Effects of front-of-pack labels on the nutritional quality of supermarket food purchases: evidence from a large-scale randomized controlled trial

Pierre Dubois, Paulo Albuquerque, Olivier Allais, Céline Bonnet, Patrice Bertail, Pierre Combris, Saadi Lahlou, Natalie Rigal, Bernard Ruffieux, Pierre Chandon

EFFECTS OF FRONT-OF-PACK LABELS ON THE NUTRITIONAL QUALITY OF SUPERMARKET FOOD PURCHASES: EVIDENCE FROM A LARGE-SCALE RANDOMIZED CONTROLLED TRIAL

Pierre Dubois*
Toulouse School of Economics, France

Paulo Albuquerque*
INSEAD, France

Olivier Allais*
INRAE, France

Céline Bonnet*
Toulouse School of Economics, France

Patrice Bertail
University Paris Nanterre, France

Pierre Combris
INRAE, France

Saadi Lahlou
London School of Economics and Political Science, UK

Natalie Rigal
University Paris Nanterre, France

Bernard Ruffieux
Grenoble INP, INRAE, France

Pierre Chandon*
INSEAD, France

Note: * contributed equally. † corresponding author: pierre.chandon@insead.edu. The authors are grateful to the Editor, Associate Editor, and the Reviewers for their helpful comments and suggestions. They would like to thank Noël Renaudin, the Chairman of the scientific committee, as well as Benoît Vallet and Christian Babusiaux, the co-heads of the steering committee of the evaluation of graphical nutrition labels in real-life conditions launched by the French Minister of Social Affairs and Health. They also thank Daniel Nairaud, the managing director the Fonds Français Alimentation Santé (FFAS), for his operational leadership of the project, Michel Chauliac (Ministry of Social Affairs and Health), Sandrine Raffin (LinkUp), and Pascale Hébel (Credoc). Funding was provided by CNAMTS, FFAS, and the Ministry of Social Affairs and Health.

EFFECTS OF FRONT-OF-PACK LABELS ON THE NUTRITIONAL QUALITY OF SUPERMARKET FOOD PURCHASES: EVIDENCE FROM A LARGE-SCALE RANDOMIZED CONTROLLED TRIAL

ABSTRACT

To examine whether four pre-selected front-of-pack nutrition labels improve food purchases in real-life grocery shopping settings, we put 1.9 million labels on 1,266 food products in four categories in 60 supermarkets and analyzed the nutritional quality of 1,668,301 purchases using the FSA nutrient profiling score. Effect sizes were 17 times smaller on average than those found in comparable laboratory studies. The most effective nutrition label, Nutri-Score, increased the purchases of foods in the top third of their category nutrition-wise by 14%, but had no impact on the purchases of foods with medium, low, or unlabeled nutrition quality. Therefore, Nutri-Score only improved the nutritional quality of the basket of labeled foods purchased by 2.5% (-0.142 FSA points). Nutri-Score's performance improved with the variance (but not the mean) of the nutritional quality of the category. In-store surveys suggest that Nutri-Score's ability to attract attention and help shoppers rank products by nutritional quality may explain its performance.

Keywords: Nutrition; labelling; supermarket; RCT; food; Field experiment; policy.

To promote healthier eating, regulatory authorities worldwide are encouraging the use of labels that provide simplified nutrition information on the front of the pack (FOP) in addition to the mandatory calorie and nutrition information already provided on the back. The European Union, for example, recently introduced a voluntary scheme for manufacturers to put graphical information about nutritional product quality on the front of the pack (Regulation EU 2011). However, there is disagreement about whether FOP nutrition labels truly improve food purchases and, if they do, about which specific label regulators should endorse, and companies adopt (Askew 2019).

In May 2018 Mondelēz, Mars, Nestlé, PepsiCo, Coca-Cola, and Unilever backed the Evolved Nutrition Label, a nutrient-specific system inspired by the British "Multiple Traffic Light" label but, a few months later, Mars and Nestlé withdrew their support and, in June 2019, Nestlé announced its support for Nutri-Score. In France, four FOP nutrition labels competed for governmental endorsement in 2016: two were analytic systems which provide information analysis for each nutrient: Nutri-Couleurs, an adaptation of the British traffic-light system, and the mono-colored Nutri-Repère, which was backed by industry groups. The other two provide a single summary indicator: Nutri-Score, which has started to be used by some manufacturers and retailers in France, Spain, Belgium, and Germany, and SENS, which was developped for French retailers and promoted by French, German, and Belgian food industry groups (Michail 2015).

Despite the importance of this issue, evidence of the effects of FOP nutrition labelling on food purchases in natural settings is sparse. A recent review and meta-analysis (Ikonen et al. 2020) concluded that, "Although most FOP nutrition labels help consumers identify healthier options within product sets, this does not directly translate into other measures of effectiveness, which show smaller effects and greater variability between different label types and product categories".

It is still unclear whether FOP nutrition labels in general—and the four labels seeking government approval in particular—improve the nutritional quality of purchases made in real-life grocery shopping settings. The literature is also silent about whether nutritional improvements come from an increase in purchases of foods with higher nutritional quality in their category, a fall in purchases of options with lower nutritional quality, or both. Moreover, the impact of FOP labels on purchases of unlabeled foods remains unknown. This is a critical issue as current regulations prevent retailers or governments from forcing manufacturers to adopt a FOP nutrition label. Studying the purchases of unlabeled products also allows us to test whether the mere presence of FOP nutrition labels in one category changes the consumer's decision process or preferences (e.g., for health versus taste) rather than simply providing information about the nutritional quality of labeled foods. Finally, disagreement exists about the role of the design, notably whether FOP nutrition labels should provide a summary vs. nutrient-specific scores, show the range of possible grades on each label, and should be color-coded.

To answer these questions, the French health authorities asked the study authors to help determine which of the four FOP labeling system mentioned earlier, if any, should receive official approval. Meanwhile, industry associations and individual retailers and manufacturers are waiting for evidence from real-life grocery shopping studies to decide which FOP system to adopt, if any. We therefore conducted a randomized controlled trial (RCT) to test the effects of these four labeling systems in sixty supermarkets over a ten-week period. To shed light on some of the mechanisms underlying the effectiveness of the labels, we also conducted a survey of shoppers in the test and control stores, both before and during the experiment. The large scale of the study, which involved adding millions of labels and tracking millions of purchases, provided the large

sample size required to obtain precise estimates for effect sizes that are known to be small (Ikonen et al. 2020).

In addition to examining whether FOP labels improve the nutritional quality of supermarket food purchases, the randomized-controlled-trial allowed us to test whether they do so by influencing the purchase of foods that were labeled high or low nutritional quality, as well as to assess their effects on other products in the category without an FOP nutrition label. The field experiment also allowed us to assess whether the effects of nutrition labels vary across four product categories that differ in terms of the mean and variance of their nutrition quality. Finally, by comparing four different FOP nutrition labels, this study examines the effects of the design choices, such as providing nutrient-specific information or a summary score.

Front-of-package nutrition labeling

FOP nutrition labels provide summary, simplified information about the calorie and/or nutrient content of foods on the front of the pack, sometimes augmented with evaluative symbols or color coding (McGuire 2012; Newman et al. 2018). They should not be confused with various other FOP information such as warning labels (e.g., "contains Sulfites"), health or structure/function claims (e.g., "calcium helps create strong bones''), unregulated food claims (e.g., "natural"), or the regulated nutrient claims such as the "low-fat", "no trans-fat" or "extra antioxidants" studied by Kiesel and Villas-Boas (2013) or by Belei et al. (2012). They also differ from "reductive" nutrient-specific labels such as "Facts Up Front" labels which simply reproduce a subset of the nutrition information available on the back of the package without further interpretation.

FOP nutrition labels are related to, but qualitatively different from the various initiatives developed to provide calorie or nutrition information on the menus of restaurants. Although the

goal is similar, the context is not. Unlike in restaurants, foods sold in supermarkets are pre-packaged (with a few exceptions that fall outside the scope of labeling laws, such as fresh products in bulk) and hence already provide ingredients, calorie, and nutrition information on the back of the pack. People make purchase decisions in supermarkets, whereas they make consumption decisions in restaurants. For these reasons, studies conducted in restaurants are suggestive and can be used for hypothesis generation. However, evidence that adding calorie labeling to menus can work in restaurant settings (Bollinger, Leslie, and Sorensen 2011; Bleich et al. 2017) cannot be directly extrapolated to the effects of FOP nutrition labels on supermarket purchases.

Effects of nutrition labels in laboratory settings

Consumers have favorable attitudes towards FOP nutrition labels and believe that they help them make healthier food choices (Cadario and Chandon 2019; Feunekes et al. 2008; Grunert and Wills 2007; Hawley et al. 2013). Reviews (Kiszko et al. 2014) and meta-analysis (Ikonen et al. 2020) shows that there is also evidence that FOP nutrition labels help consumers identify healthier products and increase intentions to choose healthier foods in online surveys (e.g., Andrews, Burton, and Kees 2011; Roberto et al. 2012), although there are many counter-examples of studies finding no effects on purchase intentions (e.g., Gorski Findling et al. 2018).

Analyses of the effects of FOP labels on actual food choices or consumption found more limited effects. For example, the recent meta-analysis by Ikonen et al. (2020) found small effect sizes for interpretive (nutrient-specific) and summary FOP labels on the choice of healthier options (respectively, Fisher's z back-transformed estimated correlation = 0.079 [0.064, 0.094] and 0.023 [0.008, 0.039]) and smaller and statistically insignificant effects on actual consumption (respectively, 0.036 [-0.022, 0.093] and 0.006 [-0.030, 0.041]). However, their meta-analysis relied mainly on laboratory studies.

A recent laboratory study by Crosetto et al. (2020), which was not included in Ikonen's meta-analysis, provides the strongest evidence that FOP nutrition labels can significantly influence consequential food choices in laboratory settings. Grounded in the paradigm of experimental economics (e.g., Muller et al. 2017), respondents were asked to shop for their households using a paper catalog of 290 products, including the full denomination of the product, color pictures, prices, bar-code, but without front-of-pack nutrition label. With a bar-code reader, the shoppers could access the information typically available when shopping online (e.g., the list of ingredients as well as nutrition information). To make the experiment incentive compatible, the participants were informed that they would have to buy a randomly determined subset of the products that they chose. After purchasing food for a day once, participants were unexpectedly asked to make a second "shopping trip" from the same catalog, but now with a front-of-pack nutrition label implemented on all products (or the same catalog without any nutrition label for the control treatment). They tested five labels, including the four that we tested and NutriMark (a French version of Australia and New Zealand's Health Star Ratings). They found that all labels significantly improved the nutritional quality of the shopping baskets purchased, in the following order (from most to least effective): Nutri-Score, NutriMark, Nutri-Couleurs, Nutri-Repère, and SENS.

