand labels have a higher impact and crowd out more successfully other information cues like ingredients lists and nutritional tables.

Keywords:
Nutrition labels, Framed field experiment, Front of Pack
JEL: C93, Q18

1. Introduction

Consumers face enormous choice sets. They make hundreds of food choices every day. A medium-sized supermarket stocks tens of thousands of food products. For each of these products, consumers can access a variety of information, some of it compulsory (price, size, ingredients, nutritional facts), some put forward by the producer (origin and benefits of the products, and a series of environmental impact, organic, fair-trade claims). An enormous amount of information is competing for the limited time, attention and cognitive resources of consumers.

Confronted to these large search costs, consumers often rely on habit and simpler heuristics. Yet, the aggregation of those many food choices determines consumers' diets. Diets, in turn, have a large impact on long-term health conditions. As a consequence, biased and sometimes mindless food choices have the potential to impact public health and treatment costs.

Front-of-Pack nutritional labels (FOPL) are a policy tool conceived for a context of choice overload (ADD REF) with the aim of facilitating consumer choice towards healthier diets. FOPL do not usually give additional information with respect to that found on the back of pack in nutritional tables and ingredients lists. Rather, they give an aggregated, summarized and focal version of the same information. They are designed to be simple and to stand out among the mass of competing signals. A clear message standing out on the front of pack has a higher chance of being seen, parsed and used by consumers.

Different FOPL formats have been shown to have differentiated impacts on consumer choices. Differences in formats may reflect competing interests. From a public policy perspective, the best FOPL is the one generating the healthiest diet changes.

Several governments (Australia, 2013a, FSA, 2013, Norden, 2010) and the EU (Council of the European Union, 2011) have hence endorsed FOPL as a policy tool. Recently, France entered the debate. Article 14 of the French Law of 26 January 2016 on the modernization of the health system recommends food producers to adopt, on a voluntary basis, a uniform front-of-pack nutritional label. Under the aegis of the health authorities, representatives of manufacturers, retailers, consumers and scientists were consulted and four candidate systems were selected: NutriCouleurs (a French version of the UK's Traffic Lights), NutriRepère (an updated version of the UK's Reference Intakes), NutriScore and SENS.

In this paper we assess in a controlled laboratory experiment the impact of five different FOPL on the nutritional value of food choices. They are the four aforementioned labels that were selected by the French health authorities plus NutriMark. Three of them are existing and widely used (NutriCouleurs corresponds to the UK's Traffic Lights, NutriMark to the Australia and New Zealand's Health Star Ratings and NutriRepère is an updated version of the UK's Reference Intakes) and two are newly proposed (NutriScore and SENS). 691 subjects perform real shopping tasks on a selected catalog of 290 food products. We use a difference-in-difference design, whereby subjects shop twice, once without and a second time, unannounced, with FOPL. By varying the label between subjects we can cleanly identify, for each subject, its effect on choices when compared to the reference shopping basket.

Several other studies have measured the relative performance of different FOPL (several reviews of the literature exist, for instance Hersey et al., 2013, Vyth et al., 2012, Cecchini and Warin, 2016, Grunert and Wills, 2007). Studies follow by and large one of four methods: surveys, randomized controlled trials, choice experiments, and willingness-to-pay studies.

Surveys set up hypothetical choice scenarios, and systematically vary the use of labels to identify effects. Subjects' task is usually to identify the healthier product among a set of two or three. These studies allow researchers to assess the understanding and parsing of FOPLs. On the other hand, they do not allow to see how this understanding interacts with preferences; moreover, by relying on scenario choices they suffer from hypothetical bias and do not take into account the role of price in the choice.

Choice experiments improve on surveys by giving the subjects dozens of choices, systematically varying the attributes of the products – price, quality, label – and analyzing the data in a structural random utility model. They allow the researcher to estimate the marginal role of each attribute in a context in which prices are taken into account. On the other hand, choice experiment usually rely on hypothetical choices and on strong behavioral assumptions.

Natural field experiments apply FOPL in grocery shops following a randomized procedure. Subjects might or might not be informed that they are part of an experiment, and their response to FOPLs is observed in a real setting and according to real preferences and budget constraints. These studies have a high degree of external validity. On the other hand, data are extremely noisy and the experimenters do not have tight control on all possible confounds (promotions, shops opening hours, . . .). As a result effects are rarely found, and when found, small and difficult to identify.

Willingness-to-pay studies use incentivized elicitation methods to assess the perceived value of different food items, with and without labels. They indirectly assess, through the value that a subject assigns to a labeled product, the importance the subject attributes to the label's message. Usually run in the lab, they feature strong control on confounding factor but forfeit most of the external validity. In addition, these experiments usually involve a small number of products.

In this paper, we use a randomized controlled trial that allows us to observe real purchasing choices from a representative sample over a large set of products, as in natural field experiments, but in a controlled laboratory setting in which we minimize noise, as in choice experiment or willingness-to-pay studies. Other studies have taken this route (among others, Cecchini and Warin, 2016, Ducrot et al., 2016, Waterlander et al., 2014, Neal et al., 2017). Nonetheless, to the best of our knowledge we are the only study providing clean, incentivized laboratory evidence on the effect of FOPLs. Our design allows us to both assess the nutritional quality of the whole shopping with and without competing labels, and to observe in detail the behavioral sources of this nutritional change.

We find that all labels have a significant positive impact on the nutritional quality of the shopping. On average, labels improve the FSA score – nutritional index from the Food Standards Agency ((Rayner, Scarborough, and Lobstein 2009) – by 1.56 points, with respect to the baseline shopping basket with no labels. Color-coded labels providing aggregate nutritional information generate nutritionally healthier choices than analytical labels using numbers. This is in line with most previous research. In particular, the NutriScore label (aggregate, color) outperforms all others, followed by NutriMark (aggregate, black and white). We further find that the alternative approach of giving the consumer information about the suggested frequency of consumption (SENS label, aggreagate, color) performs worse than more direct labels giving nutritional information.

Nutritional improvements come at an economic cost, though, as the average cost of 2000Kcal increases for all labels (but NutriMark). Nonetheless, we show that the extra cost for a unit nutritional improvement is borne mainly by richer households. This is in stark contrast with policies based on taxes and subsidies, that have been shown to have regressive effects (Muller et al., 2017).

Our experiment yields a rich data set and allows us to shed some light on the behavioral drivers of the above nutritional outcomes. For aggregate labels (i.e. rating the nutritional quality of the product as a whole), subjects tend to focus on the extreme values, treating information on nuanced 5- or 10-level scales as a binary (good-bad) or ternary (good-average-bad) information. These tendency result in baskets that contain more products that are labeled 'green' or 'five star' and less products that are labeled 'red' or 'zero stars' but do not differ much for all the in-between categories. For analytic labels (i.e., assessing the amount of nutrients), subjects disproportionally focus on fat. Labels differ in their ability to crowd out other information: some labels are used as nearly perfect substitutes to nutritional tables and ingredients lists, while others are not. We interpret this as an indicator of the trust subject give to each label. The ability of labels to crowd out other information correlates with its overall nutritional performance.

Subjects seem to comply more with labels that they trust more and focus more on extreme values. Their choices are more impacted if the label provides aggregate information and/or uses color scales. That is, labels that stand out and are cognitively easy to parse are trusted more and have a stronger effect than analytical labels that give plenty of information but no focal, clear nutritional judgment.

2. Method

Subjects were asked to shop for two days for their household, following their usual shopping habits. A paper catalog of 290 products, divided in 39 categories, was distributed to each subject. The catalog displayed for each product a full color picture, price, weight (g) or volume (l), price per kg or per liter and a bar code. Prices corresponded to those recorded before the experimental campaign in a local supermarket. Using a bar-code reader, subjects could display on their screen the product of their choice in a custom on-line e-shopping environment. They could thereby access, for each product, a list of ingredients (with outlined allergens) and a nutritional table.

Subjects were asked to shop twice; once with a benchmark catalog without nutritional labels and a second, previously unannounced, time with a new catalog, strictly identical to the previous one, but in which nutrition labels were added. In the second catalog, all products that are legally subject to labeling were labeled, i.e. all products except fresh fruits and vegetables, fresh packaged meat, and eggs.

We ran 6 between-subjects treatments: one for each of 5 nutritional labels and a benchmark neutral treatment in which no label was added and subjects shopped twice with the same catalog.

The experiment was incentive compatible. Subjects were informed that they would have to buy a subset of one of their two shopping baskets. The payoff-relevant basket would be randomly and publicly drawn at the end of the experiment. The subset of products actually put up for sale was unknown *ex-ante*. In a separate room, we had stored about a quarter of all the catalog products. The intersection of the items selected by the subjects and what we had in store was then sold, at catalog prices, to the subjects at the end of the session.

3. Measures

Our experimental design (already put to use in tis general form in Muller et al., 2017, Muller and Ruffieux, 2012) allows us to measure behavior twice. The first, unlabeled basket allows us to set a benchmark for the shopping behavior of each subject. The second basket allows us to assess, within subjects, changes with respect to the baseline basket. The comparison of individual changes across treatments allows us to cleanly assess, by exploiting differences in differences, the effect of labels by controlling for heterogeneity of individual preferences.

The main measure of interest is the aggregation per treatment of the individual change in the nutritional score between basket 2 (labeled) and basket 1 (unlabeled). We adopt as our nutritional measure the Nutrient Profiling Model developed by the UK Food and Standard Agency (Rayner et al., 2009, , henceforth FSA score). This score is computed for each product by assigning negative points for salt, saturated fatty acids, calories, and sugar, and positive points for fiber, fruit & vegetable content and proteins. The score ranges from -15 to 35, with lower numbers indicating better overall nutritional quality.

We compute the aggregate nutritional score for each shopping basket, by adding the FSA score of each item and normalizing by 100 Kilocalories. That is, for each subject i, for each basket $j \in [1, 2]$, we compute

$$FSA_{ij} = \frac{\sum_p Kcal_{pij} \cdot FSA_{pij}}{\sum_p Kcal_{pij}},$$

in which the index pij denotes each product p in caddy j for subject i.

This measure gives us a single, continuous variable to assess the nutritional quality of the whole shopping basket. Nonetheless, it relies on two assumptions: that the FSA score correctly assess the nutritional value of a shopping basket, and that a normalization by calories is not distortive. Normalization by energy content has been proven to be least distorsive. For robustness, though, we ran our analysis also using SAIN/LIM, an alternative nutritional system (Tharrey et al., 2017) and normalizing by weight.

We estimate treatment effects using a difference-in-difference regression analysis. We let the treatment and basket variables interact, and add a series of subject-specific variables to control for the socio-demographics of the sample. Our main estimation uses data from both caddies and all six treatments, and takes the following form:

$$FSA_{it} = \beta_0 + \beta_1 Cart2 + \delta Treatment \times Cart + \gamma Controls,$$

in which the benchmark treatment with no labels serves as reference category, β_0 identifies the average nutritional value of the first basket, β_1 the impact of the benchmark treatment, δ is a vector of 5×2 coefficients identifying the difference with respect to the first benchmark basket of the first and second basket for each of the five labels (NS, NR, NC, NM, SENS), and γ is a vector of subject-specific coefficients.

We supplement the analysis of nutritional impact with economic considerations. First, in line with public policy concerns for lower income households, we separately run regressions for three different income groups, declaring low ($<$ 2000€/month family disposable income), average (between 2000€ and 3000€) and high ($>$ 3000€). Second, we run an extensive analysis of results normalized by price. A price normalization allows us to directly assess the impact on overall spending of the adoption of each label, and to see whether the nutritional improvement comes at the cost of a higher price.

In addition to these analytical measures based on the FSA score, we computed a number of behavioral measures and used data from questionnaires. In particular, we recorded the number of items in each basket, the number of product entries and exits from basket 1 to 2, the number of clicks on the ingredient lists and nutritional tables. We have use label-specific behavioral variables keeping track of the qualitative change between baskets. Finally, we asked subjects to fill in two questionnaires, one aimed at assessing their understanding of the labels, and the other including socio-demographic characteristics.

4. Experimental details

4.1. Products

Subjects could shop in a printed color catalog of 290 food items grouped into 39 categories[1] While not being perfectly representative of shopping in a real supermarket, where the number of available products numbers in the thousands, the catalog covered all the needs of an average household, and included fresh, canned, packaged and frozen food, fruits and vegetables, snacks and mixed prepared dishes. The overall coherence and representativeness of the catalogue was respectively verified by nutritionists and using actual consumption data (Kantar World Panel). All products were currently on sale in local supermarkets at the time of the experiment. The prices used in the catalog were collected in a specific supermarket the week before the start of the experimental campaign, in October 2016. The catalog displayed a picture of the front-of-pack of the product, its name, price, weight or volume, and the price per kilogram (or liter).

4.2. Subjects

The study involved 691 subjects over 42 sessions (7 per treatment). Subjects were recruited among the general population of the Grenoble metropolitan area by a professional recruitment agency. Grenoble and its suburbs, located in the Alps in south-eastern France, have about four hundred thousand

[1]For screenshots of the catalog, see AppendixD. The full catalog for the benchmark treatment is available here

inhabitants. Recruited subjects were in charge of the grocery shopping for their household and regular supermarket customers. The sample was stratified by household disposable income: one third of subjects with less than two thousand Euros per month, one third between two and three thousand, and another third with more than three thousand. Following standard procedures of randomized controlled trials, individuals from each income group were randomly allocated to either one of the 5 label treatments or the benchmark treatment. Summary statistics for our sample are provided in Table A.11 in AppendixA. Overall, one in five subject is a man, the age and income structure loosely reflects those of the Grenoble Metro Area. Randomization checks do not reveal significant differences across treatments, except for the distribution of professional status.