Effects of nutrition labels in field settings

Evidence that FOP nutrition labels work in a laboratory setting does not mean that they will necessarily work in the field. For example, a meta-analysis of the effects of menu calorie labeling in restaurants found that it led to an 18-calorie reduction per meal in laboratory and online studies but to an insignificant 8-calorie reduction in studies conducted in restaurants (Long et al. 2015).

As reviewed by Hawley et al. (2013), few studies have examined the effects of FOP labeling on food purchases and those which have, focused on a few products and short periods of time made by people in supermarkets. Gaigi et al. (2015) placed an ad-hoc summary FOP label (Vita+) in two French supermarkets and found no effects on food purchases. In a study encompassing 100 stores, and eight categories over six months, Nikolova and Inman (2015) found that the introduction of the summary Nuval system, which grades each food on a 100-point scale, led people to switch to higher-scoring foods. However, that label, which was briefly used by one North American retailer before being abandoned, is unrepresentative of most FOP labels. First, it was added by the retailer to the shelf tag, and not on the packages themselves. Second and more importantly, it was available for all the products in the category, whereas EU laws, for example, require the agreement of the manufacturer. Mhurchu et al. (2017) used an RCT to deliver "traffic light" labels, Health Star Rating labels, or nutritional information via a smartphone app to 1,357 shoppers. They found no significant effect on the nutritional quality of grocery purchases over a four-week period. It remains unclear whether these results would hold in a larger-scale trial over a longer period, and when nutrition information is displayed directly on packages. A recent meta-analysis of field studies (Cadario and Chandon 2020) found similarly low effect sizes for both descriptive and evaluative (interpretive) labeling (respectively, $d = 0.10$, SE=0.07 and $d = 0.17$, SE = 0.06). Unfortunately, this study did not differentiate between food selection and consumption, included warning labels and, like Ikonen's (2020) meta-analysis, relied primarily on studies conducted in restaurant settings.

In short, we still do not know whether established FOP nutrition labels, like the British Traffic-Light system, the French Nutri-Score, or alternatives backed by research and industry groups significantly improve the nutritional quality of foods purchased over multiple shopping

trips in real-life grocery shopping conditions and across a wide variety of brands and product categories. However, absence of evidence is not evidence of absence and, based on the positive of recent laboratory settings, all four labels may have a small but positive effect on the nutritional quality of supermarket food purchases in real-life grocery shopping conditions.

Effects of nutrition labels across brands

Another important question is whether the effects of FOP nutrition labels come from boosting the purchases of brands or categories with high nutrition value or from hurting the purchases of brands or categories with low nutrition value. Ikonen's (2020) meta-analysis found a positive and statistically significant effect of FOP nutrition labels on intentions to purchase virtue food products such as almonds but a null effect for vice foods like cakes. Similarly, Nikolova and Inman (2015) found that the roll-out of Nuval had stronger effects in healthier product categories. They also hypothesized that Nuval would have the largest effect in categories with the widest within-category variance in nutrition quality, which provide greater opportunity to switch to healthier alternatives, but this hypothesis was not supported by their data.

To date, no field study has examined whether FOP labels has a different impact on the purchases of foods with a high, low, or unknown (unlabeled) nutritional quality within their product category. Crosetto et al. (2020) found that summary labels like Nutri-Score and SENS had larger effects on the purchases of brands carrying the most extreme labels (e.g., green or red compared to yellow). However, they did not examine whether the effects of analytic labels varied according to the nutritional quality of the food. Also, because all the brands in their study were labeled, they did not examine the effects on unlabeled foods. This is an important consideration because EU regulations allow firms to opt out of front-of-package nutrition labeling (only back-of-package nutrition information is compulsory).

Based on these findings, FOP nutrition labels may have stronger effects in product categories with a high mean and a high variance in nutrition quality. Given the low effects of nutrition labels overall, this would imply very small to null effect for categories with low mean and variance in nutritional quality, at least if there are few reports from one category to another. By extending this reasoning to within-category nutrition differences, nutrition labels could increase the purchases of foods with a higher nutritional quality than the average of their category but have no effects for foods with a lower than average nutritional quality. Finally, given the limited effects of FOP labels on the purchases of labeled products, they are unlikely to influence the purchases of those foods in the category without a nutrition label, whose nutritional quality is not immediately knowable.

Method

Stimuli

Four labeling systems were pre-selected following a comprehensive consultation process involving national research institutes, food manufacturers and retailers, the French national health insurance administration, consumer and patient advocacy groups, and the Ministries of Health, Agriculture, and Consumer Affairs. These labeling systems were chosen based on scientific evidence and were backed by industry and consumer associations. Each system was developed independently by competing research teams in accordance with a research and innovation contest. These contests are an effective and increasingly popular tool for encouraging innovation because they incentivize each competitor to design the best possible solution and increase the diversity of solutions (Boudreau, Lacetera and Lakhani 2011). On the other hand, the lack of central coordination is a disadvantage in terms of the ability to pinpoint why one solution outperformed another. In any case, the costs involved in conducting a large-scale randomized controlled trial in

natural grocery shopping conditions allow for a limited number of experimental conditions, which prevents the identification of the effects of each design characteristic.

We display a graphical representation of each labeling system in competition in Figure 1. SENS, in panel (a), provides a summary evaluation of the nutritional quality of the food (Maillot et al. 2016). It is based on an algorithm adapted from the nutrient profiling system SAIN/LIM (Darmon et al. 2009), which scores nutritional quality based on the relative quantity of favorable ("SAIN") and unfavorable ("LIM") nutrients. This system classifies foods into four categories represented by green, orange, blue, or purple inverted pyramids. Each level has a label specifying the consumption frequency or portion size (e.g., the purple label says "occasionally or in small quantity"). Only the proper pyramid is visible on a given product's label.

--- Insert Figure 1 here ---

The Nutri-Score labeling system in panel (b) below provides a summary evaluation based on the amount of positive and negative nutrients (Julia and Hercberg 2017). It is adapted from the British Food Standards Agency's nutrient profiling system. It grades products on a five-point scale, from A to E, and displays the assigned grade with a larger font on a sliding scale showing the five grades, colored from green to yellow to dark orange, identifying the relative nutritional quality on this scale.

Nutri-Repère, panel (c), is an analytic label that displays the amount of energy, fat, sugars, and salt per suggested portion, as well as their contribution—in percentage and as a blue bar graph—to the Guideline Daily Amounts, which represents the daily nutritional requirements of the average adult consumer. It is backed by industry bodies (Nutri-Repère 2015). We refer to it as mono-color rather than monochrome because the level and percentage information are in black, not in the same color as the bar chart, the key visual element. Finally, Nutri-Couleurs, panel (d),

is an analytic label that provides the same information as Nutri-Repère but, like the British Multiple Traffic Light label on which it is based, it color-codes each nutrient as red, amber, and green based on thresholds set by the British Food Standards Agency (2007).

Finally, Nutri-Couleurs, in panel (d), is an analytic label that provides the same information as Nutri Repère, but like the British Multiple Traffic Light label on which it is based, it color-codes each nutrient as red, amber, and green based on thresholds set by the British Food Standards Agency (2007).

According to the classification of FOP nutrition labels (Newman et al. 2018; Ikonen et al. 2020), all of these are "interpretive" because they repeat some of the descriptive nutrition information present on the back of the package but enhance it with graphical symbols (e.g., colors, bar charts, more or less filled triangles) that help to convey the nutritional quality of each nutrient or of the food product as a whole. This is true even for Nutri-Repère, the least "enhanced" label, where a bar chart allows consumers to see at a glance the contribution of the product to daily nutritional requirements.

Procedure

From September 26 to December 4 of 2016, we placed FOP labels on food products in 40 randomly selected supermarkets in France, 10 supermarkets per labeling system. In addition, 20 control supermarkets were randomly chosen in which products had no additional FOP labeling. These 60 supermarkets belonged to three of the largest retail chains in France. The protocol was made publicly available on the website of the French Ministry for Solidarity and Health in April of 2016 (Renaudin et al. 2016) and the intervention was authorized by ministerial decree. The field experiment was registered at the International Standard Randomized Controlled Trials Number

Registry (ISRCTN 58212763). The operational settings are summarized in Figure A1, in the appendices.

Before the randomization procedure started, the number of stores in the treatment and control groups was chosen based on implementation costs (hence the higher number of stores in the control group) and market conditions (e.g., the market share of each retailer chain). Each treatment condition included four supermarkets from Carrefour (for a total of 16 stores), three from Simply Market (total: 12 stores), and three from Casino (total: 12 stores), while the control condition included eight supermarkets from Carrefour, six from Simply Market, and six from Casino (total: 20 stores).

The random sample of 60 stores included in the study, out of the universe of the 174 French supermarkets of these three chains, was obtained through several steps. First, we categorized the universe of stores into groups characterized by two attributes: retailer chain (Casino, Simply Market, and Carrefour) and whether the store was in a privileged or underprivileged geographical area. The latter was operationalized as the store's catchment area being in the bottom two quintiles in terms of the proportion of unskilled laborers, the only measure that was available to us for all stores. This was done to ensure that we had enough shoppers from lower socio-economic status, because prior research has shown that nutrition labeling tends to have lower effects for this population (e.g., Elbel et al. 2009). This resulted in six groups, three levels by two levels. Second, for each of the six groups, the stores were a-priori assigned a number from 1 to N, where N was the total number of stores in that group. For example, for Carrefour stores in underprivileged areas, it was decided to have 12 stores (with 2 for in each condition and 4 in the control group) out of 22 possible stores, and so N was 22 for that group and each store was assigned a number from 1 to 22. For each treatment or control conditions, we then obtained random numbers from 1 to N using

the www.randomizer.org site. In the Carrefour example, we obtained 12 numbers, 2 random numbers for each treatment condition and 4 for the control group, from the set of 1 to 22 (e.g., we obtained numbers 4, 12, 15, and 21 for the control condition). We then looked up each of these randomly drawn numbers to the a-priori assigned numbers given to each store, hence obtaining the set of stores assigned to the study for each condition.