4.3. *Tested Labels*

We tested 5 nutritional labels (see Table 1 for an overview). Omitting binary recommendation labels as the keyhole (Norden, 2010) or the Heart Foundation Tick (Australia, 1989), these labels span the whole space of those in use today in different countries, and follow one or both of the main approaches in nutritional labels: providing detailed analytic information about a set of key nutrients, or displaying an aggregate assessment of the food item. The labels use colors, numbers and letters to convey their meaning; they rely on absolute, relative, or frequency assessment; and they are based upon different algorithms to translate bare nutritional values into marks and ranks.

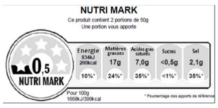
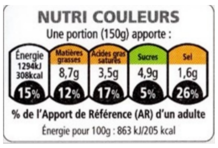
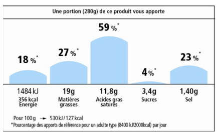
Label	Name	A.K.A.	Information	Color	Base	Reference	Use
	NutriScore (NS)	5Couleurs	Aggregate	Yes	FSA score	Julia et al. (2014)	(since 2017) France
	NutriMark (NM)	Health Star Rating	Aggregate and analytic	No	HSR score	Australia (2013a,b)	Australia, New Zealand, Leclerc retailer
	NutriCouleur (NC)	Multiple Traffic Light	Analytic	Yes	nut.values	FSA (2013)	United Kingdom, Ireland, Spain
	NutriRepère (NR)	Guideline Daily Amount, Reference Intake	Analytic	No	nut.values	Rayner et al. (2004)	EU indsutry standard, US facts up front, . . .
	SENS	–	Aggregate	Yes	LIM score	Tharrey et al. (2017)	–

Table 1: Nutritional labels tested in the experiment

NutriScore (NS) was developed and validated by an independent research team (Equipe de Recherche en Epidémiologie Nutritionnelle Julia and Hercberg, 2017). NS is based on the nutrient profiling system developed by the UK Food Standards Agency (Rayner et al., 2009, FSA score,) later adapted to the French context de la Santé Publique (2015). Its final graphical format was developed on a dedicated study (Nugier et al., 2016). NS is an aggregate, color-coded label, similar to the energy efficiency labels used in the home appliance sector, which gives coarse but salient information, in the form of a letter (from A to E) color-coded from green to red. Studies have shown it to be effective in evaluating the healthiness of foods (Ducrot et al., 2015) and to impact purchasing intentions in an online supermarket (Ducrot et al., 2016, Julia et al., 2016), in a controlled laboratory setting (Crosetto et al., 2016b). The Nutri-Score was elected as the official French front-of-pack labeling system in October, 2017 (République, 2017).

NutriMark (NM) was adapted by the French retailer Leclerc based on the Australian Government Health Star Rating system (Cooper et al., 2017, Australia, 2013a,b). NM displays both aggregate information, in the form of stars, ranging from 0.5 to 5, and nutrient-specific information in the form used by Reference Intake labels. NNM was tested on Leclerc's online shopping platform in the fall of 2016, involving 3,000 Leclerc brand products at 84 collection points in France. Results show a slight improvement in the average nutritional quality of purchases, but only for middle-class and under-30s.

NutriCouleur (NC) was developed under the original name Multiple Traffic Light system by the UK Food and Standard Agency (FSA, 2013). Its introduction in France was supported by Nestlé. NC presents analytical information for energy and four key nutrients (fat, saturated fatty acids, sugar and salt) in three different ways: as a percentage contribution to the daily reference intake, in absolute amount per serving (in grams) and in color on three levels (red, amber, green). Several studies have investigated the impact of Multiple Traffic Lights on choices (Aschemann-Witzel et al., 2013, Crosetto et al., 2016a, Julia et al., 2016, Bialkova and van Trijp, 2010, Kelly et al., 2009), showing overall greater effectiveness with respect to labels that do not use colors.

NutriRepère (NR) was proposed by over 500 food companies in France. It is based on the Reference Intake label (RI).[2] RI has been adopted by the Australian food and beverage industry in 2006 (as Daily Intake Guide), by the European Union in 2009 as an industry standard and has been introduced in the US following Michelle Obama's initiative in 2012 (Facts Up Front). NR presents analytical nutritional information for energy and four key nutrients (fat, saturated fatty acids, sugar and salt). This information is displayed in three different ways: in percentage, absolute value, and by means of light blue histograms.

SENS was developed by nutritionist Nicole Darmon and her group (Tharrey et al., 2017) based on the previously established SAIN/LIM nutritional profiling (Darmon et al., 2009). SENS is backed by the organisation of French distributors, grouping 50 supermarket chains. SENS presents nutritional information as a color-coded recommendation of eating frequency, using a red-blue-orange-purple palette. That is, after assessing the food nutritional quality, SENS translates this index into a frequency, and tells the consumer if a food item might be eaten anytime, often, from time to time, or rarely.

Overall, the labels vary across three dimensions that have been shown to be crucial by previous research (Muller and Ruffieux, 2012, Crosetto et al., 2016a, Muller and Prevost, 2016, Drichoutis et al., 2008, Shogren, 2011): *(i)* whether the information is aggregated (NS, SENS, in part NM) or analytical (NR, NC, in part NL); *(ii)* whether the logo uses color (NR, NC, SENS) or not (NM, NR); and ()whether the recommendation focuses on nutritional content (NS, NM, NC, NR) or on consumption frequency (SENS). More generally, labels trade off saliency with detail, with NS and SENS choosing to deliver a salient but coarse message with no readily available reference to nutritional values, and NR and NC choosing to give more detailed information in a less salient package. NM includes both approaches in a combined design, but forfeits the use of color.

4.4. Laboratory procedures

Subjects were invited for sessions lasting approximately 1.5 hours and received 32€as show-up fee. Their task was to shop to cater to the needs of their family for two days, but were otherwise not directed in their choices. This frame was given to reach some uniformity in the task that the subjects faced, but it was not enforced. If subjects asked what "two days's of consumption" meant, they were told that each must decide according to their household's taste and needs. Subjects were allowed to shop more, or less, or not at all.

Instructions were then showed on each subject's screen, as well as projected overhead, and read aloud by the experimenter[3]. Questions were asked and answered publicly all along the reading of the instructions. The English translation of the original French instructions is available in AppendixC.

[2] Reference Intake have been previously known as Guideline Daily Amount (Rayner et al., 2004). Although the principles behind GDA and RI are the same, the major difference is that GDA existed for men, women and children; there is only one set of *Reference Intakes* for an average adult.

[3] Due to the size of the experimental campaign, four different experimenters were involved. Results are robust to controlling for the identity of the experimenter.

After all instructions had been given, including a screenshot-based demonstration of the software interface[4], subjects were asked to do their shopping. Subjects were not given a time limit, and each could shop at his or her own rhythm. When all subjects had finished their shopping, they were asked to shop again in an unannounced second shopping exercise.

For all treatments but the Benchmark, after the first shopping and before the second subjects were given a one-page explanation of the nutritional label that they would face in the second shopping (see AppendixC). These explanation were based on the flyers distributed in the 60 supermarkets of a natural field experiment comparing the same labels performed in December 2016. Subjects received a paper copy of these one-page explanatory sheets; the same was projected on a screen and its contents read aloud. Any question was replied before proceeding with the second shopping period.

After each subject submitted his or her choices for the second shopping period, one of the two was randomly selected as binding using a physical urn for increased transparency. Then subjects were exposed to a socio-demographic questionnaire and, for all treatments but the Baseline, to a qualitative survey asking them to rate over several dimensions the nutritional label they had been exposed to. Subjects left then the room individually, and bought the items they had chosen in the binding basket in a separate room.

5. Results

5.1. Nutritional results

The benchmark treatment

In the benchmark treatment, subjects shopped twice under the same conditions. This treatment allows us to assess the degree of variability of diets across repetitions. Indeed, a repetition *ceteris paribus* does not necessarily have to generate the same choices. First, as basket 1 was not recalled to the screen when basket 2 was created, subjects had to reconstruct it from memory. Second, subjects might wish to change some items because of a taste for variety. Third, a basket does not reflect all eating habits, but represents only a limited sample so it is natural to expect a second sample to differ. Finally, subjects could find repetition boring and make changes to escape boredom. Although we expect variability in all treatments, we do not expect any systematic nutritional effect in the benchmark treatment.

This is indeed what we observe. The average FSA score for basket 1 is 5.22 (s.d. 3.01), and 5.34 (2.94) for basket 2. The difference is not significant (Wilcoxon Sgned Rank test, p-value 0.77). Figure 2 reports the cumulative distribution function of changes in the FSA score between basket 2 and 1 (negative changes mean healthier diets), for all treatments. The distribution for the Benchmark treatment (light grey) is roughly symmetric around its center. Despite a large variability in nutritional quality between baskets 1 and 2 within the benchmark treatment, there is no net effect on overall nutritional quality, as individual deteriorations and improvements cancel each other out.

The label treatments

The nutritional quality of basket 2 improved across the board with respect ot basket 1 in all treatments with a label. Averaging over all label treatments, the mean effect was of -1.56 points (Wilcoxon Signed Rank test, p-value < 0.001) Individually all treatments but NutriRepère showed a significant effect at 5% (WSRT, all p-values < 0.02), while NR was just short of 5% (WSRT, p-value $= 0.0579$). A graphical representation of this main result can be seen in Figure 1, in which error bars represent 95% confidence intervals. Mean treatment effect (absolute difference in FSA score between basket 2 and 1) are given along the diagonal in Table 2. The table also shows the results of pairwise comparing any two treatments. The upper triangle of the Table gives the result of Wilcoxon Rank-sum tests, the lower diagonal of t-tests. Results are robust to the tests used. Among the label treatments, NutriScore performs better than any other label. NutriMark comes in second, outperforming all other remaining labels but NutriCouleur. SENS and NutriRepère are not statistically different from each other and come last.

As noted for the benchmark treatment, there is a natural between-cart variation in the nutritional quality of the shopping. In the labeled treatments, though, there is a switch to healthier products that is directly imputable to the label. The distribution of FSA score change by treatment (Figure 2) shows that in every treatment there is a share of subjects that worsened the overall nutritional score of their shopping. This is, as in the benchmark treatment, an effect of the natural variability in the subjects' shopping. None of the labels reduces the variance of the FSA score changes (all changes in variance are

[4]The source code for the interface, written in PHP, is available upon request.

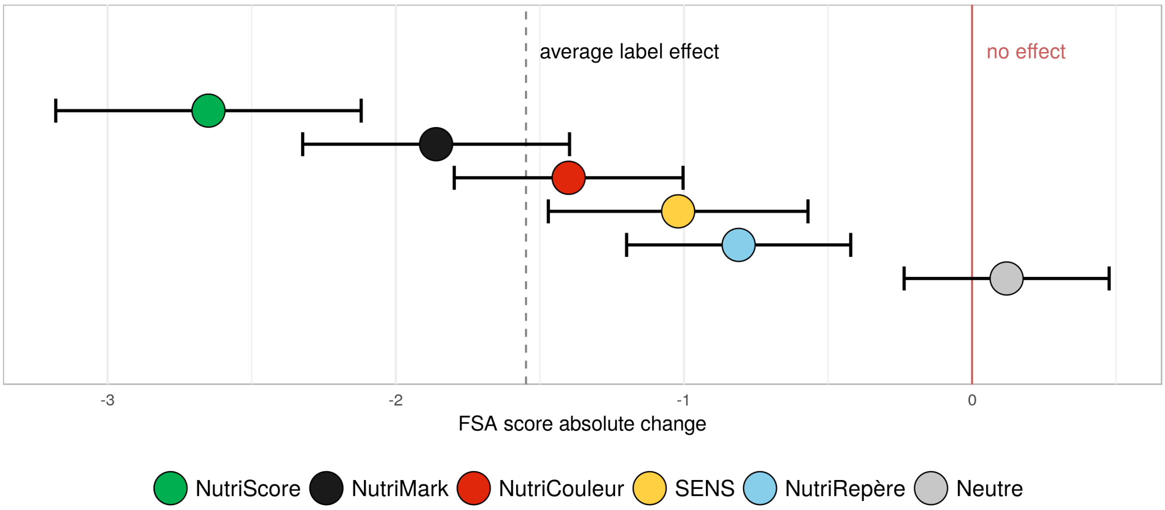


Figure 1: Average absolute FSA score change, basket 2 vs. 1, by treatment

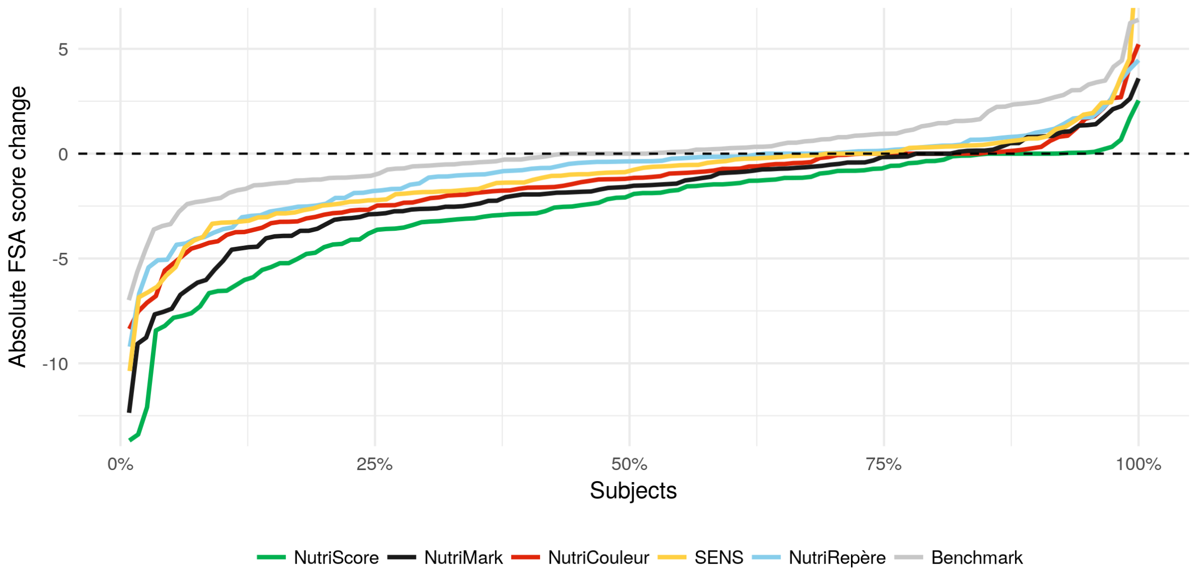


Figure 2: distribution of FSA score change, basket 2 vs. 1, all treatments

	Benchmark	NS	NM	NC	SENS	NR
Benchmark	0.12	<0.001	<0.001	<0.001	<0.001	**0.001**
NS	<0.001	-2.65	**0.034**	**0.001**	<0.001	<0.001
NM	<0.001	**0.048**	-1.86	0.26	**0.014**	<0.001
NC	<0.001	**0.001**	0.394	-1.40	0.142	**0.006**
SENS	**0.002**	<0.001	**0.041**	0.459	-0.81	0.222
NR	**0.019**	<0.001	**0.006**	0.236	0.492	-1.02