Consumers were informed of the local intervention in each treatment supermarket through leaflets and displays that were rigorously identical, except for the explanation of each labeling system. Shoppers were informed about the labels present in their store (but not about the other types of labeling systems) in three ways. First, leaflets were made available in each store, describing the labeling system, the four product categories where it could be found, and two hypothetical examples (one of a high and one of a low nutritional quality food). The leaflets also described the goal of the intervention, the sponsors (French Government, National Health Care, French Fund for Food and Health) and the regions in which the study was conducted. The four leaflets were identical in structure and design and the text describing each labeling system was validated by a scientific committee to ascertain its validity and fairness (see exhibit A2 for an example of leaflet). Second, information about the label was made available in the aisles themselves thanks to self-standing signs ("totems"). As can be seen in Figure A3 in the appendices, some of these displayed included additional leaflets. Third, stickers were manually placed on the front of each of the products themselves (Figure A4). After we had checked that all the communication was put in place properly at the start of the intervention, responsibility for in-store signage was transferred to the stores and no additional measures were taken to ensure that the signs remained visible or would be replaced if they went missing. This was done on purpose, to replicate what would happen when a national labelling system is put in place by the retailers.

The operational management of the study was done by the French Fund for Food and Health (FFAS), an organization jointly created by the French Nutrition Institute (Institut Français pour la Nutrition) and the National Association of Food Industries (Association Nationale des Industries Alimentaires), with the aim of developing a partnership between the academic community and economic actors, for a better service of public health. The FFAS raised both private and public funding for the study, selected the company in charge of its implementation and oversaw the coordination between all the parties involved.

Stickers were affixed to food products in four categories: fresh prepared foods (e.g., pizzas, quiches), pastries (e.g., croissants, brioches), breads (e.g., sliced breads, baguettes), and canned prepared meals (e.g., cooked beans, ravioli). These categories were selected because they are consumed regularly by a large percentage of shoppers and because it is relatively easy to place stickers on their packages. These are all processed or ultra-processed foods, classified in the third or fourth NOVA category (Monteiro et al. 2018), and therefore representative of most foods targeted by FOP nutrition labels (nutrition labelling is not mandatory for minimally processed foods, like fruits and vegetables). As we report in the data section, these four categories provide good cross section of products in terms of nutritional quality compared to the French diet.

Participation by manufacturers was voluntary, as per E.U. regulations. Allowing firms to decline to participate matches the regulations of most markets but means that our results are not illustrative of a case of mandatory enforcement. A large majority of manufacturers (29 firms) and all three retailers agreed to participate, leading to a total of 1,266 tested products.

More than 1.9 million stickers were affixed. Sixty research assistants printed the labels on site by directly scanning the product's barcode and using an open-source web-based application designed for this study (https://getiq.inra.fr). Daily quality checks were carried out by supermarket

personnel, with additional checks performed bi-weekly by 24 trained dieticians. Seven independent professional auditors oversaw quality control of the intervention.

Data

Retailers provided purchase data from their loyalty cardholders for two time periods: the ten weeks during which the study was implemented in 2016, and the corresponding ten weeks in the previous year, 2015. Because labels were only affixed on some products from the fifth week onwards, we restrict the analysis to weeks 5 through 10 (for both years). This allowed us to examine the effects of labeling after the initial curiosity and trial phase was over. After removing transactions where product information was missing, our data set included 1,668,301 purchases of 3,586 products, of which 1,266 were labeled products, made by 171,827 consumers.

We assessed the nutritional quality of purchased food using the Ofcom nutrient profiling score developed by the British Food Standards Agency (FSA). The FSA score allocates positive points between 0 and 10 according to the amount of four 'A' components: energy, sugars, saturated fat, and sodium per 100g or 100mL (UK Department of Health 2011); negative points between -5 and 0 according to the amount of three 'C" components: percentage of fruits, vegetables and nuts; fibers, and protein. The FSA score can range from -15 (best) to +40 (worst nutritional quality). We chose the FSA score as a measure of nutritional quality because it is one of the most used in scientific literature. It is also the only system that has been validated in a French context by, among others, prospective associations with the onset of metabolic syndrome, cancer, and cardiovascular risks (Labonté et al. *2*018).

The FSA score can be computed using nutrient and energy information that is mandatory on food packages, but also requires additional information on the amount of fruits, vegetables, and fiber, which we obtained from the manufacturers or, when missing, using standard recipes for the

category. We computed the FSA score for the 1,266 products participating in the RCT, but it was not possible to compute the FSA score for most of the unlabeled products.

As shown in Table 1, canned prepared foods and breads have the best nutritional quality (lowest FSA score), pastries have the worst, and fresh prepared foods are close to the average of the four categories. Overall, the mean FSA score of the chosen products (5.62, when computed across both years) was slightly better than the average FSA score of the French diet, estimated to be 7.67 for men and 7.47 for women (Julia et al., 2016). However, the four categories encompassed a large range of FSA scores, from low (high nutrition quality) for canned prepared meals and breads (respectively, -0.10 and 0.27), to medium for fresh prepared foods (6.1), to high for pastries (14.4). The four categories also differed in terms of the variance of FSA scores across the products of the category, which was low for breads (1.8), moderate for pastries and canned prepared meals (respectively, 3.1 and 3.4), and high for fresh prepared foods (7.23). The average calories per product is about 1,000 and the average weight of each product about 400g.

--- Insert Table 1 here ---

The overall FSA score ranges between 5.41 in 2015 in the SENS stores and 5.75 in the control stores. Table 1 shows that the largest declines (which means the largest improvements) from 2015 to 2016 happened in the fresh prepared foods category, with the largest decline emerging in the Nutri-Score stores, from 6.36 to 5.79.

Outcome measures

We assessed changes in the nutritional quality of food purchased at the purchase incidence level (whether or not to buy foods with high, medium, low, or unlabeled nutritional quality) and at the purchase quantity level (the nutritional quality of the basket of foods purchased, weighted by calories).

To compute the dependent variable of the purchase incidence models, we categorized products as unlabeled or labeled, and further divided labeled products into nutrition terciles based on the FSA score in each category. We then computed the number of purchases over the weeks in our data set for the products in each of four brackets, *b = {tercile 1, tercile 2, tercile 3, unlabeled}*. The terciles were computed for each product category *c* separately, with FSA score cutoffs of {0,2}, {-1,1}, {1,5}, and {12,16} for breads, canned prepared foods, fresh prepared foods and pastries, respectively.

Formally, the outcome variable is defined as:

$$N_{icb,2016} - N_{icb\ 2015} = \sum_t \sum_{j_c \in b} \left(X_{ij_cbt,2016} - X_{ij_cbt,2015} \right) \quad (1)$$

In equation (1), $X_{ij_cbt,2016}$ and $X_{ij_cbt,2015}$ take the value of 1 if shopper *i* purchased product *j* of category *c* in bracket *b* (one of the three terciles or unlabeled), on week *t*, and 0 otherwise. Hence, $N_{icb,2016}$ and $N_{icb\ 2015}$ measure the number of purchases for each individual *i* over all products in each bracket *b*, and so the difference $N_{icb,2016} - N_{icb\ 2015}$ evaluates if that particular shopper increased or decreased purchases of products in the various brackets in product category *c*. By computing the difference between years, we are effectively controlling for individual fixed effects (each shopper acts as her own control). For robustness, we also performed our analysis with number of units purchased (e.g., taking into account if consumers bought more than one unit of each product) instead of number of purchases. The results are substantively similar to the ones presented here.

By categorizing labeled products into three nutrition tiers, as opposed to just two, we test whether nutrition labels have the intended effect of increasing the purchases of products with a higher nutritional quality than average in the category and of decreasing the purchases of products with a lower nutritional quality than average. This allows us to test for unintended purchase

consequences of labeling, such as promoting foods with low nutritional quality. This is important because even if an unintended positive effect on the purchases of low nutritional foods is more than compensated by a positive effect on high nutrition foods, one of the first principles of labeling is that it should do no harm. By incorporating unlabeled products as a fourth tier, the analysis provides information for manufacturers considering adopting each specific FOP labeling system. Studying unlabeled products also allows us to test whether the mere presence of FOP labels in the category changes the consumer's decision processes in the category rather than just providing information about the nutritional quality of labeled foods.

As a second outcome measure, like Crosetto et al. (2020), we computed the nutritional quality of the average basket of foods purchased in each product category. Given the FSA score for each labeled product *j* purchased in category *c*, FSA_{j_c} , we compute the difference between the two years in the quantity-weighted average FSA score of labeled purchased goods for each shopper *i*. In other words, our unit of analysis is the individual-specific basket of labeled goods purchased in each category. Formally, for shopper *i* who purchased quantity q_{ij_c} of product *j* of product category *c,* the difference in the weighted average (across products) FSA score of the basket of goods in that category is:

$$FSA_{ic,2016} - FSA_{ic,2015} = \frac{\sum_j q_{ij_c,2016} FSA_{j_c,2016}}{\sum_j q_{ij_c,2016}} - \frac{\sum_j q_{ij_c,2015} FSA_{j_c,2015}}{\sum_j q_{ij_c,2015}} \qquad (2)$$

Following earlier studies, the weighing quantity q_{ijt} is measured in calories (as a robustness check, the analysis was also done using the quantity in kilograms as weights, with substantively similar results). In terms of interpretation, the dependent variable $FSA_{i,2016} - FSA_{i,2015}$ measures the change in the average FSA score across the shopping basket of products purchased. Because a low FSA score indicates a high nutrition quality, a reduction in the FSA score is an improvement in the nutrition quality of the basket of foods purchased.

The main advantage of the basket measure is that it summarizes the effects of the label on shoppers' buying of high-, medium-, and low-nutritional-quality products into one number, while considering the quantity bought. For this reason, it is the measure of choice in epidemiological studies of the effects of the nutritional quality of food purchased on health (e.g., Donnenfeld et al. 2015). The drawback is that it can only be computed for consumers who bought labeled products in the category both before and during the intervention. It cannot tell whether consumers entered or exited the category. In addition, it cannot tell whether an improvement in average nutritional quality is driven by increasing the purchase of high nutritional quality products, reducing the purchase of products of low nutritional quality, or both. Therefore, we used both measures.

Econometric specifications

We estimate the average treatment effect (ATE) of nutrition labels using a difference-in-differences (DiD) approach, which is commonly used for policy evaluation (Bertrand et al. 2004). In the purchase incidence regressions, we include as covariates the difference in the number of store visits and prices between 2016 and 2015, and fixed effects for each tercile and unlabeled products, and food category. We include, for both regressions, consumers with at least one store visit in any week during the 2015 and 2016 periods where a visit is inferred from having at least one purchase in any of the four categories. The coefficients of interest are those of the interactions between the variables coding the intervention and the nutritional quality tier or unlabeled product. For the purchase incidence model, the difference-in-differences equation is:

$$N_{icb,2016} - N_{icb,2015} = \alpha_c + \beta\left(X_{icb,2016} - X_{icb,2015}\right) + \gamma Z_{iLb} + e_{icb}, \qquad (3)$$

where, as described earlier, the dependent variables $N_{icb,2016}$- $N_{icb,2015}$ measure if the specific consumer *i* increased or decreased their purchases of products in the various brackets in category *c*. On the right-hand side, α_c includes an intercept and fixed effects for each product tier and

category. The term $X_{icb,2016} - X_{icb,2015}$ is the difference between the two years in observed characteristics, such as the difference in the average price paid by the individual and in the number of visits to the store. The variable Z_{iLb} is the interaction term between label system L faced by individual i and bracket b. We used a dummy coding to examine the effect of each label compared to the control condition. Each observation is assigned to only one condition, as each consumer shops at one unique store from the set of stores in our data. All standard errors are store clustered.

For the nutritional basket analysis, the equation is the following:

$$FSA_{ic,2016} - FSA_{ic,2015} = \alpha_c + \beta\left(X_{ic,2016} - X_{ic,2015}\right) + \gamma Z_{iL} + e_{icb}. \qquad (4)$$

As described earlier, the dependent variable $FSA_{i,2016} - FSA_{i,2015}$ measures the change in the average FSA score across the shopping basket of products purchased. The term α_c includes an intercept and category-specific intercepts while $\left(X_{ic,2016} - X_{ic,2015}\right)$ includes the difference between years of the number of purchases in the category c for individual i and the difference in the average prices of individual baskets. The term Z_{iL} includes the variables coded for the case faced by individual i, where L is the FOP label system. We computed the basket's FSA for all consumers with at least one purchase of a labeled product in the category, in both years. As for the purchase incidence analyses, all standard errors are store clustered.

Results

Purchase incidence analyses

The estimates and standard errors are reported in Figure 2 and Table 2. Figure 2 shows the estimated effect of label characteristics and of label systems on the number of purchases of products in the first (in green), second (in amber), and third (in red) nutrition quality tier as well as for unlabeled products (in white).

--- Insert Figure 2 and Table 2 here ---

Looking at each label separately, but for all product categories together, the second column of Table 2 shows that Nutri-Score, and to a lesser degree Nutri-Couleurs significantly increased the purchases of products with high nutritional quality, with coefficients of 0.021 and 0.012 respectively. Given that consumers bought on average 0.142 high-nutrition products in 2015, these estimates correspond to an increase of 14.4 percent and 8.0 percent in purchases of healthier tiered products for Nutri-Score and Nutri-Couleurs, respectively. Second, the effects of Nutri-Score were appropriately ordered, decreasing from high- to medium- to low-nutritional-quality products. In comparison, Nutri-Couleurs had the undesirable effect of slightly increasing the purchase of low-quality products (directionally, not statistically significant) and SENS had the undesirable effect of reducing the purchases of products of medium nutritional quality more than those of products of low nutritional quality. In terms of performance, Nutri-Score is best, followed by Nutri-Couleurs, with SENS and Nutri-Repère having essentially no beneficial effect.

The last four columns of Table 2 show the results of the purchase incidence analyses for each of the four categories separately. It shows that the overall results are mostly driven by the prepared foods category, which has the widest variance of FSA scores across products in that category. Looking at each category separately, the results are consistent with those reported for all products together (second column in Table 2): Nutri-Score is the only label that always has a positive impact on the purchase of products with the highest nutrition quality. Unsurprisingly given the lower number of observations, the reliability of the estimates is lower when each category is estimated separately. We return to category differences when discussing the results of the shopping basket analyses.

In a final analysis, we examined whether the nutrition labels led shoppers to change the total number of purchases by estimating the same model as in Equation 3 but without the four

brackets *b*. Nutri-Score led to a small but statistically insignificant increase in the total number of purchases (B = .006, t = 1.52, p = .13), whereas the other three labels all led to small and statistically insignificant decreases in the total number of purchases (for Nutri-Couleurs: B = -.004, t = -.85, p = .40, for Nutri-Repère: B = -.003, t = -.63, p = .53, for SENS: B = -.008, t = -1.82, p = .07).

Nutritional basket analyses

When considering all products together, the coefficients of the four labels are all in the expected (negative) direction (second column of Table 3) but are all statistically insignificant at the 5% level. The most reliable decrease is for Nutri-Score, with a reduction in FSA score by 0.142 (t = -1.66, p = 0.097), confirming the results from the purchase incidence models that favored this system over others. The reduction of statistical significance of the effects, when compared to the purchase incidence analyses, was expected given the reduction in the number of observations from four per shopper (the changes in the number of purchases of products of high, medium, low, and unlabeled nutritional quality) to just one (the change in the average nutritional quality of the shopping basket score). Another contributing factor is that the FSA score is averaged over all tiers and that purchase incidence analyses found a positive effect on the highest tier of products but no effect on the lowest tier. The robustness of the ordering of the four systems in both analyses is worth noticing: Nutri-Score is best, followed by Nutri-Couleurs, while SENS and Nutri-Repère have the weakest effects. This is the same ordering as for the increase in the purchases of high nutrition quality products shown in Table 2: Nutri-Score is first, followed by Nutri-Couleurs, with SENS and Nutri-Repère having essentially no effect.

--- Insert Table 3, Table 4, and Figure 3 here ---

The diagonal of Table 4 shows the effect sizes associated with each label in the shopping basket analysis. We report Cohen's f^2, calculated based on the semi-partial correlation coefficients (Cohen 2013). All of these effect sizes can be labeled as "very small" according to Sawilowsky's (2009) classification. The lower triangle of Table 4 shows the t-values of pairwise comparison tests for the four labels. These t-values are obtained by estimating the same model as in equation 4 but with a different coding of the treatment conditions in order to measure the difference between two treatment groups, as opposed to the difference between the control group and one treatment group. The upper triangle of Table 4 provides similar t values but comparing the effects of labels on the purchases of high nutrition products in the purchase incidence analysis. These results show that the only statistically significant differences between the labels are those for Nutri-Score.

The last four columns of Table 3 show the shopping basket results separately for each product category. As for the purchase incidence analysis, the largest effects on the nutritional quality of baskets are observed for fresh prepared foods and are the weakest for breads, the categories with, respectively, the largest and lowest variance in nutritional quality across foods. To examine whether nutrition labels are more effective in categories with a higher variance in nutrition quality or in those with a higher nutrition quality, we computed the squared correlation between the four regression coefficients of a particular label (one per category) and the mean and standard deviation of the FSA score of that category before the intervention (in 2015, shown in Table 1).

The results of this analysis are shown for Nutri-Score in Figure 3, which plots the regression coefficient (the estimated decrease in FSA score) as a function of the within-category standard deviation in FSA score (top panel) and the mean FSA score of the category (bottom panel). As can be seen from the linear trendline, the effectiveness of Nutri-Score increased linearly

with the standard deviation in the nutritional quality of the product category ($R^2 = 0.90$). On the other hand, the effectiveness of Nutri-Score is unrelated to the average nutritional quality of the category: the trend is flat but and mean FSA scores do have low explanatory power ($R^2 = 0.05$).

The pattern observed for Nutri-Score was also observed for SENS ($R^2_{SD} = 0.68$ vs. $R^2_M = 0.16$) and for Nutri-Couleurs ($R^2_{SD} = 0.47$ vs. $R^2_M = 0.28$). This means, that the variance, rather than the mean nutritional quality is linked to the performance of three most effective labels. Interestingly, it was the opposite pattern for Nutri-Repère, whose limited effects were less strongly associated with variance than with the mean category nutrition quality ($R^2_{SD} = 0.07$ vs. $R^2_M = 0.76$). One explanation is that the 5 nutrient-specific and mono-colored bar charts in Nutri-Repère make it hard to distinguish between foods with high and low nutritional quality within a category, but that taller bars overall signal to shoppers that the category, as a whole, has a low nutritional quality.

Shopper survey results

The goal of the survey was to shed some light on the differences in the effectiveness of the nutrition labels, particularly between the most effective (Nutri-Score) and its immediate follower (Nutri-Couleurs) by assessing their effects on shopper behaviors that are not captured by transaction data, such as their ability to attract attention and to identify the healthiest options. To achieve this goal, we conducted face-to-face shopper surveys in two waves, one in early September before the start of the experiment, the second in late November, while the labels were in place. We also sent an online questionnaire in late December using the email addresses collected during the two waves but chose not to use it because of concerns about the low response rate (13.6%), potential contamination from earlier responses, and as it coincided with the holiday season.

We randomly selected 20 stores among the 60 that participated in the study to obtain 4 stores per experimental group and a balance across the three retail chains, geographical location,

and income level of the store's catchment area. As with the transaction data, we oversampled shoppers in the control group, aiming for 95 respondents in each of the 4 control stores and 70 in each of the test stores, for a total of 1,500 different respondents per wave.

The survey was conducted on our behalf by CREDOC, a market-research company. Trained interviewers stationed themselves in one aisle of the tested product categories and moved from aisle to aisle during the day. They approached shoppers who appeared to be considering purchasing products in those categories. The response rate was 54.4% in the first wave, yielding 1,844 observations, and 52.2% in the second wave, yielding 1,737 observations. The interviewers asked about respondents' age, which they estimated if they refused to answer. The data were weighted by gender and age group in order to match the respondent samples with the distribution of the entire sample of shoppers approached in the stores (for more details about the procedure, see CREDOC 2017).

To understand the differences between the FOP labels, two of the dimensions examined by the survey are particularly illuminating: 1) their ability to attract attention, 2) their ability to help shoppers rank foods by their nutritional quality (other results are available online, CREDOC 2017). To measure the capacity of each FOP nutrition to attract attention, participants in the second wave were asked: "Did you see that some products on the shelf had a nutrition label on the front of their packages?", which is a standard measure when eye-tracking information is not available (van Herpen and van Trijp 2011). The two summary systems were noticed by more people (48% and 46% of the respondents for Nutri-Score and SENS) than the two analytic systems (respectively 31% and 37% for Nutri-Couleur and Nutri-Repère). A logistic regression of attention on label system controlling for retail chain and socio-demographic variables showed that the difference

between Nutri-Score and SENS was not statistically significant, but that Nutri-Score was more visible than both Nutri-Couleurs and Nutri-Repère at the 1% level.