Mean treatment effect (FSA score absolute difference) on the diagonal (greyed).
P-values from two-tailed t-tests (lower triangle) and from Wilcoxon rank-sum tests (upper triangle).
Significant tests in bold.

Table 2: Average treatment effects (FSA score) and p-values from t-tests and Wilcoxon rank-sum tests

positive, but none exceeds < 5.6%). Hence, no label reduces the intrinsic variability of diets. Rather, the whole distribution moves down (towards healthier scores).

The descriptive results are confirmed by our difference-in-difference estimation, as shown in Table

3, first column. In the regression, NR and SENS fail to reach significance, but NS, NM and NC all have a significant impact. The results are robust to controlling for demographics (see Table B.13 in AppendixB, column 1). Summarising, even if all labels induce nutritional change, we see a group of labels that consistently show significant results across different methods (NS, NM and NC), and two labels (NR and SENS) that show smaller, barely significant or not significant results.

	All	Low income	Middle income	High income
Intercept	**5.225 *****	**5.372 *****	**5.323 *****	**4.945 *****
	(0.305)	**(0.514)**	**(0.547)**	**(0.508)**
Basket 2	0.115	0.064	0.440	-0.117
	(0.431)	(0.728)	(0.774)	(0.719)
NutriScore	-0.487	-0.627	-0.074	-0.725
	(0.437)	(0.752)	(0.769)	(0.729)
NutriMark	-0.054	-0.194	0.662	-0.781
	(0.434)	(0.724)	(0.759)	(0.758)
NutriCouleur	-0.846	-0.836	-0.820	-0.907
	(0.437)	(0.739)	(0.774)	(0.740)
NutriRepere	-0.430	-1.487	0.636	-0.279
	(0.444)	(0.767)	(0.774)	(0.740)
SENS	-0.656	**-1.763 ***	0.367	-0.340
	(0.441)	**(0.767)**	(0.786)	(0.719)
NutriScore × Basket 2	**-2.766 *****	**-2.584 ***	**-3.294 ****	**-2.487 ***
	(0.619)	**(1.064)**	**(1.087)**	**(1.031)**
NutriMark × Basket 2	**-1.974 ****	-1.916	**-3.049 ****	-0.832
	(0.613)	(1.024)	**(1.073)**	(1.072)
NutriCouleur × Basket 2	**-1.513 ***	-1.104	**-2.274 ***	-1.316
	(0.619)	(1.046)	**(1.095)**	(1.046)
NutriRepere × Basket 2	-0.924	-0.705	-1.485	-0.645
	(0.627)	(1.085)	(1.095)	(1.046)
SENS × Basket 2	-1.140	-0.639	-1.812	-1.079
	(0.624)	(1.085)	(1.111)	(1.016)
N	1382	536	424	422
R 2	0.078	0.073	0.126	0.088
logLik	-3632.179	-1441.235	-1093.776	-1074.627
AIC	7290.357	2908.469	2213.553	2175.255

Significance thresholds: *** p < 0.001; ** p < 0.01; * p < 0.05.

Table 3: Difference in difference treatment effect estimations, overall and by income class

Robustness checks

	Benchmark	NS	NM	NC	SENS	NR
LIM	0.15 (3.34)	-3.94^{***} (4.91)	-2.57^{***} (4.22)	-2.31^{**} (3.9)	$-1.39^{\dagger}$ (4.29)	$-1.44^{\dagger}$ (3.68)
Fat	0.37 (12.83)	-6.85^{**} (11.57)	-6.49^{**} (12.69)	-7.25^{**} (12.46)	$-2.74^{\dagger}$ (15.13)	-3.65 (12.75)
Saturated Fatty Acids	-0.01 (9.33)	-5.96^{***} (9.92)	-5.17^{***} (10.66)	-5.3^{***} (8.81)	-0.64 (12.43)	-3.11^{*} (10.52)
Sugar	0.23 (23.49)	-3.13 (18.97)	-1.75 (19.97)	-0.9 (22.02)	-2.84 (14.06)	-3.72 (18.03)
Salt	0.02 (0.67)	0.03 (0.71)	-0.15 (0.59)	-0.14 (0.61)	-0.01 (0.61)	-0.03 (0.71)
Normalized by weight	0.02 (1.04)	-1.21^{***} (1.45)	-0.85^{***} (1.24)	-0.52^{*} (1.09)	-0.46^{*} (0.98)	-0.54^{*} (1.2)

Wilcoxon Signed Rank tests, significance thresholds: p-values $< 0.10 :^{\dagger} < 0.05 :^{*}; < 0.01 :^{**}; < 0.001 :^{***}$

Table 4: Robustness checks. Mean (st.dev.) absolute change in indicator, by treatment.

Results might depend on the nutritional indicator or normalization used for the analysis. We hence run robustness checks for both the descriptive and regression analyses.

The FSA score may favor some labels over others. In particular, for aggregate labels, NS and NM are directly built on the FSA score, while SENS is built on the LIM nutritional score (Darmon et al., 2009). Insofar as the LIM and the FSA are not perfectly correlated, adoption of the FSA score as the determining indicator could adversely affect the performance of SENS relatively to NS and NM. For analytical labels, NC and NR give nutrient contents and are therefore dependent on the weightings made by the FSA score. We run robustness checks using as independent variable the LIM index for 100Kcal, and four individual nutrients: Salt, Saturated Fatty Acids, Sugar and Fat.

The normalization used to aggregate the shopping basket might also prove distortive. Labels might correlate differently to quantities than to calories, and hence reporting the shopping to energy intakes rather than weight could be a less than innocent exercise. We hence run the FSA score analysis normalizing by weight (100 grams) rather than calories.

Results are summarised in Table 4 (descriptive statistics and tests) and in Table B.12 in AppendixB (diff-in-diff regressions). Results are in general not impacted by the change in indicator or normalization: NS, NM and NC consistently show significant results, while SENS and NR less so. The use of LIM instead of the FSA score does not change the results nor the ranking of labels change. Paradoxically, SENS performs even worse with LIM than with the FSA score. Labels do not induce significant changes in salt nor sugar content. Fat and saturated fatty acids contents are instead impacted. As in the FSA score analysis, NS, NM and NC show significant impacts. NC performs particularly well, likely because of its analytical nature, that gives the consumer interested in a particular nutrient a salient detailed information. Normalizing by weight does not impact the ranking of labels either.

5.2. Economic results

Shopping cost

Nutritional gains might come at an economic cost, which could undermine the long-term impact of labels, as households would face a trade off between nutritional and economic value of their shopping. We start by testing the difference in the total cost of a basket, across shopping baskets, by treatment (Table 5, first two columns). We can exploit the within-subject variation across baskets, and run a Wilcoxon Signed-Rank test (WSRT). The total cost of a basket stayed the same for the Banchmark, NS, NR and SENS treatments, and decreased for NM and NC. Overall spending is not a good indicator, though, since households vary in size and products vary in caloric and nutritional content. To take this into account, we look at the total amount spent for 2000 Kcal, the recommended daily intake of an average adult (Table 5, last two columns). Again we can exploit within-subject variation and run a WSRT.

	Overall cost		Cost of 2000Kcal	
	Difference	p-value	Difference	p-value
Benchmark	0.05	0.65	0.09	0.06
NutriScore	-0.73	0.46	0.39^{***}	0.00
NutriMark	-0.96^{*}	0.03	0.13	0.06
NutriCouleur	-1.11^{***}	0.00	0.21^{**}	0.01
SENS	0.21	0.90	0.16^{*}	0.03
NutriRepère	0.54	0.86	0.23^{**}	0.01

Table 5: Average cost difference and Wilcoxon Signed-Rank test, by treatment.

All labels but NutriMark show a significant increase in the price of a daily calorie intake. The nutritional gain comes indeed at an economic cost – with the exception of NutriMark. This increased cost has an ambiguous effect on the long term impact of labels. If households consume more calories than recommended, a calorie cut would not do harm. If on the other hand the caloric constraint binds then households would face a calorie/cost trade off. Our sample households are well above the recommended number of calories. The average number of calories per household member is 2310 Kcal in the first basket and 2214 in the second (statistically different, Wilcoxon Signed-Rank p-value < 0.001). This is above the 2200 and 2000 Kcal threshold per adult male and female respectively per day. We can conclude that the negative impact of labeling on the amount of calories per euro spent should not impact negatively the long term effect of labels. This should be taken with caution as buying behaviors are probably not a homothetic projection of eating behaviors.

Besides the bare cost of the shopping basket, our data allow us to directly analyze the relationship between the nutritional gain and the economic cost, across treatments. We can cleanly compare treatments using a difference in difference approach as described above. Table 6 reports the results of the estimation

$$\Delta cost_i = \beta_0 + \gamma(-\Delta FSA_i) \times treatment$$

,

in which the between-baskets change in cost, Δcost, is regressed on the interaction of treatment and the between-baskets change in FSA score, ΔFSA. The Benchmark treatment acts as baseline, and, to ease interpretation, we inverted the sign of ΔFSA so that higher values imply better nutritional content. The coefficients can be directly interpreted as extra Euro spent for each 2000Kcal to obtain a one-point improvement in the score FSA. Only NutriCouleur (12 Euro cents) and more markedly SENS (22 Euro cents) show a significantly higher cost of the nutritional adjustment with respect to the Benchmark.

		Δcost	
	Coeff.	Std. Err.	p-value
Intercept	0.10	0.07	0.19
ΔFSA	0.04	0.04	0.32
NutriScore	0.14	0.13	0.28
NutriCouleur	-0.12	0.12	0.31
NutriRepère	0.02	0.11	0.85
NutriMark	-0.11	0.12	0.36
SENS	-0.21	0.11	0.07
ΔFSA × NutriScore	0.02	0.05	0.67
ΔFSA × NutriCouleur	0.12^{*}	0.05	0.02
ΔFSA × NutriRepère	0.10	0.05	0.06
ΔFSA × NutriMark	0.04	0.05	0.45
ΔFSA × SENS	0.22^{***}	0.05	0.00

For clarity, the FSA score sign has been inverted

Table 6: Difference in difference estimation of the cost of FSA improvement

Summarising, the cost of 2000Kcal significantly increased in all treatments but NutriMark. This price increase financed differentiated nutritional gains. When interacting the two dimensions most treatments do not differ from the Benchmark, with the exception of NutriCouleur and in particular of SENS, that displays a much higher cost of adjustment – a sign that SENS generated changes towards more expensive goods for little nutritional gain.

Income groups

The impact of nutritional policies on low income households is usually the key to their success, since nutrition-related problems are often correlated with low income. Our recruitment was stratified by income. Our subjects can hence be divided into three income groups (< 2000, $2000 - 3000$, > 3000 euro/month disposable income) of roughly equal size (see Table A.11 in Appendix AppendixA for the breakdown by treatment).

The nutritional impact of the labels is summarized in Table 3 above, right panels, where our difference in difference specification is run separately by income group. The ranking of the point estimates of label effects stays constant, but for low and high income groups only NutriScore stays significant. This loss of significance is partly due to the lower number of subjects, but since the middle income group shows significant impacts for NS, NM and NC, it also reflects real differences in the subsamples.

Different behavior by income group could also be reflected in the total cost paid for a basket, and the cost associated with nutritional gains. We have reason to expect this, since, as noted elsewhere in the literature (Drewnowski and Eichelsdoerfer, 2010, Kozlova, 2016) low income households buy more caloric food on average. The average cost over all treatments of 2000Kcal in the first basket is of 4.99 euro for lower, 5.13 for medium, and 5.57 for higher income subjects (statistically different, Kruskal-Wallis test, p-value< 0.001).

We now replicate the analysis carried out above, broken down by income group. Analogously to Table 5 above, Table 7 shows the between-basket difference in overall cost and in cost by 2000Kcal by income group. The stars indicate the result of Wilcoxon Signed-Rank Tests (p-values are not reported for readability).

	Overall cost			Cost of 2000Kcal		
	< 2000	2000-3000	> 3000	< 2000	2000-3000	> 3000
Benchmark	0.39	0.70	-0.98	-0.133	0.209	0.28*
NutriScore	-0.72	0.09	-1.57	0.374**	0.478***	0.316*
NutriMark	-0.44	-2.49***	0.10	0.055	0.121	0.245
NutriCouleur	-1.68	-1.11***	-0.33	0.129	0.358*	0.151
SENS	0.09	-1.14*	1.50	0.227	0.027	0.201**
NutriRepère	2.35	-1.20	0.20	0.306	0.006	0.378

Wilcoxon Signed-Rank Test, p-values: ***=0.001, **=0.01, *=0.05

Table 7: Change in cost for 2000Kcal, Basket 2 vs Basket 1, by treatment and income group

The application of labels results, as in the merged sample, of a decrease in overall cost (significant only for the middle income group) but an increase of the cost for 200OKcal for all groups and treatments, but significant only for NS, across all groups, for NC, medium income, and SENS high income groups. A differential effect by income group is therefore not present: labels induce a smaller or bigger increase in cost irrespective of the income group. Formally testing by means of a Mann-Whitney test over all pairwise combinations of income groups within each treatment never yields significant results (all p-values > 0.13), but for a significant difference between medium and high income for NR (p-value 0.02)

The amount of nutritional adjustment bought by the increased cost is estimated using difference in difference in Table 8. Given the complexity of interpreting coefficients from a three-interaction difference in difference regression, we estimate separately for each treatment *j*:

$$\Delta cost_{ij} = \beta_0 + \gamma(-\Delta FSA_i) \times income$$

,

in which, as above, we inverted the sign of the FSA score so that nutritional improvements have a positive sign. Low income subjects act as reference. As above, the coefficient of the interaction between ΔFSA and income group can be directly interpreted as extra Euro for 2000Kcal spent with respect to the low income group to obtain a one-point FSA score improvement.

While there are no consistent effects of income across the board, there are notable exceptions. With NutriScore, high income subjects pay an extra 13 euro cents per 2000Kcal for a one-point improvement in FSA score with respect to the low income category. The same is true for NutriMark (18 cents), and for the medium income group for NutriCouleurs (21 cents). On the other hand, nutritional adjustment is cheaper for higher income groups with SENS (-21 and -17 cents).

Summarising, while a nutritional adjustment is costly, and lower income groups start with higher caloric intensity, the adjustment induced by the label is by and large not income-specific. For the best performing labels, NutriScore and NutriMark, the adjustment is indeed cheaper for lower income groups. This is in stark contrast with subsidies and tax approaches, for which it has been shown that lower income groups pay a higher cost as a result of the policy (see Muller et al., 2017).

5.3. Behavioral results

The laboratory context and the computerized shopping platform allow us to shed some light on the behavioral changes that underlie the nutritional and economic results.

Number of products

Over all treatments, the size of neither basket did significantly change across label treatments, with the exception of the benchmark, that is statistically different from all others. (Table 9, first two rows). This confirms that subjects complied with the instructions. We can exclude boredom as a source of bias for the second basket. The number of items across baskets was not significantly different for any treatment but NC and NM (at 10%). Despite the absolute number of items in each basket being mostly constant, subjects did perform quite some substitutions. On average, a quarter of the products of each basket was substituted between basket 1 and 2 (Table 9, second row). Again, this is an indication that subjects did take into account the information displayed by the labels. The fact that even in the baseline treatment, with no labels, subjects substituted roughly a quarter of all products is a sign that subjects

		Δcost		
		estimate	std.error	p.value
Benchmark	Intercept	-0.13	0.12	0.29
	ΔFSA	0.05	0.05	0.32
	medium	0.32	0.19	0.10
	high	0.40	0.18	0.03
	ΔFSA × medium	-0.09	0.10	0.36
	ΔFSA × high	0.06	0.12	0.61
NutriScore	Intercept	0.40	0.16	0.02
	ΔFSA	-0.01	0.04	0.77
	medium	-0.24	0.27	0.38
	high	-0.40	0.25	0.12
	ΔFSA × medium	0.12	0.07	0.09
	ΔFSA × high	0.13	0.06	0.04
NutriMark	Intercept	-0.04	0.12	0.73
	ΔFSA	0.05	0.03	0.13
	medium	-0.01	0.21	0.95
	high	0.06	0.18	0.74
	ΔFSA × medium	0.02	0.06	0.81
	ΔFSA × high	0.18	0.07	0.01
NutriCouleur	Intercept	0.06	0.14	0.68
	ΔFSA	0.07	0.06	0.23
	medium	-0.21	0.23	0.36
	high	-0.10	0.22	0.64
	ΔFSA × medium	0.21	0.09	0.02
	ΔFSA × high	0.07	0.09	0.46
SENS	Intercept	0.03	0.14	0.85
	ΔFSA	0.35	0.05	0.00
	medium	-0.18	0.23	0.42
	high	-0.04	0.22	0.85
	ΔFSA × medium	-0.21	0.09	0.02
	ΔFSA × high	-0.17	0.10	0.09
NutriRepère	Intercept	0.27	0.12	0.03
	ΔFSA	0.06	0.06	0.36
	medium	-0.45	0.19	0.02
	high	-0.04	0.18	0.84
	ΔFSA × medium	0.12	0.09	0.20
	ΔFSA × high	0.13	0.08	0.10

Table 8: Difference in difference estimations of the cost of FSA improvements, by income groups

varied their choices between the two baskets, by choosing in absence of any change a different menu out of their habits for the second basket.[5]

Use of back-of-pack information

We recorded each time subjects visualized the ingredients list or the nutritional table (Table 9, third and fourth rows). Usually available on the back of packages in grocery shops, subjects could access such information by clicking on a button on the product screens. For basket 1, subjects looked on average between 1 and 2 times at the ingredient lists, and between 0.5 and 1 at nutritional tables, depending on the treatment. Ingredient view frequency significantly decreased in basket 2 for all treatments, including the benchmark; nutrition views decreased for all treatments but the benchmark. On the one hand, this is a straightforward knowledge effect: having acquired the needed information with the first basket, subjects no longer need it for the second. The best estimation of this effect is given by the benchmark treatment, in which no further information was given between baskets. The average number of ingredient and nutrition views decreased by 50% and 32% respectively.

On the other hand, for label treatments, the extent of the decrease measures how much the label crowds out the information contained in the nutritional table and ingredients list. Labels can crowd

[5] Note that the number of product changes is not directly comparable to the difference in the average number of products. This is because the numbers reported are averages of individual values.

		Benchmark	NS	NC	NR	NM	SENS
# products	Basket 1	21.20***	17.90	18.20	19.90	19.80	19.40
	Basket 2	21.00***	17.50	17.60	19.90	19.20	19.40
	Abs difference	-0.30	-0.30	-0.60***	0.00	-0.60†	0.00
# product changes	mean entry	5.30	4.58	3.67	5.80	4.64	4.27
	mean exit	5.63	4.90	4.20	5.20	4.99	4.22
# of ingredient views	Basket 1	1.06	0.88	1.89	1.18	0.90	1.45
	Basket 2	0.53	0.16	0.35	0.10	0.16	0.45
	% difference	-49.6%***	-82.3%***	-81.6%***	-91.5%***	-82.2%***	-68.9%***
# of nutrition views	Basket 1	0.64	0.57	1.14	0.77	0.48	0.73
	Basket 2	0.43	0.17	0.10	0.06	0.13	0.46
	% difference	-31.73%	-69.96%***	-90.83%***	-91.74%***	-72.17%***	-37.04%**

Tests: across treatments Wilcoxon Rank-Sum test; between baskets by treatment: Wilcoxon Signed-Rank test;
Significance thresholds: p-values < 0.10 :† < 0.05 :*; < 0.01 :**; < 0.001 :***;

Table 9: Number of clicks, products, nutrition and ingredient views by treatment. Individual averages.

out back-of-pack information both because they contain it already (as it is the case of NM, NC and NR which basically reproduce the nutritional table) or because they summarize it in a focal and credible way (it is the case of NS and SENS). Thus, the drop in the rate of ingredient and nutrition views is an indicator of both the amount of information given by the label and the trust subjects have in that label. All labels crowd out additional information, but to different extents. NR and NC are nearly perfect substitutes of back-of-pack information and as a result lead to a massive decrease in views (between 82% and 92%). At the other end, SENS does result in a smaller (-37%) decrease, that is not significantly different from the benchmark (Wilcoxon Rank-Sum tast, p-value = 0.29). This means that SENS fails to give nutritional information or that subjects do not trust the label enough to forfeit looking at the nutritional table.

Label-specific behavioral change

Labels have different granularity, focality, and aims. It is hence possible to analyze the behavioral change of subjects with respect to the characteristics of each label. Figure 3 summarizes our indicators of label-specific behavioral change. While these indicators are not strictly comparable, focusing on them allows us to assess how subjects used the cues given by the labels, and to extract some regularities across treatments.

For aggregated labels (NM, SENS and NS), subjects are highly compliant. Nonetheless, they have a tendency to oversimplify the already simple messages carried by the labels by limiting their attention to extreme values or neglecting the nuances of labels. For NM, only the products in the first two categories (five and four and a half stars) are mostly chosen in basket 2 to replace products in basket 1. For SENS, while green products (for every day) replace purple products (occasionally), the two intermediate categories generate little change. For NS, the average number of replaced red and orange products is virtually the same, and light and dark green are very close: despite the fact that the scale has 5 values, subjects behave as if it had 3. Across all aggregated labels, subjects transform more nuanced messages into "good/bad (or neutral)" signals.

Subjects are also compliant with analytical labels (NC and NR) but to varying degrees depending on nutrients. Subjects pay most attention to fat, some to sugar, and not much to salt. Overall, products with red nutrients decrease and products with green nutrients increase with NC. Similarly, the frequency of products with lower bad nutrient content decreases with NR. Given the continuous nature of NR, in Figure 3 we show the average change in % reference intake points across the two baskets. Subjects seem to favor fat and saturated fatty acid information.

Label comprehension and assessment

At the end of the experiment, all subjects, except those who took part in the benchmark treatment, completed a questionnaire on the qualitative assessment of the label. Subjects only assessed the label with which they were confronted during the experiment. Accordingly results do not reflect a *comparative*, but rather *absolute* assessment. Each response was on a 4-item likert scale ("yes", "mostly", "not really", "no"). Results are given in Table 10. The Table reports summary indicator, that is computed by giving scores of 1, 0.75, 0.25 and 0 to each of the four possible answers. Results are robust to changing the values of the intermediate options. Bold indicates the label that has the higher aggregate score for each question.