The key objective of FOP labels is to help people identify the nutritional quality of food products. To measure the performance of the four systems on this key criterion, we first selected three food products in each of the four categories tested, one per nutrition tercile, showed them side by side to the respondents, and asked them to rank them by nutritional quality. This approach is an extension to three products of the classic nutrition comprehension test which asks people to say which of two products presented side by side is nutritionally better (Gorski Findling et al. 2018; Roberto et al. 2012).

To control for the effects of trends in nutrition comprehension or seasonality, we used a difference-in-difference approach, just like for the main study. As mentioned before, we interviewed people who were shopping in control stores, where no FOP label was ever available, as well as in test stores, when FOP labels were absent in the first wave of the survey but were present in the second. In the first wave of the study, as well as in both survey waves for shoppers in control stores, interviewers invited respondents to look at the nutrition fact panel on the back of the packages to help them rank the products. In the second wave, people shopping in test stores were invited to look at the FOP labels when ranking the products. By comparing the changes in the percentage of test stores shoppers who correctly ranked the three products in the first and second waves, we measure whether the FOP systems improved people's ability to identify healthier products. By comparing these changes to those of the respondents who shopped in the control stores, we control for the effects of trends.

--- Insert Table 5 here ---

The first two rows of Table 5 show the percentage of respondents who correctly ranked the three products by nutritional quality in each wave and experimental condition. The change in that percentage between waves is provided in the third row (first difference). Accuracy remained stable in control stores where respondents could only rely on back-of-pack nutrition information in both waves. In the Nutri-Score and SENS stores, accuracy improved between the first and the second wave, suggesting that these labels helped shoppers identify the relative nutritional quality of products. In the Nutri-Couleur and Nutri-Repère stores, however, accuracy decreased between the first and second waves, suggesting that the two analytic labels confused shoppers. All the results were robust and statistically significant at the 5% level when looking at the double-difference shown in the bottom row. The improvement driven by Nutri-Score was the largest and was twice as large as the improvement driven by the second-best system, SENS.

Overall, the in-store surveys suggest that the superior performance of Nutri-Score stems from its ability to draw as much attention as SENS, and significantly more than the two analytic systems, and from a superior ability (compared with the analytic system and with SENS) to help shoppers rank products by nutritional quality. The survey results also suggest that the null effect of Nutri-Score on the purchase of products of low nutritional quality does not arise from shoppers' confusion about their relative nutritional quality.

Conclusion

Although FOP labels have been investigated at length in laboratory experiments and in hypothetical settings, there was still a lot of uncertainty about their ability to improve the nutritional quality of supermarket foods purchases over a significant time period and across a wide variety of brands and product categories. This study also allowed measuring whether FOP labels primarily impact the purchase of brands with a label signaling high or low nutrition quality, their

effect on brands without any FOP label, and whether the effects differ across product categories. Comparison of the four labels was used to determine which nutrition label should receive government approval.

The key overall conclusion is that, compared to the encouraging findings of laboratory-based studies with exhaustive products implementation in all products, FOP nutrition labels had disappointingly modest effects on the nutritional quality of the foods purchased in four categories in real-life grocery shopping conditions. Despite the controls and large number of observations, their impact on the nutritional quality of the shopping basket of labeled products was not statistically significant at the customary 5% level. Although they slightly increased purchases of the best tercile of products in terms of nutritional quality, they slightly decreased purchases of products in the second tercile and had no effect on products in the lowest tercile. They also had insignificant effects on the purchases of unlabeled products.

A second conclusion is that Nutri-Score is the best nutrition label, closely followed by Nutri-Couleurs, with SENS and Nutri-Repère significantly behind. Nutri-Score had the largest and the only statistically significant (at 10 percent) improvement in the nutritional quality of the basket of labeled products purchased, thanks to its positive impact on the purchase of high-nutritional-quality products, followed by a monotonically decreasing effect on the purchases of products of medium and low nutritional quality. This last result is important given that other labels had the undesirable property of having a larger influence on products of medium quality than on products of low-nutritional-quality products. The shopper survey suggests that this happened because Nutri-Score is among the top two most visible labels but provides the easiest way to gauge the relative nutritional quality of the food. The effectiveness of Nutri-Score was unrelated to the mean

nutritional quality of the four product category studied, and increased with the variance in the nutritional quality of the category.

Effect sizes of FOP nutrition labels in the field and in the lab

Compared to what could have been expected based on the results of recent laboratory studies, the effects of even the best nutrition label, Nutri-Score, were disappointingly small. Nutri-Score led to a 14.4 percent increase in purchases of high-nutritional-quality products (an increase by 0.021 FSA points over an average 0.148 purchases of these products). In terms of the nutrition quality of the overall basket of labeled products purchased, the improvement was of 0.142 FSA points, a 2.5 percent improvement of the average FSA score of 5.61. Given that the FSA score has a standard deviation of 7.31, this corresponds to a standardized mean deviation (Cohen's *d*) of only 0.02. In addition, the effects of Nutri-Score, like those of the other three labels, were mostly driven by the fresh prepared food category, a category with the widest variance in nutrition quality. Nutri-Score did not reliably improve the nutrition of food basket in the other three categories, in which foods do not differ as much in terms of nutrition quality.

The Crosetto et al. (2020) study described earlier provides the best lab-based comparison to our results because it collected incentive-compatible purchase decisions, used the same FSA metric to measure nutritional quality, tested the same four labels (implemented on all products) with a similar population, and used the same within-subjects difference-in-difference architecture. The good news is that the correlation between the effectiveness estimates (changes in FSA scores of shopping baskets compared to the control condition) between both studies is 0.82. Nutri-Score was the winner in both studies, followed by Nutri-Couleur (but the order of SENS and Nutri-Repère was different). In terms of effect sizes however, there is a 17 to 1 difference between their estimates and ours. For Nutri-Score, their estimate of the change in basket FSA score is -2.65,

which is 18.6 times larger than our estimate of -0.142. The gap was nearly as wide for Nutri-Couleur (-1.40 vs. -0.12, 12 times larger) and SENS (-0.81 vs. -0.06, 13 times larger) and even bigger for Nutri-Repère (-1.02 vs. -0.24, 43 times larger).

The much lower effects that we observed in the field could be driven by any of the differences between the two studies. First, Crosetto et al. (2020) studied two consecutive purchase decisions whereas this study encompasses multiple purchase decisions over several weeks. It is possible that initial interest for healthier products may have decayed over time as people revert to their habitual behaviors. Another important factor may be that their participants paid stronger attention to the nutrition labels because they had just seen the same catalogue without the labels minutes earlier and because the each label was present on all the products, unlike in our study in which the participation of manufacturers was voluntary. Another important difference is that their study was a "framed natural experiment" as opposed to a "natural experiment" like ours, in which people are unaware that their choices are being studied (Levitt and List 2007). In the food domain, and especially for nutrition-related decisions, there can be a large difference between what people choose when they are being watched and when they are not (Herman, Roth, and Polivy 2003, Holden, Zlatevska and Dubelaar 2016, Vartanian 2015).

That nutrition labels have much smaller effects in the field than in the lab has important implications for public health. A recent paper (Egnell et al. 2019) used the estimates of the Crosetto et al. (2020) study to simulate the effects of nutrition labels on dietary intakes (using data from an observational study) and then, through a macro-simulation using the PRIME model (Scarborough et al. 2014), estimated the effects on mortality from diet-related non-communicable diseases. They estimated that approximately 7,680 deaths (3.4% of all deaths from diet-related non-communicable diseases in France) would be avoidable if the Nutri-Score was used. Clearly, it would be useful to

run these simulations with our much smaller estimates. On the other hand, even small changes in FSA scores can have significant health outcomes. Prospective studies have found that a one-point increase in the FSA score computed over a total diet is associated with a 16% higher risk of obesity among men (Julia et al. 2015) and with a higher risk (hazard ratio = 1.08) of cardiovascular disease (Adriouch et al. 2016).

More generally, it would be important to compare the costs and benefits of FOP nutrition labeling with those of other cognitive-focused nudges, such as shelf display changes, but also of affect-focused nudges that motivate people to eat better by using pleasure or social pressure, and of behavior-focused nudges that directly change behaviors by changing portion sizes, for example. A recent meta-analysis has shown that the effectiveness of nudges increases significantly as their focus shifts from cognition to affect to behavior (Cadario and Chandon 2020).

The relative effectiveness of nutrition labels: The role of design characteristics

According to a recent Delphi analysis of expert opinion, the most important guideline for effective FOP communication is "to keep communication simple and clear and avoid jargon" (Hung et al. 2019). This recommendation is consistent with earlier work (e.g., Ratner and Riis 2014) and underpins the entire reasoning in favor of FOP labeling in the first place (McGuire 2012).

One of the key design choices influencing the simplicity of a nutrition label is whether it provides a single summary score, like Nutri-Score or SENS, or nutrient-specific (analytic) information, like Nutri-Couleurs or Nutri-Repère. Research has shown that consumers can more easily assess the healthfulness of a product with summary labels than with analytic ones (Ikonen et al. 2020). A few studies (Borgmeier and Westenhoefer 2009; Ducrot et al. 2016) also found summary labels to be more effective than analytic ones when it came to intentions to make

healthier purchases. However, Ikonen et al.'s (2020) meta-analysis did not find any statistically significant differences between analytic and summary labels in terms of intentions to purchase healthy foods, healthy food choices, or consumption.

At least three other design choices could explain the relative performance of the four labels. One is whether the label used a color-coding with a well-established ordinal meaning, like the green-amber-red traffic light color coding (used by Nutri-Score and Nutri-Couleurs), another color coding without the same ordinal meaning (like the green-blue-orange-purple color code used by SENS) or no color coding at all (like Nutri-Repère). Although there are exceptions (e.g., Borgmeier and Westenhoefer 2009), a recent review (Hersey et al. 2013) found that meaningful (e.g., traffic light) color coding facilitates understanding and food selection compared to labels without color. Other design choices are whether to displays the full range of possible scores in addition to the food's own and whether to provides grades (like Nutri-Score's A to E grades) or behavioral recommendations. Studies suggest that labels that help people assess the relative nutritional quality of products are more effective than those that only provide its absolute nutritional quality (Ducrot et al. 2016; Feunekes et al. 2008; Helfer and Shultz 2014; Hersey et al. 2013).

Given the multiple differences between the four labels studied and the lack of data on the choice process and goals of shoppers, this study cannot provide definite answers to the question of why some labels performed better than others. Still, our results show that being a summary label rather than an analytic label, is not, in itself, enough to reliably impact performance. An additional test comparing the two summary labels (Nutri-Score and SENS) to the two analytic ones (Nutri-Couleurs and Nutri-Repère) found no statistically significant difference in terms of the purchases of high-nutrition foods ($t = 0.61$, $p = .54$) or the nutritional quality of the basket of food purchased

(t = -0.55, p = .58). On the other hand, although our survey results show that Nutri-Score is the most effective to help people rank products by nutritional quality, we cannot tell whether it is because of its color coding, use of letter grades, use of five categories (vs. four for SENS), or because its ruler design shows the full range of possible grades. Future research is necessary to examine the effects of these design characteristics on the performance of nutrition labels.

Implications for future research

Because participation in the study by manufacturers was on a voluntary basis, our results must be interpreted in this context and cannot be easily extrapolated to contexts that would force all firms to adopt a label. We found no systematic effect of labeling on the purchases of unlabeled products, ruling out the possibility that consumers drew any statistically significant inferences, negative or positive, from the fact that a product did not carry a nutrition label. Under the plausible assumption that the nutritional quality of unlabeled products is below the category average (otherwise, its manufacturers would have chosen to participate in the study), it indicates that the mere presence of labels did not increase healthy eating motivation or attention to back-of-pack nutrition information. Relatedly, our study did not look at the impact of labeling systems on the decision of manufacturers to reformulate their products, which has been identified as one of the most relevant and cost-effective public health nutrition strategies (Dobbs et al. 2014), and has occurred in response to other labeling initiatives (Vyth et al. 2010).

The strengths of the trial include its large-scale real setting condition and rigorous RCT design, which are increasingly called for given that most true effect sizes are likely to be small and hard to estimate precisely in under-powered studies (Funder and Ozer 2019). However, it is not without limitations. The study was restricted to four food categories. Further research using multiple socioeconomic indicators is necessary to examine the effects of graphical labeling

systems on specific populations and for a larger variety of foods, including less processed ones. Future research should also examine the underlying purchasing process in order to study whether nutrition labels operate by enhancing attention to nutrition information, reducing information-processing costs (Russo et al. 1986), or by increasing the importance attached to health over other benefits of food, like taste. To help manufacturers and retailers determine whether they should adopt these labels and reformulate their foods, it would also be useful to estimate the short and long-term effects of these labels at the brand level on consumer price sensitivity, on product-level substitution patterns. Finally, given the distinctiveness of French attitudes to food, it remains to be seen whether these results would hold in different countries.

REFERENCES

Adriouch, S., Julia, C., Kesse-Guyot, E., Méjean, C., Ducrot, P., Péneau, S., Donnenfeld, M., Deschasaux, M., Menai, M., Hercberg, S., Touvier, M., & Fezeu, L. K. (2016). Prospective association between a dietary quality index based on a nutrient profiling system and cardiovascular disease risk. *European Journal of Preventive Cardiology*, *23*(15), 1669-1676.

Andrews, J. C., Burton, S., & Kees, J. (2011). Is simpler always better? Consumer evaluations of front-of-package nutrition symbols. *Journal of Public Policy & Marketing*, *30*(2), 175-190.

Askew, K. (2019). NutriScore under fire: Debate rages over nutrition labeling in Germany. *FoodNavigator, 2*6 April 2019.

Belei, N., Geyskens, K., Goukens, C., Ramanathan, S., & Lemmink, J. (2012). The Best of Both Worlds? Effects of Attribute-Induced Goal Conflict on Consumption of Healthful Indulgences. *Journal of Marketing Research*, *49*(6), 900-909.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, *119*(1), 249–75.

Bleich, S. N., Economos, C. D., Spiker, M. L., Vercammen, K. A., VanEpps, E. M., Block, J. P., Elbel, B., Story, M., & Roberto, C. A. (2017). A Systematic Review of Calorie Labeling and Modified Calorie Labeling Interventions: Impact on Consumer and Restaurant Behavior. *Obesity*, *25*(12), 2018-2044.

Bollinger, B., Leslie, P., & Sorensen, A. (2011). Calorie posting in chain restaurants. *American Economic Journal*, 3:91-128.

Borgmeier, I., & Westenhoefer, J. (2009). Impact of different food label formats on healthiness evaluation and food choice of consumers: A randomized-controlled study. *BMC Public Health*, *9*(1), 184.

Boudreau, K.J., Lacetera, N., & Lakhani, K.R. (2011). Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management Science,* 57(5), 843-863.

Cadario, R., & Chandon, P. (2020). Which healthy eating nudges work best? A meta-analysis of field experiments. *Marketing Science,* Forthcoming.

Cadario, R., & Chandon, P. (2019). Effectiveness or consumer acceptance? Tradeoffs in selecting healthy eating nudges. *Food Policy*, *85*, 1-6.

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

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences.* Routledge.

CREDOC. 2017. *Les enquêtes complémentaires : Etude quantitative évaluation ex ante de systèmes d'étiquetage nutritionnel graphique simplifié*. Retrieved from https://solidarites-sante.gouv.fr/IMG/pdf/rapport-credoc-fev_2017.pdf

Crosetto, P., Lacroix, A., Muller L., & Ruffieux, B. (2020). Nutritional and economic relative impacts of 5 alternative front-of-pack nutrition labels: experimental evidence. *European Review of Agricultural Economics, 47*(2), 785-818.

Darmon, N., Vieux, F., Maillot, M., Volatier, J. L., & Martin, A. (2009). 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*, *89*(4), 1227–1236.

Dobbs, R., Sawers, C., Thompson, F., Manyika, J., Woetzel, J. R., Child, P., & Spatharou, A. (2014). Overcoming obesity: An initial economic analysis. *McKinsey Global Institute*.

Donnenfeld, M., Julia, C., Kesse-Guyot, E., Mejean, C., Ducrot, P., Peneau, S., Deschasaux, M., Latino-Martel, P., Fezeu, L., Hercberg, S., & Touvier, M. (2015). Prospective association between cancer risk and an individual dietary index based on the British Food Standards Agency Nutrient Profiling System. *British Journal of Nutrition*, *114*(10), 1702–1710.

Ducrot, P., Julia, C., Méjean, C., Kesse-Guyot, E., Touvier, M., Fezeu, L. K, Hercberg, S., & Péneau, S. (2016). Impact of different front-of-pack nutrition labels on consumer purchasing intentions: A randomized controlled trial. *American Journal of Preventive Medicine*, *50*(5), 627–636.

Egnell, M., Crosetto, P., d'Almeida, T., Kesse-Guyot, E., Touvier, M., Ruffieux, B., Hercberg, S., Muller, L., & Julia, C. (2019). Modelling the impact of different front-of-package nutrition labels on mortality from non-communicable chronic disease. *International Journal of Behavioral Nutrition and Physical Activity*, *16*(1), 56.

Elbel, B., Kersh, R., Brescoll, V. L., & Dixon, L. B. (2009). Calorie Labeling And Food Choices: A First Look at The Effects On Low-Income People In New York City. *Health Affairs*, *28*(6), w1110-1121.

Feunekes, G. I., Gortemaker, I. A., Willems, A. A., Lion, R., & Van Den Kommer, M. (2008). Front-of-pack nutrition labeling: Testing effectiveness of different nutrition labeling formats front-of-pack in four European countries. *Appetite*, *50*(1), 57–70.

Food Standards Agency (2007). *Front-of-pack Traffic Light Signpost Labelling Technical Guidance*. Issue 1, January 2007, available at: https://www.foodwatch.org/fileadmin/Themen/Ampelkennzeichnung/guidance_ampel_issue_1_januar_2007.pdf

Funder, D.C., & D.J. Ozer. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156-168.

Gaigi, H., Raffin, S., Maillot, M., Adrover, L., Ruffieux, B., & Darmon, N. (2015). Expérimentation d'un fléchage nutritionnel dans deux supermarchés à Marseille «le choix Vita+». *Cahiers de Nutrition et de Diététique, 50*(1), 16–24.

Gorski Findling, M.T., Werth, P.M., Musicus, A.A., Bragg, M.A., Graham, D.J., Elbel, B., & Roberto, C.A. (2018). Comparing five front-of-pack nutrition labels' influence on consumers' perceptions and purchase intentions. *Preventive Medicine*, *106,* 114-121.

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

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

Helfer, P., & Shultz, T. R. (2014). The effects of nutrition labeling on consumer food choice: A psychological experiment and computational model. *Annals of the New York Academy of Sciences*, *1331*(1), 174–185.

Herman, C. P., Roth, D. A., & Polivy, J. (2003). Effects of the Presence of Others on Food Intake: A Normative Interpretation. *Psychological Bulletin*, *129*(6), 873-886.

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

Holden, S. S., Zlatevska, N., & Dubelaar, C. (2016). Whether Smaller Plates Reduce Consumption Depends on Who's Serving and Who's Looking: A Meta-Analysis. *Journal of the Association for Consumer Research*, *1*(1), 134-146.

Hung, Y., Hieke, S., Grunert, K. G., & Verbeke, W. (2019). Setting policy priorities for front-of-pack health claims and symbols in the European union: Expert consensus built by using a Delphi method. *Nutrients*, *11*(2), 403.

Ikonen, I., Sotgiu, F., Aydinli, A., & Verlegh, P. W. (2020). Consumer effects of front-of-package nutrition labeling: an interdisciplinary meta-analysis. *Journal of the Academy of Marketing Science*, https://doi.org/10.1007/s11747-019-00663-9.

Julia C., Méjean, C., Touvier, M., Péneau, S., Lassale, C., Ducrot, P., Hercberg, S., & Kesse-Guyot, E. (2016). Validation of the FSA nutrient profiling system dietary index in French adults—findings from SUVIMAX study. *European Journal of Nutrition, 55,* 1901-1910.

Julia, C., & Hercberg, S. (2017). Development of a new front-of-pack nutrition label in France: The five-colour nutri-score. *Public Health Panorama*, *3*(4), 712–725.

Julia, C., P. Ducrot, C. Lassale, L. Fézeu, C. Méjean, S. Péneau, M. Touvier, S. Hercberg, E., & Kesse-Guyot, E. (2015). 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*, *81,* 189-194.

Kiesel, K., & Villas-Boas, S. B. (2013). Can information costs affect consumer choice? Nutrition labels in a supermarket experiment. *International Journal of Industrial Organization*, *31,* 153–163.

Kiszko, K. M., Martinez, O. D., Abrams, C., & Elbel, B. (2014). The Influence of Calorie Labeling on Food Orders and Consumption: A Review of the Literature. *Journal of Community Health*, *39*(6), 1248-1269.