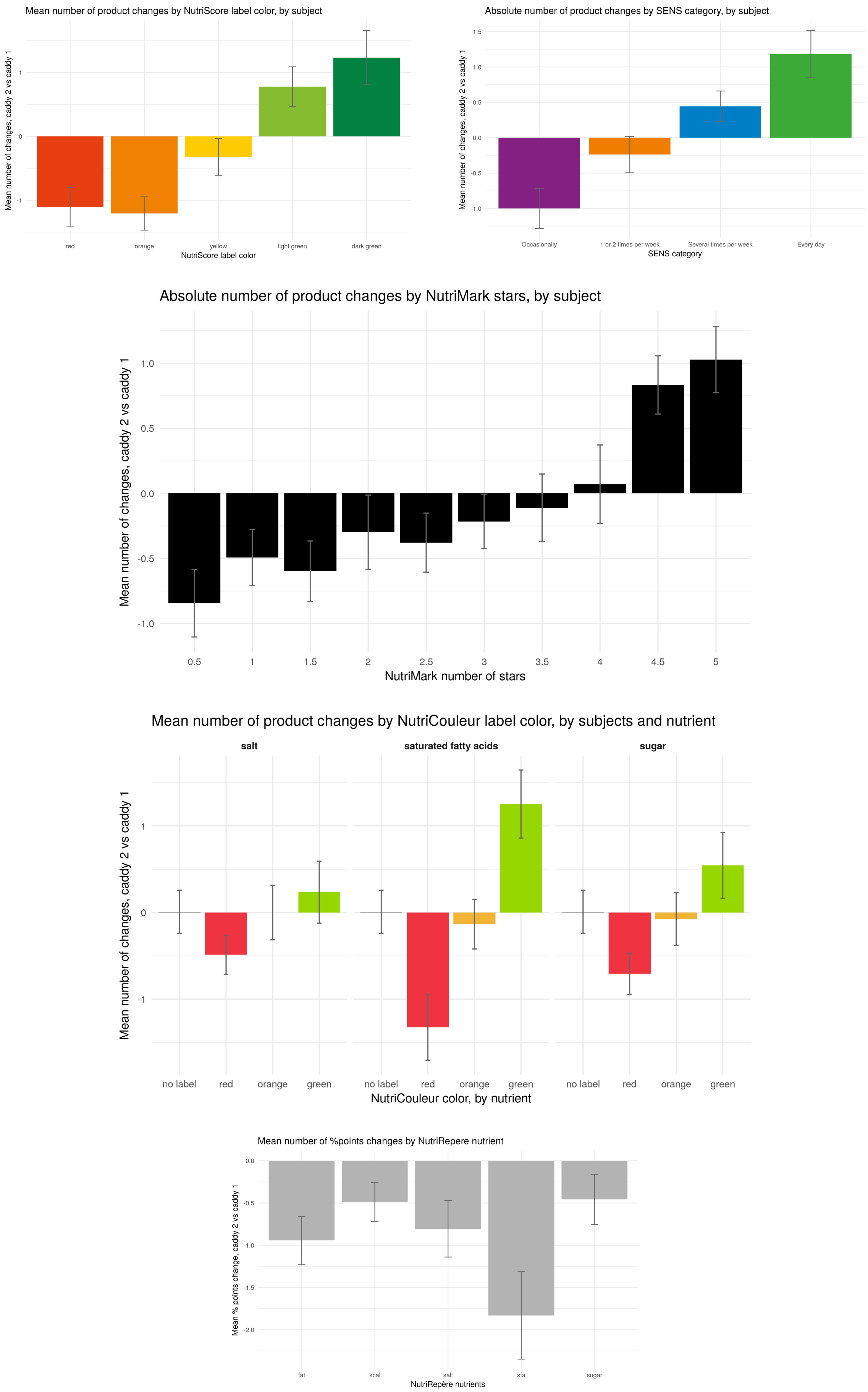


Figure 3: Label-specific behavioral change

	NS	NC	NR	NM	SENS
The nutritional label is...					
easy to understand	**0.94**	0.89	0.86	0.91	0.91
useful	**0.94**	0.91	0.88	0.89	0.85
precise	0.71	0.76	0.74	**0.79**	0.57
reassuring	**0.75**	0.70	0.69	0.74	0.67
a tip for good choices	0.73	0.65	0.61	**0.74**	0.65
just advertisement	0.30	0.36	0.34	0.34	**0.41**
The nutritional label does...					
give information about food items to limit	0.79	0.72	0.74	**0.81**	0.77
give information on the nutritional composition	0.60	**0.75**	0.68	0.71	0.46
show me the nutritional quality	**0.79**	**0.79**	**0.79**	**0.79**	0.69
The nutritional label will...					
influence my shopping	**0.71**	0.65	0.67	**0.71**	0.63
help build a better diet	**0.81**	0.72	0.73	0.76	0.62
help in following health recommendations	**0.81**	0.76	0.78	0.78	0.71

Table 10: Qualitative label assessment, aggregate score by treatment

While differences are small, some patterns are recognizable. NutriScore ranks first for simplicity, usefulness, and is considered influential and useful in building better diets that follow nutritional recommendations. On the other end of the spectrum, SENS is the label that most appears as being just an 'advertisement stunt' for products and is ranked last according to several criteria.

6. Discussion

Results indicate that FOPLs can be an effective policy instrument to guide consumers towards healthier diets. Nutritional effects are significant and surprisingly large, especially for the aggregate, color-coded labels like NutriScore. If nutritional changes of the magnitude recorded in this experiment were to occur out of the laboratory and survive in the long run, the incidence of nutrition-related diseases would be substantially reduced (see epidemiological evidence as detailed in Adriouch et al., 2016, Julia et al., 2015, Donnenfeld et al., 2015). Moreover, our data show that labels do not have the regressive effects that characterize other policies, as taxes and subsidies (as indicated in a similar exercise by Muller et al., 2017). Lower income subjects faced, especially for NutriScore and NutriMark, *lower* cost of nutritional adjustment than medium and high-income subjects. Thus clear, focal, aggregated labels prove able to generate shopping changes that lead to nutritionally improved choices at low economic cost, especially so for poorer households.

Nonetheless, the external validity of these results needs to be taken with caution. Although we have endeavored to create a context as close as possible to real purchasing situations (sample of regular shoppers; sale at market prices of a wide range of real products; computer interface very close to online or drive-through shopping websites, etc.), a crucial difference with real shopping remains: in our setting, labels are highly focal. Baskets 1 and 2 only differ in the presence of labels. As a result, the subjects' attention is automatically captured by the labels. While this feature reinforces the internal validity of our experiment, it may exaggerate the absolute impact of labels compared to real-world situations where consumer attention is limited. It is therefore possible that we identify here an upper bound of the potential effect of the labels.

On the other hand, it is possible to argue otherwise. Labels might in the long run interact with the processes of habit formation (Zhen et al., 2011, Daunfeldt et al., 2012, Verplanken and Wood, 2006), and their effect might be hence in the long run – that we cannot measure in the lab – larger than in the short run. Moreover, choices are usually made in a social context, and the willingness to signal compliance or superior eating habits could be a further force inflating the effects of the label that we do not capture in our laboratory, in which full privacy is granted (Etilé and Teyssier, 2016, Teyssier et al., 2014).

A second potential problem with laboratory experiments lies in experimenter demand effect. The aim of our experiment – testing nutritional label – was transparent to subjects by the time they had to compose the second basket. Moreover, subjects knew that our study had been financed by the Ministry for Health, and a heated political debate about labels was running in France roughly in parallel with our experimental campaign. It is possible that subjects could show appreciation to the experimenters' work by complying with the labels. Moreover, politically motivated subjects could have inflated their reaction to labels to support their political agenda.

Nonetheless, we believe these effect on the one hand to be minor, and on the other hand to be controlled by our design. Recent evidence shows that experimental demand effects are likely low (Mummolo and Peterson, 2017, De Quidt et al., 2017) and crowded out by incentives. In our experiment, subjects faced real purchasing decisions, and they had to leave with the *products* themselves; the cost of sending a political message was hence relatively high. Moreover, to the extent that demand effects and self serving bias are not label-specific, their impact is fully controlled for in our difference in difference estimations.

While remaining cautious about the absolute magnitude of the labels' impact, the experiment's magnifying glasses allow us to emphasize their relative effects. Hence, the key result of our study is that color-coded, aggregate, easy to understand label have stronger nutritional impacts than analytic, detailed, numeric labels; the cost of the adjustment is low and does not entail the regressive effects shown to appear with taxes and subsidies. This core result is strong and likely to survive intact outside of the laboratory.

The best policy tool from a public health perspective appears to be a label that gives synthetic, relevant and focal information, and that can win the trust of consumers. It should be able to stand out among competing stimuli, to be simple to understand and easy to include in each consumers' choice function; it need not necessarily give accurate nutritional information, as long as it is trusted by the consumers to do so. If simple and direct labels work best in a laboratory setting in which subjects had the time and incentive to ponder their choices, we expect them to *a fortiori* work in the real world, where attention, time and cognitive resources are limited.

Disclaimer

Part of the data presented here have already been object of a short publication with strict nutritional focus, in French, as Crosetto et al. (2017). This paper, besides being in English, contains an updated and enlarged nutritional analysis, while the economic and behavioral analyses are entirely new.

Acknowledgements

The study was financed by the French Ministry of Health and Social Affairs. The authors would like to thank Marie Cronfalt-Godet, Mariane Damois, Anne Lacroix, Jean-Loup Dupuis and Sabine Pique for invaluable technical and logistic support before and during the experiments, as well as participants to seminars in Parma, Paris, Göttingen, for valuable comments. All errors are the authors'.

References

Adriouch, Solia, Chantal Julia, Emmanuelle Kesse-Guyot, Caroline Méjean, Pauline Ducrot, Sandrine Péneau, Mathilde Donnenfeld, Mélanie Deschasaux, Mehdi Menai, Serge Hercberg et al., "Prospective association between a dietary quality index based on a nutrient profiling system and cardiovascular disease risk," *European journal of preventive cardiology*, 2016, *23* (15), 1669–1676.

Aschemann-Witzel, Jessica, Klaus G Grunert, Hans CM van Trijp, Svetlana Bialkova, Monique M Raats, Charo Hodgkins, Grazyna Wasowicz-Kirylo, and Joerg Koenigstorfer, "Effects of nutrition label format and product assortment on the healthfulness of food choice," *Appetite*, 2013, *71*, 63–74.

Balcombe, Kelvin, Iain Fraser, and Salvatore Di Falco, "Traffic lights and food choice: A choice experiment examining the relationship between nutritional food labels and price," *Food Policy*, 2010, *35* (3), 211 – 220.

Becker, Mark W., Nora M. Bello, Raghav P. Sundar, Chad Peltier, and Laura Bix, "Front of pack labels enhance attention to nutrition information in novel and commercial brands," *Food Policy*, 2015, *56* (Supplement C), 76 – 86.

Berning, Joshua P., Hayley H. Chouinard, and Jill J. McCluskey, "Do Positive Nutrition Shelf Labels Affect Consumer Behavior? Findings from a Field Experiment with Scanner Data," *American Journal of Agricultural Economics*, 2011, *93* (2), 364–369.

Bialkova, Svetlana and Hans van Trijp, "What determines consumer attention to nutrition labels?," *Food Quality and Preference*, 2010, *21* (8), 1042 – 1051. Eighth Pangborn Sensory Science Symposium.

Boztuğ, Yasemin, Hans Jørn Juhl, Ossama Elshiewy, and Morten Berg Jensen, "Consumer response to monochrome Guideline Daily Amount nutrition labels," *Food Policy*, 2015, *53* (Supplement C), 1 – 8.

Cecchini, M and L Warin, "Impact of food labelling systems on food choices and eating behaviours: a systematic review and meta-analysis of randomized studies," *Obesity reviews*, 2016, *17* (3), 201–210.

Cooper, SL, FE Pelly, and JB Lowe, "Assessment of the construct validity of the Australian Health Star Rating: a nutrient profiling diagnostic accuracy study," *European Journal of Clinical Nutrition*, 2017.

Council of the European Union, "Regulation (EU) No 1169/2011," 2011. http://data.europa.eu/eli/reg/2011/1169/oj.

Crosetto, Paolo, Anne Lacroix, Laurent Muller, and Bernard Ruffieux, "Modification des achats alimentaires en réponse à cinq logos nutritionnels," *Cahiers de Nutrition et de Diététique*, 2017, *5262* (3), 107–159.

_ , Laurent Muller, and Bernard Ruffieux, "Helping consumers with a front-of-pack label: Numbers or colors?: Experimental comparison between Guideline Daily Amount and Traffic Light in a diet-building exercise," *Journal of Economic Psychology*, 2016, *55*, 30 – 50. Special issue on Food consumption behavior: Economic and psychological perspectives.

_ , _ , and _ , "Réponses des consommateurs à trois systèmes d'étiquetage nutritionnel face avant," *Cahiers de Nutrition et de Diététique*, 2016, *5011* (3), 111–160.

Darmon, Nicole, Florent Vieux, Matthieu Maillot, Jean-Luc Volatier, and Ambroise Martin, "Nutrient profiles discriminate between foods according to their contribution to nutritionally adequate diets: a validation study using linear programming and the SAIN, LIM system," *The American journal of clinical nutrition*, 2009, *89* (4), 1227–1236.

Daunfeldt, Sven-Olov, Jonas Nordström, and Linda Thunström, "Habit formation in food consumption," 2012.

de la Santé Publique, Haut Conseil, "Avis relatif à l'information sur la qualité nutritionnelle des produits alimentaires," Technical Report, HCSP November 2015.

Donnenfeld, Mathilde, Chantal Julia, Emmanuelle Kesse-Guyot, Caroline Méjean, Pauline Ducrot, Sandrine Péneau, Mélanie Deschasaux, Paule Latino-Martel, Léopold Fezeu, Serge Hercberg et al., "Prospective association between cancer risk and an individual dietary index based on the British Food Standards Agency Nutrient Profiling System," *British Journal of Nutrition*, 2015, *114* (10), 1702–1710.

Drewnowski, Adam and Petra Eichelsdoerfer, "Can low-income Americans afford a healthy diet?," *Nutrition today*, 2010, *44* (6), 246.

Drichoutis, Andreas C., Panagiotis Lazaridis, and Rodolfo M. Nayga Jr, "Nutrition knowledge and consumer use of nutritional food labels," *European Review of Agricultural Economics*, 2005, *32* (1), 93–118.

Drichoutis, Andreas C, Panagiotis Lazaridis, Rodolfo M Nayga, Maria Kapsokefalou, and George Chryssochoidis, "A theoretical and empirical investigation of nutritional label use," *The European Journal of Health Economics*, 2008, *9* (3), 293–304.

Ducrot, Pauline, Caroline Méjean, Chantal Julia, Emmanuelle Kesse-Guyot, Mathilde Touvier, Léopold K Fezeu, Serge Hercberg, and Sandrine Péneau, "Objective understanding of front-of-package nutrition labels among nutritionally at-risk individuals," *Nutrients*, 2015, *7* (8), 7106–7125.

_ , Chantal Julia, Caroline Méjean, Emmanuelle Kesse-Guyot, Mathilde Touvier, Léopold K Fezeu, Serge Hercberg, and Sandrine Péneau, "Impact of different front-of-pack nutrition labels on consumer purchasing intentions: a randomized controlled trial," *American journal of preventive medicine*, 2016, *50* (5), 627–636.

Etilé, Fabrice and Sabrina Teyssier, "Signaling Corporate Social Responsibility: Third-Party Certification versus Brands," *The Scandinavian Journal of Economics*, 2016, *118* (3), 397–432.

FSA, "Guide to creating a front of pack (FoP) nutrition label for pre-packed products sold through retail outlets," Guidance of the Department of Health of the United Kingdom of Great Britain and Northern Ireland June 2013.

Grunert, Klaus G. and Josephine M. Wills, "A review of European research on consumer response to nutrition information on food labels," *Journal of Public Health*, 2007, *15* (5), 385–399.