Labonté, M.-È., Poon, T., Gladanac, B., Ahmed, M., Franco-Arellano, B., Rayner, M., & L'Abbé, M. R. (2018). Nutrient profile models with applications in government-led nutrition policies aimed at health promotion and noncommunicable disease prevention: A systematic review. *Advances in Nutrition, 9*(6), 741–788.

Levitt, S. D. & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic perspectives*, *21*(2), 153-174.

Long M.W., Tobias, D.K., Cradock, A.L., Batchelder, H., & Gortmaker, S.L. (2015). Systematic review and meta-analysis of the impact of restaurant menu calorie labeling. *American Journal of Public Health, 105*(5), e11–e24.

Maillot, M., Braesco, V., & Darmon, N. (2016). Relevance of the nutrient profiling SENS* system in relation to the overall nutritional quality of observed diets. *EuroFIR Food Forum 2016* in Brussels.

McGuire, S. (2012). Institute of medicine. Front-of-package nutrition rating systems and symbols: Promoting healthier choices. *Advances in Nutrition*, *3*(3), 332-333.

Michail, Niamh (2015), French retailers unveil alternative nutrition label, *Food Navigator* (November 2). Available online: https://www.foodnavigator.com/Article/2015/11/02/French-retailers-unveil-alternative-nutrition-label [Accessed 20 March 2020].

Mhurchu, C. N., Volkova, E., Jiang, Y., Eyles, H., Michie, J., B. Neal, T. Blakely, B. Swinburn, & Rayner, M. (2017). Effects of interpretive nutrition labels on consumer food purchases: The Starlight randomized controlled trial. *The American Journal of Clinical Nutrition*, *105*(3), 695–704.

Monteiro, C. A., Cannon, G., Moubarac, J.-C., Levy, R. B., Louzada, M. L. C. & Jaime, P. C. (2018). The UN Decade of Nutrition, the NOVA food classification and the trouble with ultra-processing. *Public Health Nutrition*, *21*(1), 5-17.

Muller L., Lacroix A., Lusk J., & Ruffieux B. (2017). Distributional Impacts of Fat Taxes and Thin Subsidies. *The Economic Journal*, *127* (604), 2066–2092.

Newman, C. L., Burton, S., Andrews, J. C., Netemeyer, R. G., & Kees, J. (2018). Marketers' use of alternative front-of-package nutrition symbols: An examination of effects on product evaluations. *Journal of the Academy of Marketing Science*, *46*(3), 453–476.

Nikolova, H. D., & Inman, J. J. (2015). Healthy choice: The effect of simplified POS nutritional information on consumer food choice behavior. *Journal of Marketing Research*, *52*(6), 817–835.

Nutri-Repère. (2015). Understand food labels better with Nutri-Repère. *http://wwwnutrireperefr/en/*.

Ratner, R. K., & Riis, J. (2014). Communicating science-based recommendations with memorable and actionable guidelines. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(Suppl 4), 13634–13641.

Regulation EU (2011), No 1169/2011 of the European Parliament and of the Council of 25 October, on the provision of food information to consumers. Retrieved from http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32011R1169.

Renaudin, N., Albuquerque, P., Bertail, P., Boirie, Y, Chandon, P., Combris, P., Lahlou, S., Micard, V., Oppert, JM., Rigal, N., & Ruffieux, B. (2016). Evaluation ex ante de systèmes d'etiquetage nutritionnel graphique simplifié : Rapport scientifique. French Ministry for Solidarity and Health, 16 April 2016. Retrieved from http://solidarites-sante.gouv.fr/IMG/pdf/rapport_du_comite_scientifique_11_avril_2016.pdf.

Roberto, C.A., Bragg, M.A., Schwartz, M.B., Seamans, M.J., Musicus, A., Novak, N. & Brownell, K.D. (2012). Facts up front versus traffic light food labels: A randomized controlled trial. *American Journal of Preventive Medicine*, *43*(2), 134-141.

Russo, J.E., Staelin, R., Nolan, C.A., Russell, G.J., & Metcalf, B.L. (1986). Nutrition information in the supermarket. *Journal of Consumer Research*, *13*(1), 48-70.

Scarborough, P., Harrington, R. A., Mizdrak, A., Zhou, L. M., & Doherty, A. (2014). The Preventable Risk Integrated ModEl and Its Use to Estimate the Health Impact of Public Health Policy Scenarios. *Scientifica*, 1-21.

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*(2), 26.

UK Department of Health (2011), *Nutrient Profiling Technical Guidance*. Retrieved from https://www.gov.uk/government/publications/the-nutrient-profiling-model.

van Herpen, E., & van Trijp, H.C.M. (2011). Front-of-pack nutrition labels. Their effect on attention and choices when consumers have varying goals and time constraints. *Appetite*, *57*(1), 148-160.

Vartanian, L. R. (2015) Impression management and food intake. Current directions in research. *Appetite*, *86,* 74-80.

Vyth, E. L., Steenhuis, I. H., Roodenburg, A. J., Brug, J., & Seidell, J. C. (2010). Front-of-pack nutrition label stimulates healthier product development: A quantitative analysis. *International Journal of Behavioral Nutrition and Physical Activity*, *7*(65), 1-7.

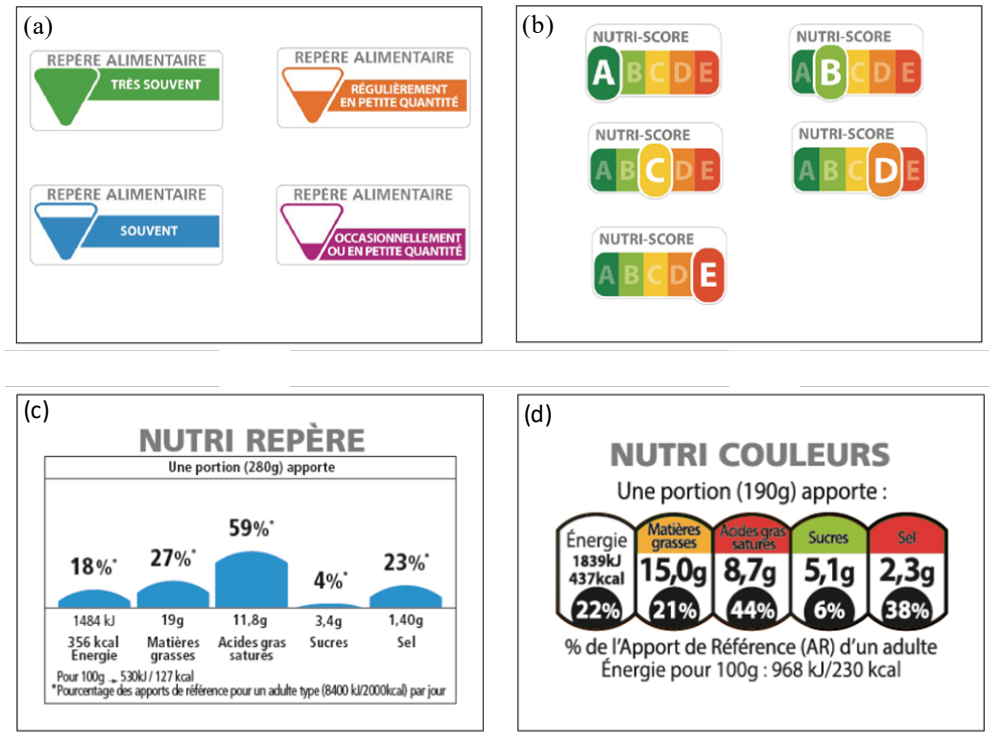

Figure 1. FOP label tested: SENS (a), Nutri-Score (b), Nutri Repère (c), Nutri-Couleurs (d)

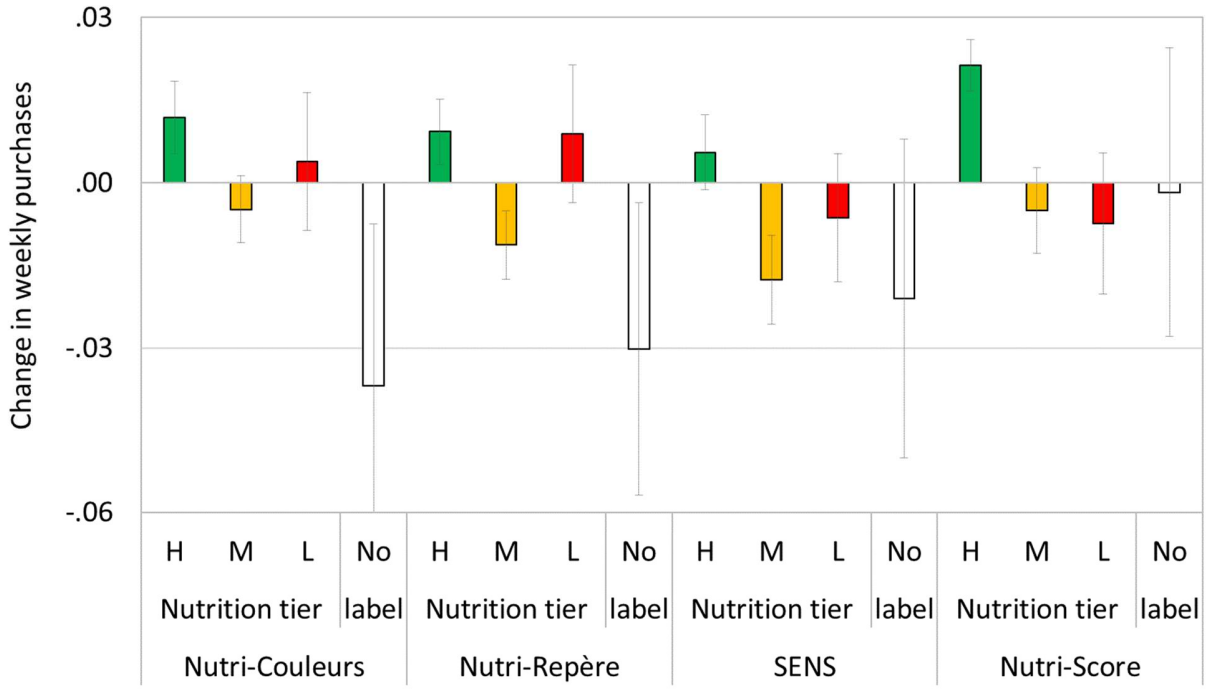

Figure 2. Effects of labels on purchases of labeled products with high (green), medium (amber), and low (red) nutrition tier and on purchases of unlabeled products (white)