Hersey, James C, Wohlgenant Kelly C, Arsenault Joanne E, Kosa Katherine M, and Muth Mary K, "Effects of front-of-package and shelf nutrition labeling systems on consumers," *Nutrition Reviews*, 2013, *71* (1), 1–14.

Julia, Chantal and Serge Hercberg, "Development of a new front-of-pack nutrition label in France: the five-colour Nutri-Score," *PANORAMA*, 2017, p. 712.

_ , Emmanuelle Kesse-Guyot, Mathilde Touvier, Caroline Méjean, Léopold Fezeu, and Serge Hercberg, "Application of the British Food Standards Agency nutrient profiling system in a French food composition database," *British Journal of Nutrition*, 2014, *112* (10), 1699–1705.

_ , Olivier Blanchet, Caroline Méjean, Sandrine Péneau, Pauline Ducrot, Benjamin Allès, Léopold K. Fezeu, Mathilde Touvier, Emmanuelle Kesse-Guyot, Eric Singler, and Serge Hercberg, "Impact of the front-of-pack 5-colour nutrition label (5-CNL) on the nutritional quality of purchases: an experimental study," *International Journal of Behavioral Nutrition and Physical Activity*, Sep 2016, *13* (1), 101.

_ , Pauline Ducrot, Camille Lassale, Léopold Fézeu, Caroline Méjean, Sandrine Péneau, Mathilde Touvier, Serge Hercberg, and Emmanuelle Kesse-Guyot, "Prospective associations between a dietary index based on the British Food Standard Agency nutrient profiling system and 13-year weight gain in the SU. VI. MAX cohort," *Preventive medicine*, 2015, *81*, 189–194.

Kelly, Bridget, Clare Hughes, Kathy Chapman, Jimmy Chun-Yu Louie, Helen Dixon, Jennifer Crawford, Lesley King, Mike Daube, and Terry Slevin, "Consumer testing of the acceptability and effectiveness of front-of-pack food labelling systems for the Australian grocery market," *Health Promotion International*, 2009, *24* (2), 120–129.

Kozlova, Olga, "Is Healthy Food a Luxury for the Low-Income Households in the US?," 2016.

Marchi, Elisa De, Vincenzina Caputo, Rodolfo M. Nayga, and Alessandro Banterle, "Time preferences and food choices: Evidence from a choice experiment," *Food Policy*, 2016, *62* (Supplement C), 99 – 109.

Mørk, Trine, Klaus G Grunert, Morten Fenger, Hans Jørn Juhl, and George Tsalis, "An analysis of the effects of a campaign supporting use of a health symbol on food sales and shopping behaviour of consumers," *BMC public health*, 2017, *17* (1), 239.

Muller, L. and M. Prevost, "What cognitive sciences have to say about the impacts of nutritional labelling formats," *Journal of Economic Psychology*, 2016, *55*, 17 – 29. Special issue on Food consumption behavior: Economic and psychological perspectives.

Muller, Laurent and Bernard Ruffieux, "Modification des achats en réponse à l'apposition de différents logos d'évaluation nutritionnelle sur la face avant des emballages," *Cahiers de Nutrition et de Diététique*, 2012, *47* (4), 171 – 182.

_ , Anne Lacroix, Jayson L. Lusk, and Bernard Ruffieux, "Distributional Impacts of Fat Taxes and Thin Subsidies," *The Economic Journal*, feb 2017, *127* (604), 2066–2092.

Mummolo, Jonathan and Erik Peterson, "Demand Effects in Survey Experiments: An Empirical Assessment," 2017.

Neal, Bruce, Michelle Crino, Elizabeth Dunford, Annie Gao, Rohan Greenland, Nicole Li, Judith Ngai, Cliona Ni Mhurchu, Simone Pettigrew, Gary Sacks et al., "Effects of Different Types of Front-of-Pack Labelling Information on the Healthiness of Food Purchases—A Randomised Controlled Trial," *Nutrients*, 2017, *9* (12), 1284.

Norden, "The Keyhole: Healthy choices made easy," Technical Report, Nordic Cooperation – Nordic Council of Ministries 2010.

Nugier, A, AJ Serry, and VN Thanh, "Système d'information nutritionnelle à cinq couleurs sur la face avant des emballages alimentaires: comparaison de trois formats de logo," *Bulletin Epidémiologique Hebdomadaire*, 2016, *34*, 598–607.

of Australia, Commonwealth, "Proposed Front-of-pack food labeling designs: qualitative research otucomes," Technical Report, Hall & Partners Open Mind 2013.

_ , "Proposed Front-of-pack food labeling designs: quatitative research otucomes," Technical Report, Hall & Partners Open Mind 2013.

of Australia, National Heart Foundation, "The Hearth Foundation Tick," Technical Report, National Heart Foundation of Australia 1989.

Quidt, Jonathan De, Johannes Haushofer, and Christopher Roth, "Measuring and Bounding Experimenter Demand," Technical Report, National Bureau of Economic Research 2017.

Rayner, M, Peter Scarborough, and Tim Lobstein, "The UK Ofcom Nutrient Profiling Model: Defining 'healthy'and 'unhealthy'foods and drinks for TV advertising to children," *London: OfCom*, 2009.

Rayner, Mike, Peter Scarborough, and Carol Williams, "The origin of Guideline Daily Amounts and the Food Standards Agency's guidance on what counts as 'a lot'and 'a little'," *Public Health Nutrition*, 2004, *7* (4), 549–556.

République, Française, "Arrêté du 31 octobre 2017 fixant la forme de présentation complémentaire à la déclaration nutritionnelle recommandée par l'Etat en application des articles L. 3232-8 et R. 3232-7 du code de la santé publique.," *Journal Officiel de la République Française.*, 3 November 2017, *0257* (16).

Shogren, Jason F, *The Oxford handbook of the economics of food consumption and policy*, Oxford University Press, 2011.

Teyssier, Sabrina, Fabrice Etilé, and Pierre Combris, "Social-and self-image concerns in fair-trade consumption," *European Review of Agricultural Economics*, 2014, *42* (4), 579–606.

Tharrey, Marion, Matthieu Maillot, Véronique Azaïs-Braesco, and Nicole Darmon, "From the SAIN,LIM system to the SENS algorithm: a review of a French approach of nutrient profiling," *The Proceedings of the Nutrition Society*, 2017, pp. 1–10 (First view). Proceedings of The Nutrition Society Summer Meeting 2016 held at University College, Dublin on 11–14 July 2016 Conference on 'New technology in nutrition research and practice' Nutrient profiling as a tool to respond to public health needs Uses of nutrient profiling to address public health needs: from regulation to reformulation.

Variyam, Jayachandran N., "Do nutrition labels improve dietary outcomes?," *Health Economics*, 2008, *17* (6), 695–708.

Verplanken, Bas and Wendy Wood, "Interventions to Break and Create Consumer Habits," *Journal of Public Policy & Marketing*, 2006, *25* (1), 90–103.

Volkova, Ekaterina and Cliona Ni Mhurchu, "The influence of nutrition labeling and point-of-purchase information on food behaviours," *Current obesity reports*, 2015, *4* (1), 19–29.

Vyth, Ellis L, Ingrid H M Steenhuis, Hella E Brandt, Annet J C Roodenburg, Johannes Brug, and Jacob C Seidell, "Methodological quality of front-of-pack labeling studies: a review plus identification of research challenges," *Nutrition reviews*, 2012, *70*, 709–20.

_ **, Ingrid HM Steenhuis, Jessica A Vlot, Anouk Wulp, Meefa G Hogenes, Danielle H Looije, Johannes Brug, and Jacob C Seidell**, "Actual use of a front-of-pack nutrition logo in the supermarket: consumers' motives in food choice," *Public Health Nutrition*, 11 2010, *13*, 1882–1889.

Waterlander, Wilma Elzeline, Cliona Ni Mhurchu, and Ingrid HM Steenhuis, "Effects of a price increase on purchases of sugar sweetened beverages. Results from a randomized controlled trial," *Appetite*, 2014, *78*, 32–39.

Zhang, Donglan, Yan Li, Guijing Wang, Andrew E. Moran, and José A. Pagán, "Nutrition Label Use and Sodium Intake in the U.S.," *American Journal of Preventive Medicine*, 2017, *53* (6, Supplement 2), S220 – S227. The Economics of Hypertension and Cardiovascular Disease.

Zhen, Chen, Michael K. Wohlgenant, Shawn Karns, and Phillip Kaufman, "Habit Formation and Demand for Sugar-Sweetened Beverages," *American Journal of Agricultural Economics*, 2011, *93* (1), 175–193.

AppendixA. Subject sample properties and randomization checks

		N	%	NS	NC	NR	NM	SENS	Base	χ^2
Total number of subjects		691	0	115	115	109	119	111	122	
gender	male	140	20.2	22.4	19.0	20.9	20.2	16.2	22.1	0.87
	female	554	79.8	77.6	81.0	79.1	79.8	83.8	77.9	
age	<30	141	20.3	24.1	20.7	18.2	18.5	19.8	20.5	0.49
	30-44	352	50.7	44.0	45.7	55.5	52.1	52.3	54.9	
	45-59	187	27.0	29.3	28.5	25.5	29.4	26.1	23.0	
	>60	14	2.0	2.6	5.2	0.9	0	1.8	1.6	
income	low	270	38.9	37.1	40.5	37.3	42.0	36.0	40.2	0.99
	middle	212	30.6	31.0	30.2	31.8	31.9	29.7	28.7	
	high	212	30.6	31.9	29.3	30.9	26.1	34.2	31.2	
education	<high school	198	28.5	22.4	28.5	33.6	24.4	29.7	32.8	0.35
	high school	94	13.5	17.2	15.5	14.6	16.0	8.1	9.8	
	university or >	402	57.9	60.3	56.0	51.8	59.7	62.2	57.4	
profession	handicraft	14	2.0	2.6	0	0.9	1.7	4.5	2.5	**0.01**
	managers	98	14.1	13.8	11.2	20.0	8.4	12.6	18.9	
	white collar	436	62.8	65.5	67.2	56.4	74.0	59.5	54.1	
	blue collar	21	3.0	0.9	0	7.3	1.7	4.5	4.1	
	professionals	83	12.0	12.1	12.9	12.7	7.6	15.3	11.5	
	no	42	6.1	5.2	8.6	2.7	6.7	3.6	9.0	
occupation	unemployed	72	10.4	12.1	10.3	10.9	10.9	10.8	7.4	0.81
	student	28	4.0	3.5	5.2	3.6	5.9	2.7	3.3	
	housewife	65	9.4	5.2	11.2	9.1	13.5	4.5	12.3	
	looking for	5	0.7	0.9	0.9	0.9	0.8	0	0.8	
	retired	7	1.0	1.7	1.7	0	0	0.9	1.6	
	working	517	74.5	76.7	70.7	75.5	68.9	81.1	74.6	
family	couple with children	437	63.0	63.8	51.7	65.5	63.9	65.8	67.2	0.47
	couple without children	59	8.5	9.5	14.7	7.3	6.7	8.1	4.9	
	alone with children	143	20.6	20.7	21.6	20.0	22.7	18.9	19.7	
	alone without children	55	7.9	6.0	12.1	7.3	6.7	7.2	8.2	

Table A.11: Demographics of the sample and distribution across treatments

AppendixB. Robustness checks

	LIM	Salt	Sugar	Fat	SFA	Weight
Intercept	**17.964 *****	**2.114 *****	**71.331 *****	**41.489 *****	**19.724 *****	**1.609 *****
	(0.517)	**(0.084)**	**(2.915)**	**(1.690)**	**(1.157)**	**(0.438)**
Caddy 2	0.146	0.023	0.233	0.368	-0.014	0.022
	(0.731)	(0.119)	(4.122)	(2.390)	(1.636)	(0.249)
NutriScore	-0.732	-0.097	**-9.391 ***	**-4.772 ***	-1.857	-0.257
	(0.742)	(0.121)	**(4.184)**	**(2.427)**	(1.661)	(0.254)
NutriMark	-0.258	0.050	**-8.501 ***	0.953	1.584	0.070
	(0.735)	(0.120)	**(4.148)**	(2.406)	(1.647)	(0.252)
NutriCouleur	-1.002	-0.236	-7.879	-3.746	0.176	-0.314
	(0.742)	(0.121)	(4.184)	(2.427)	(1.661)	(0.255)
NutriRepere	-0.853	0.186	-4.276	-1.186	-0.842	-0.203
	(0.752)	(0.123)	(4.243)	(2.461)	(1.684)	(0.257)
SENS	-1.048	0.110	**-11.976 ****	-2.249	-0.744	-0.372
	(0.748)	(0.122)	**(4.223)**	(2.449)	(1.676)	(0.257)
NutriScore × Basket 2	**-4.086 *****	0.010	-3.366	**-7.220 ***	**-5.949 ***	**-1.231 *****
	(1.049)	(0.171)	(5.917)	**(3.432)**	**(2.349)**	**(0.357)**
NutriMark × Basket 2	**-2.719 ****	-0.172	-1.982	**-6.858 ***	**-5.159 ***	**-0.868 ***
	(1.040)	(0.169)	(5.866)	**(3.402)**	**(2.329)**	**(0.354)**
NutriCouleur × Basket 2	**-2.457 ***	-0.165	-1.137	**-7.616 ***	**-5.291 ***	-0.542
	(1.049)	(0.171)	(5.917)	**(3.432)**	**(2.349)**	(0.358)
NutriRepere × Basket 2	-1.584	-0.057	-3.955	-4.023	-3.095	-0.560
	(1.063)	(0.173)	(6.001)	(3.480)	(2.382)	(0.362)
SENS × Basket 2	-1.534	-0.029	-3.071	-3.111	-0.630	-0.478
	(1.058)	(0.173)	(5.972)	(3.463)	(2.371)	(0.360)
N	1382	1380	1382	1382	1382	1380
R 2	0.060	0.028	0.019	0.043	0.039	0.133
logLik	-4361.625	-1849.324	-6752.937	-5999.972	-5476.186	-2860.267
AIC	8749.249	3724.649	13531.875	12025.944	10978.373	5778.534

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table B.12: Difference in difference robustness regression: all alternative indicators and normalizations

	Score FSA			
	All	**Low income**	**Middle income**	**High income**
Intercept	**5.445** ***	**5.713** ***	**5.909** ***	0.207
	(0.736)	**(1.103)**	**(1.379)**	(1.899)
Basket 2	0.115	0.064	0.440	-0.117
	(0.418)	(0.710)	(0.742)	(0.688)
NutriScore	-0.259	-0.372	0.400	-0.598
	(0.427)	(0.749)	(0.750)	(0.707)
NutriMark	0.091	0.188	0.782	-0.249
	(0.423)	(0.714)	(0.739)	(0.739)
NutriCouleur	-0.480	-0.299	-0.430	-0.660
	(0.429)	(0.736)	(0.755)	(0.730)
NutriRepere	-0.384	-1.051	0.488	-0.280
	(0.432)	(0.757)	(0.757)	(0.713)
SENS	-0.478	**-1.556** *	0.502	-0.081
	(0.431)	**(0.762)**	(0.761)	(0.695)
NutriScore × Basket 2	**-2.766** ***	**-2.584** *	**-3.294** **	**-2.487** *
	(0.600)	**(1.038)**	**(1.041)**	**(0.986)**
NutriMark × Basket 2	**-1.974** ***	-1.916	**-3.049** **	-0.832
	(0.595)	(0.999)	**(1.028)**	(1.026)
NutriCouleur × Basket 2	**-1.513** *	-1.104	**-2.274** *	-1.316
	(0.600)	(1.020)	**(1.049)**	(1.000)
NutriRepere × Basket 2	-0.924	-0.705	-1.485	-0.645
	(0.609)	(1.059)	(1.049)	(1.000)
SENS × Basket 2	-1.140	-0.639	-1.812	-1.079
	(0.606)	(1.059)	(1.065)	(0.972)
Middle income	0.222			
	(0.253)			
High income	**-0.612** *			
	(0.278)			
Age 30 -44	0.159	0.562	0.233	-0.782
	(0.257)	(0.403)	(0.450)	(0.588)
Age 45 -59	**-0.777** **	-0.650	-0.887	**-1.294** *
	(0.281)	(0.449)	(0.513)	**(0.618)**
Age > 60	**-2.555** **	-2.646	**-4.625** **	-1.101
	(0.917)	(1.497)	**(1.677)**	(1.633)
Family size	**0.171** *	0.108	-0.022	**1.002** ***
	(0.082)	(0.113)	(0.194)	**(0.202)**
# of children	0.070	-0.001	**0.526** *	**-0.594** **
	(0.116)	(0.196)	**(0.218)**	**(0.213)**
Body Mass Index	-0.010	0.014	-0.054	0.007
	(0.018)	(0.029)	(0.033)	(0.037)
Worker	0.236	-0.464	0.687	**2.874** *
	(0.334)	(0.494)	(0.556)	**(1.124)**
Unemployed	**0.934** *	-0.035	1.252	**4.050** **
	(0.412)	(0.653)	(0.665)	**(1.334)**
Student	**-1.371** *	**-2.278** **	1.896	**6.420** **
	(0.565)	**(0.737)**	(1.395)	**(2.393)**
Retired	1.631	2.317	2.894	2.233
	(1.273)	(2.335)	(2.768)	(2.240)
High school degree	-0.366	-0.383	**-1.202** *	0.411
	(0.295)	(0.475)	**(0.525)**	(0.606)
University degree	-0.384	-0.641	-0.748	0.045
	(0.221)	(0.364)	(0.398)	(0.440)
N	1382	536	424	422
R 2	0.142	0.141	0.225	0.194
logLik	-3582.712	-1420.807	-1068.250	-1048.540
AIC	7223.424	2895.613	2190.500	2151.080

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table B.13: Difference in difference FSA score estimation, overall and by income class, with controls

AppendixC. Experimental instructions

The original instructions were made up of a Power Point slideshow, using several visual cues as to make them appealing and easy to understand for the varied and heterogeneous population of subjects we faced. The French version of the slideshow is available here. Here we provide a translation of all the words, plus the most relevant pictures.

General instructions

Welcome. This experiment is run by the Grenoble Applied Economics Laboratory (GAEL), part of the University Grenoble-Alpes (UGA). This study is financed by the Ministry for Social Affairs, Health and Womens' Rights.

The study is about individual food consumption behavior. Instructions will be given to you as we go along. During the whole session, you will have to take some simple decisions. Nonetheless, should you have any difficulty or misunderstandings, do not hesitate to ask.

In order to protect your privacy during the session and in data analysis, you have been assigned a code. No data allowing us to identify you will be collected. Thus, it will be impossible for us to link your replies and decisions to your name. Data will be kept for statistical analysis and publication, but always in their anonymous format.

Communication between participants is not allowed, nor are comments about what should or should not be done during the experiment. Keep concentrated on your own computer screen and keep silence for the whole session. If you have a question, feel free to raise your hands and ask anytime.

You will see on your desk an envelop containing 32€in cash, rewarding you for your participation. This sum is yours to keep. During the experiment, you will have the opportunity to buy some products. It is important that you understand why and how you will be able to shop.

We ask you to shop because we want to observe your actual shopping behavior and not just your shopping intentions. As behavioral scientists, we know that intentions can differ from actual purchasing decisions. Our guiding principle is simple: put you in a real shopping situation. You will have to buy some products, that you will have chosen yourself. In no circumstance our research group makes profits out of the eventual sales. The products will be sold at supermarket prices, as recorded by us on October 5th, 2016. You will discover these prices later in the experiment.

This study is composed of several phases. After the start of each phase, instructions for the following phase will be given. A new phase will start only after all participants have completed the previous phase. This session will not exceed one hour and thirty minutes.

Phase 1: food shopping

During this first phase, you will be shopping for food. You will have to compose your basket in the same way you would do in a supermarket. In order to observe your real shopping behavior, your decisions will have real implications (namely, the purchase of some products).

Imagine you are in the only shop open today. Please do not take into account the shop you usually go to. You will have to shop for the coming 2 days (no week-end) for your household, choosing along the products available in our shop. Please think that your cupboards at home are empty. Think that they contain only some basic products: butter, flour, oil, spices, coffee, tea, seasoning, water, alcohol, wine, sugar, vinegar, sauces. These products will not be on sale here.

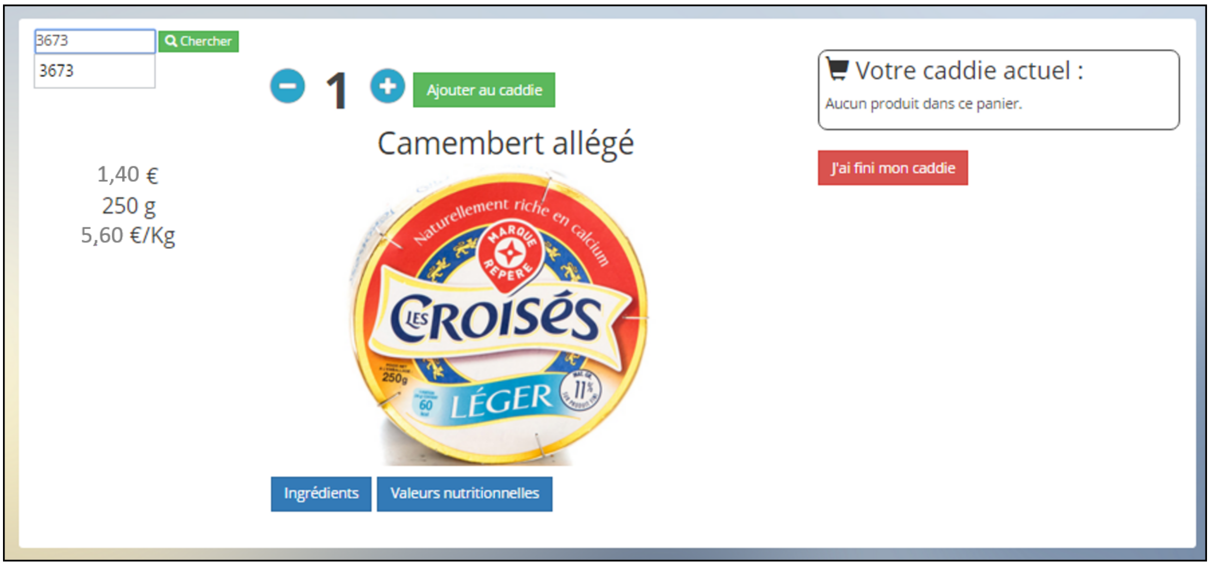
In order to fill your shopping basket, you will be able to choose among 290 food products, presented to you in the paper catalogue that was just handed over to you[6]. Each page of the catalog contains products of a specific category. Each category is composed of 3 to 9 products. A table of contents is available at the beginning of the catalog.

On the screen, you will see this page:

[6][see AppendixD for a sample page of the catalog in each treatment]

In order to select a product and display it on the screen, you can use the barcode scanner available on the desk in front of you. Pass the scanner over the barcode of the product and press the button. Should you meet any difficulty with the scanner, you can manually type the 4-digit product code in the search bar and click on 'search'.

Once you selected the product, it will be shown on screen as seen in a supermarket shelf, showing its front of pack.



You will see price, price per kilogram or liter, and its weigh or volume. The prices shown have been recorded in a Grenoble area supermarket on October 5th, 2016. You can click on the buttons "Ingredients" and "Nutritional values" to access the information that you normally find on the back of pack.

If you click on "Ingredients" you will access the substances or food additives used in the production of that food item. These ingredients are present in the final product in a more or less transformed state. On your screen (as on backs of pack), ingredients are ranked according to their relative importance in the product recipe. As the lax dictates, the ingredients potentially resulting in intolerances or allergies must be highlighted. Here, they will be written in ALL CAPS.

If you click on "Nutritional values" on your screen (as on backs of pack) you will find the nutritional values, expressed for 100 grams of product. These values will not be available for some products (fruits, vegetables...), as for these products the law does not impose their display.

Once you have selected the product you are free to choose a quantity, as you would do in normal shopping. Once you have chosen its quantity, you can add it to the basket by clicking on "add to basket". Even after you have added the product to your basket you can change its quantity using the buttons "+" and "-" and then confirming with the button "Change quantity". You can remove the product from the basket clicking on "Remove from basket". Once you have finished to fill your basket, you can click on "I finished my basket".

Your decisions will have real implications. In order to observe your actual shopping behavior, your decisions will have real implication (namely, the purchase of some products). This will be done in the following way:

- in another room we store about one quarter of the 290 products displayed in the catalog. In order not to influence your choices, you will not know which products we store or not.
- at the end of the experiment we will identify the products of your basket that we stock; in all likelihood, a quarter of the products you chose.

- you will have to buy these products. For example, if your basket contains 20 products, you will likely buy 4 or 5 of them, for a cost of some Euro. You will leave the room with these products. Please note that the products are sold at the prices shown on the catalog.

Any questions?

[During the shopping task, the overhead projectors showed a reminder of the main features of the task. This is its text:] Reminder: you shop for the next 2 days for your household. YOu are free to choose the products and quantity you prefer; to spend the amount you prefer. You can retrieve the informations that you would find on the backs of packs, namely ingredients and nutritional information, on screen.

Phase 2: food shopping, a second basket

Your task. We ask you now to compose a second shopping basket. We remind you that you have access to the nutritional values and ingredients of the products by clicking on the corresponding buttons below the product image.

[*all treatments but Baseline:* we distribute now a new catalog. It contains the same products. Each of the products now features a nutritional label. We will explain the label next.]

Monetary implications. Just one of the two baskets will be taken into account for the product purchases at the end of the experiment. The binding basket will be randomly drawn.

Why a second basket? In order to help you in making better health choices when shopping for food, the Health Ministry is planning to introduce a graphical representation aimed at simplifying the display of nutritional values. Within this plan, our study has the aim of testing the impact of different graphical representation systems. The principle of our test consists in adding (or not) on the products' front of pack a graphical representation that might be understood with a quick look.

[*Baseline:*] We test with you the impact of repetition absent any graphical representation. You belong to the **benchmark group**.

[*All other treatments:*] We test with you the impact of the [*name of the label*] system.

[*For all treatments but baseline, at this point we distributed to the subjects one-page description of the labeling scheme applied in their respective treatment. We projected the content of this leaflet overhead on two slides, and read aloud its content. Any questions were answered before moving on to the second shopping basket. The one-page descriptions are displayed next*]

Any questions?

[During the shopping task, the overhead projectors showed a reminder of the main features of the task. This is its text:] Reminder: you shop for the next 2 days for your household. You are free to choose the products and quantity you prefer; to spend the amount you prefer. You can retrieve the informations that you would find on the backs of packs, namely ingredients and nutritional information, on screen.

Ending

Phase 3: questionnaire and end of the experiment.

You will first be given two short questionnaires to fill in. Once the questionnaire filled, you will be asked to fill in the receipt for payment. Once everyone will be finished, an experimenter will call your code, and you will proceed to another room to purchase the food products of the randomly chosen basket.

Thank you for your participation.

▪ Qu'est ce que le Nutri Couleurs?

Cet étiquetage est une représentation graphique mise au point par des experts en nutrition qui permet de connaître :

- l'apport en nutriments à limiter dans notre alimentation, en grammes (g) par portion,

- la contribution de ces mêmes nutriments aux apports quotidiens de référence (en %).

Une couleur est associée à chacun de ces nutriments en fonction de leurs apports par portion (si la portion est inférieure à 100g, les couleurs sont données pour 100g de produit) :

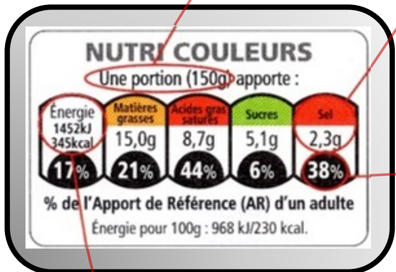
- 🔴 Si la teneur est élevée
- 🟠 Si la teneur est moyenne
- 🟢 Si la teneur est faible
- L'énergie est indiquée en blanc

▪ Ce qu'il faut comprendre

Nutri Couleurs permet de comparer les apports nutritionnels des produits par portion.

Une indication de portion

Produit 1

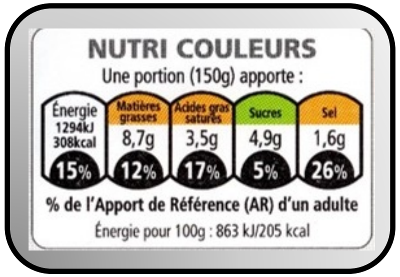


Une portion du produit 1 contient 2,3 g de sel.

Une portion du produit 1 représente 38% des apports de référence en sel sur 1 journée.

Une portion du produit 1 apporte 345 Kcal.

Produit 2



Les couleurs indiquent par exemple que pour une portion du produit 2, la teneur en sel est moyenne et la teneur en sucres est faible.

Les couleurs indiquent aussi que les teneurs en acides gras saturés et en sel du produit 2 sont inférieures à celles du produit 1 (orange pour le produit 2 contre rouge pour le produit 1).

Figure C.4: One page label explanation: NutriCouleur

Qu’est ce que le Nutri-Score ?

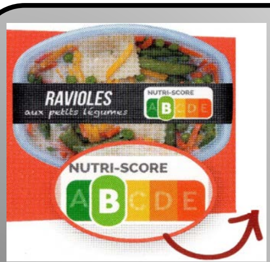
Cet étiquetage vous informe sur la qualité nutritionnelle d’un produit grâce à une lettre associée à une couleur. Pour chaque produit, les chercheurs en santé publique ont mis au point un score nutritionnel qui prend en compte la teneur en éléments à favoriser (fibres, protéines, fruits et légumes) et en éléments à limiter (énergie, acides gras saturés, sucre et sel).

Chaque produit est positionné sur une échelle à 5 niveaux allant du produit :

- Le plus favorable sur le plan nutritionnel (classé A / vert foncé)
- Au produit le moins favorable sur le plan nutritionnel (classé E / orange foncé)

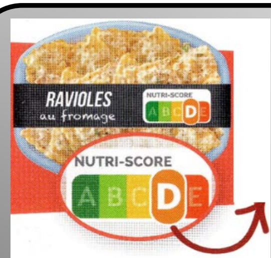
Ce qu’il faut comprendre

Produit 1



Le produit 1 est classé B car il contient une quantité intéressante de légumes et de fibres, mais aussi parce qu’il contient peu d’éléments à limiter comme le sel ou les acides gras saturés.

Produit 2



Le produit 2 est classé D car il est plus calorique que le produit 1, contient davantage de sel et d’acides gras saturés et moins de légumes et de fibres.

1 logo et 5 niveaux différents de qualité nutritionnelle des produits :

Figure C.5: One page label explanation: NutriScore

Qu'est ce que le Nutri Mark?

Cet étiquetage, déjà testé en Australie et en Nouvelle-Zélande, vous informe sur la qualité nutritionnelle d'un produit de deux façons :

- De façon synthétique : chaque produit est globalement positionné par un chiffre allant de 0,5 à 5 associé à des étoiles. Les produits les plus favorables sur le plan nutritionnel ont 5 étoiles. Les produits les moins favorables sur le plan nutritionnel ont 0,5 étoiles.

- De façon détaillée : les données sur l'énergie, les matières grasses, les acides gras saturés, les sucres et le sel sont fournies pour une portion et exprimées d'une part en valeur absolue et d'autre part en pourcentage des apports de référence.

Ce qu'il faut comprendre

Le nombre d'étoiles est calculé en prenant en compte la teneur en éléments à favoriser (fibres, protéines, fruits et légumes) et en éléments à limiter (énergie, acides gras saturés, sucre et sel).

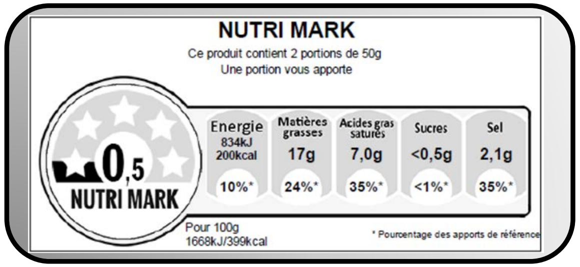
Produit 1

Les informations sont données pour une portion et indiquent combien de portions sont contenues dans le paquet

Pour chacun des nutriments (énergie, matière grasse, acides gras saturés, sucre et sel), l'apport d'une portion est indiqué en grammes (g).

La contribution de ces mêmes nutriments est rapportée aux apports quotidiens de référence (en %)

Produit 2



Le nombre d'étoiles indique que le produit 1 est plus favorable sur le plan nutritionnel que le produit 2. La contribution en sel de la portion du produit 2 est plus importante que celle du produit 1.

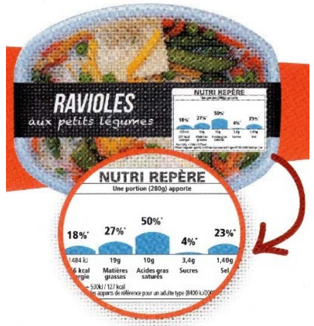
Figure C.6: One page label explanation: NutriMark

Qu’est ce que le Nutri Repère?

Cet étiquetage est une représentation graphique mise au point par des experts en nutrition qui permet de visualiser en % la contribution d’une portion d’aliment aux apports quotidiens de référence pour un adulte.

Ainsi Nutri Repère vous indique ce qu’une portion d’aliment apporte en énergie, ainsi que pour les nutriments suivants par rapport à ce qu’il est recommandé de consommer :

- Matières grasses
- Acides gras saturés
- Sucres
- Sel

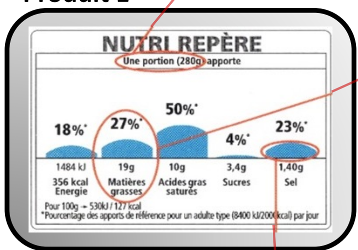


Ce qu’il faut comprendre?

Nutri Repère permet de comparer les apports nutritionnels des produits par portion.

Les informations sont données pour une portion

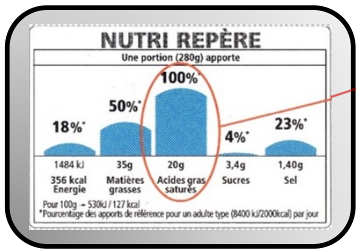
Produit 1



Une portion du produit 1 contient 19 g de matières grasses, ce qui représente 27% des apports recommandés sur 1 journée (soit plus d’un quart)

Plus la colonne est haute, plus l’apport (ici en sel) est élevé

Produit 2



Une seule portion du produit 2 permet d’atteindre le seuil à ne pas dépasser en acides gras saturés sur une journée (soit 100%)

Une portion du produit 2 contient 2 fois plus d’acides gras saturés que le produit 1

Figure C.7: One page label explanation: NutriRepère

- **Qu'est ce que le repère alimentaire SENS ?**

Cet étiquetage a été mis au point par des chercheurs nutritionnistes qui ont réalisé pour chaque produit un calcul prenant en compte :

- Les nutriments dont la consommation est **à favoriser** (fibres, oméga 3, fruits et légumes, ...)
- Les nutriments dont la consommation est **à modérer** (acides gras saturés, sucre, sel,...)

Après calcul, une indication de fréquence de consommation est attribuée à chaque produit.

- **Ce qu'il faut comprendre**

Produit 1

Le produit 1 peut être consommé souvent car il contient une quantité intéressante de légumes et de fibres, mais aussi parce qu'il contient peu d'éléments dont la consommation est à limiter comme le sel ou les acides gras saturés

Produit 2

Le produit 2 peut être consommé régulièrement mais en petite quantité car il est assez calorique, contient une quantité importante de sel et d'acides gras saturés et moins de légumes et de fibres.

4 représentations indiquent la fréquence de consommation suggérée pour chaque produit

Du plus fréquent — au plus modéré

Figure C.8: One page label explanation: SENS

AppendixD. Product catalog screenshots

Figure D.9

Snacks

Cacahuètes grillées très pauvres en sel

1,00€
200g
5,00€/Kg
2872

Cacahuètes délicatement salées

1,09€
220g
4,95€/Kg
2849

Chips de maïs nature

2,49€
230g
10,83€/Kg
2859

Chips paysannes nature

2,48€
300g
8,27€/Kg
2898

Soufflés de maïs goût cacahuète

2,25€
250g
9,00€/Kg
2851

Biscuits apéritif à l'emmental

0,52€
50g
10,40€/Kg
2810

30

Figure D.10

Snacks

Cacahuètes grillées
très pauvres en sel

1,00€
200g
5,00€/Kg
2872

NUTRI MARK 5,0

Cacahuètes délicatement salées

1,09€
220g
4,95€/Kg
2849

NUTRI MARK 4,0

Chips de maïs nature

2,49€
230g
10,83€/Kg
2859

NUTRI MARK 3,0

Chips paysannes nature

2,48€
300g
8,27€/Kg
2898

NUTRI MARK 2,0

Soufflés de maïs goût cacahuète

2,25€
250g
9,00€/Kg
2851

NUTRI MARK 2,0

Biscuits apéritif à l'emmental

0,52€
50g
10,40€/Kg
2810

NUTRI MARK 0,5

30

Figure D.11

Snacks
Cacahuètes grillées très pauvres en sel
1,00€
200g
5,00€/Kg
2872
Carrefour
NUTRI REPERE
Cacahuètes délicatement salées
1,09€
220g
4,95€/Kg
2849
Benenuts
Cacahuètes
NUTRI REPERE
Chips de maïs nature
2,49€
230g
10,83€/Kg
2859
Doritos
NUTRI REPERE
Chips paysannes nature
2,48€
300g
8,27€/Kg
2898
Lay's
Chips Paysannes
Nature
NUTRI REPERE
Soufflés de maïs goût cacahuète
2,25€
250g
9,00€/Kg
2851
menguy's
NUTRI REPERE
Biscuits apéritif à l'emmental
0,52€
50g
10,40€/Kg
2810
BeLIN
Monaco
à l'emmental
NUTRI REPERE
30

Figure D.12


Snacks
Cacahuètes grillées très pauvres en sel
1,00€
200g
5,00€/Kg
2872
Carrefour
NUTRI-SCORE
A B C D E
Cacahuètes délicatement salées
1,09€
220g
4,95€/Kg
2849
Benenuts
Cacahuètes
NUTRI-SCORE
A B C D E
Chips de maïs nature
2,49€
230g
10,83€/Kg
2859
Doritos
NUTRI-SCORE
A B C D E
Chips paysannes nature
2,48€
300g
8,27€/Kg
2898
Lay's
Chips Paysannes
NUTRI-SCORE
A B C D E
Soufflés de maïs goût cacahuète
2,25€
250g
9,00€/Kg
2851
menguy's
NUTRI-SCORE
A B C D E
Biscuits apéritif à l'emmental
0,52€
50g
10,40€/Kg
2810
BeLIN
Monaco
à l'emmental
NUTRI-SCORE
A B C D E
30

Figure D.13

Snacks

Cacahuètes grillées très pauvres en sel

1,00€
200g
5,00€/Kg
2872

REPÈRE ALIMENTAIRE
RÉGULIÈREMENT EN PETITE QUANTITÉ

Cacahuètes délicatement salées

1,09€
220g
4,95€/Kg
2849

REPÈRE ALIMENTAIRE
OCCASIONNELLEMENT OU EN PETITE QUANTITÉ

Chips de maïs nature

2,49€
230g
10,83€/Kg
2859

REPÈRE ALIMENTAIRE
RÉGULIÈREMENT EN PETITE QUANTITÉ

Chips paysannes nature

2,48€
300g
8,27€/Kg
2898

REPÈRE ALIMENTAIRE
OCCASIONNELLEMENT OU EN PETITE QUANTITÉ

Soufflés de maïs goût cacahuète

2,25€
250g
9,00€/Kg
2851

REPÈRE ALIMENTAIRE
OCCASIONNELLEMENT OU EN PETITE QUANTITÉ

Biscuits apéritif à l'emmental

0,52€
50g
10,40€/Kg
2810

REPÈRE ALIMENTAIRE
OCCASIONNELLEMENT OU EN PETITE QUANTITÉ

30

Figure D.14