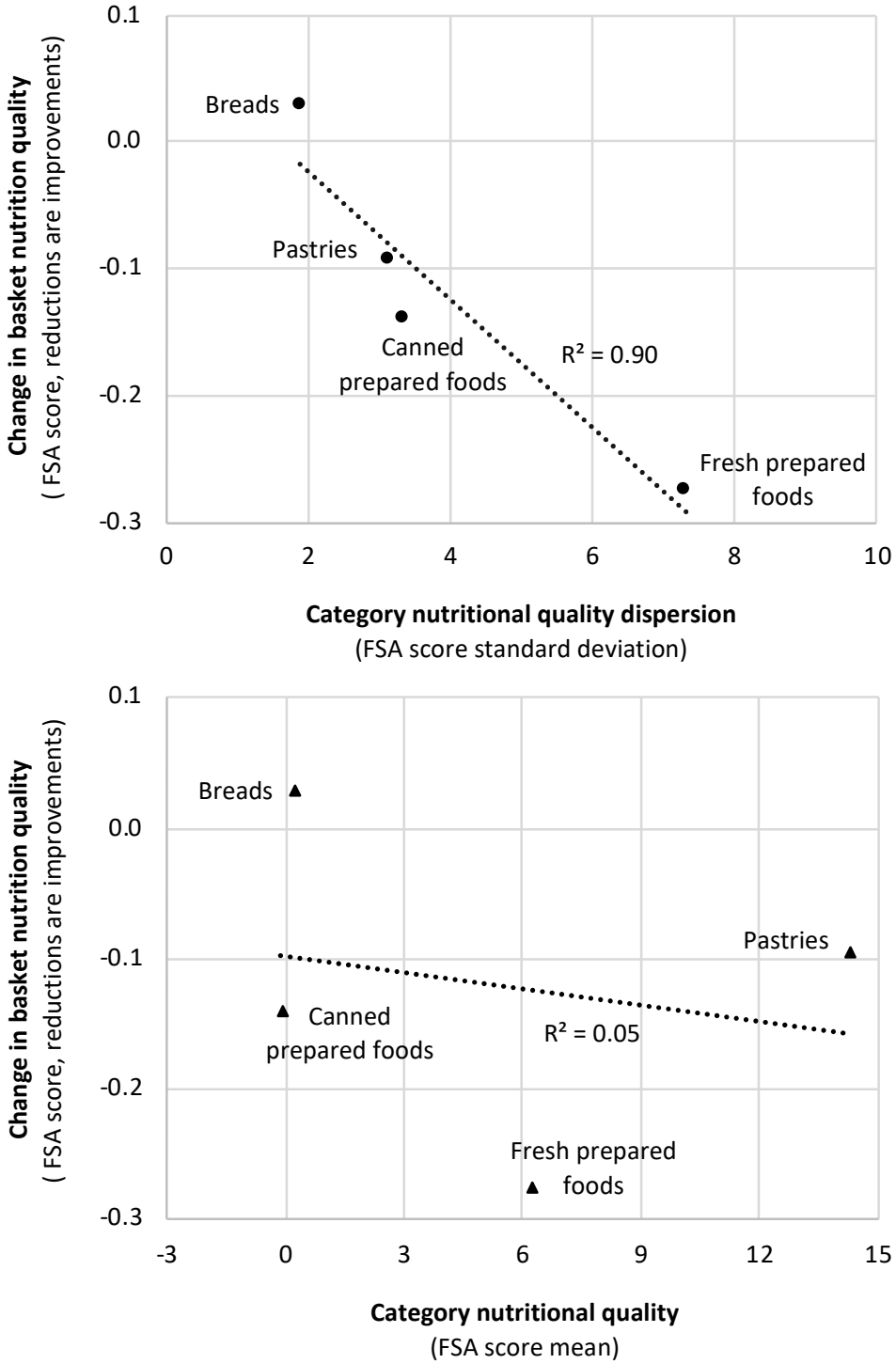

Figure 3. Effects of Nutri-Score by category nutritional quality: Standard Deviation (top) vs. mean (bottom)

Table 1. Descriptive statistics: Average FSA score (labeled products)

	All		Fresh prep. foods		Pastries		Breads		Canned prep. foods	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
All Stores										
Year: 2015	5.61	7.31	6.22	7.31	14.29	3.13	.21	1.89	-.12	3.36
Year: 2016	5.63	7.23	5.93	7.14	14.43	3.08	.33	1.63	-.09	3.39
Control stores										
Year: 2015	5.71	7.34	6.16	7.33	14.22	3.10	.22	1.89	-.19	3.29
Year: 2016	5.75	7.25	5.95	7.16	14.40	3.06	.36	1.60	-.04	3.41
Nutri-Couleurs stores										
Year: 2015	5.53	7.23	6.24	7.22	14.14	3.26	.31	1.94	-.11	3.40
Year: 2016	5.50	7.16	5.86	7.08	14.35	3.02	.39	1.55	-.09	3.40
Nutri-Repère stores										
Year: 2015	5.61	7.30	6.29	7.27	14.32	3.11	.18	2.09	-.01	3.63
Year: 2016	5.63	7.23	6.09	7.16	14.46	3.06	.32	1.72	-.11	3.47
SENS stores										
Year: 2015	5.41	7.30	6.10	7.29	14.45	3.10	.15	1.86	-.07	3.40
Year: 2016	5.51	7.26	5.93	7.15	14.55	3.04	.28	1.65	-.12	3.40
Nutri-Score stores										
Year: 2015	5.65	7.35	6.36	7.37	14.39	3.11	.19	1.70	-.13	3.20
Year: 2016	5.59	7.22	5.79	7.06	14.44	3.15	.32	1.63	-.13	3.32
Weight (grams)	395	167	311	108	449	136	457	149	554	259
Energy (kcal)	1060	436	955	440	1550	180	1148	95	460	144
N (labeled products)	1,068,857		518,087		194,662		250,023		104,755	
N (unlabeled products)	599,444		276,119		99,457		117,030		106,838	

Note: Decreasing FSA scores indicate an improvement in nutritional quality.

Table 2. Impact of nutrition labels on purchase incidence

	All products			Fresh prepared foods			Pastries			Breads			Canned prep. foods		
	Est.	S.E.	*t*	Est.	S.E.	*t*	Est.	S.E.	*t*	Est.	S.E.	*t*	Est.	S.E.	*t*
Nutri-Couleurs x T1 (H)	.012*	.007	1.79	.045*	.024	1.84	-.006	.009	-.66	.000	.011	.02	.006	.006	.98
Nutri-Couleurs x T2 (M)	-.005	.006	-.80	-.007	.016	-.42	-.009	.012	-.73	-.009	.014	-.65	.004	.010	.40
Nutri-Couleurs x T3 (L)	.004	.013	.31	.000	.037	.00	.006	.022	.27	.001	.018	.04	.007	.008	.86
Nutri-Couleurs x No label	-.037	.029	-1.26	-.041	.061	-.68	-.051	.053	-.96	-.009	.017	-.52	-.046*	.024	-1.94
Nutri-Repère x T1 (H)	.009	.006	1.58	.027	.019	1.42	-.001	.008	-.09	.011	.014	.75	.001	.008	.12
Nutri-Repère x T2 (M)	-.011*	.006	-1.81	-.019	.020	-.94	-.003	.015	-.17	-.030	.018	-1.66	.006	.007	.89
Nutri-Repère x T3 (L)	.009	.013	.71	.018	.036	.49	.001	.020	.06	.028	.024	1.16	-.011	.009	-1.28
Nutri-Repère x No label	-.030	.027	-1.14	-.013	.057	-.23	-.042	.049	-.87	-.034	.021	-1.59	-.031	.017	-1.84
SENS x T1 (H)	.006	.007	.81	.010	.021	.46	-.004	.012	-.33	.012	.019	.61	.003	.008	.39
SENS x T2 (M)	-.018**	.008	-2.21	-.023	.017	-1.38	-.011	.015	-.73	-.038	.016	-2.32	.000	.011	.00
SENS x T3 (L)	-.006	.012	-.55	-.010	.029	-.34	-.017	.021	-.83	.012	.024	.52	-.012	.009	-1.33
SENS x No label	-.021	.029	-.72	.013	.080	.16	-.021	.047	-.44	-.039	.027	-1.41	-.037	.020	-1.88
Nutri-Score x T1 (H)	.021***	.005	4.58	.044***	.015	2.96	.018*	.010	1.89	.014	.018	.76	.010	.007	1.36
Nutri-Score x T2 (M)	-.005	.008	-.65	.007	.012	.56	-.004	.016	-.28	-.029	.017	-1.70	.006	.006	.95
Nutri-Score x T3 (L)	-.007	.013	-.58	-.027	.039	-.69	-.001	.017	-.05	.000	.017	.02	-.003	.006	-.44
Nutri-Score x No label	-.002	.026	-.07	-.018	.063	-.28	-.010	.050	-.21	.033	.022	1.53	-.011	.018	-.59
R^2	.064			.092			.081			.076			.038		
N	1,644,176			411,044			411,044			411,044			411,044		

Note: *** $p<.01$, ** $p<.05$, * $p<.10$. An intercept, fixed effects for nutrition tercile 2, nutrition tercile 3 and unlabeled products, and changes in the number of purchases in the category and in average basket prices were also included in the regressions but are not shown here.

Table 3. Impact of nutrition labels on calorie-weighted basket FSA scores

	All products			Fresh prepared foods			Pastries			Breads			Canned prep. foods		
	Est.	S.E.	*t*	Est.	S.E.	*t*	Est.	S.E.	*t*	Est.	S.E.	*t*	Est.	S.E.	*t*
Nutri-Couleurs	-.115	.102	-1.13	-.195	.208	-.94	.031	.132	.23	-.095	.100	-.95	-.084	.103	-.82
Nutri-Repère	-.024	.077	-.31	-.059	.130	-.45	.126	.108	1.17	-.025	.104	-.24	-.109	.130	-.84
SENS	-.062	.070	-.88	-.140	.123	-1.14	-.084	.110	-.77	.029	.108	.27	-.093	.093	-1.00
Nutri-Score	-.142*	.086	-1.66	-.275*	.164	-1.68	-.094	.064	-1.47	.029	.082	.35	-.140	.103	-1.35
R^2		.008			.019			.006			.004			.031	
N		105,245			48,673			18,801			28,835			8,936	

Note: *** $p<.01$, ** $p<.05$, * $p<.10$. An intercept and changes in the number of purchases in the category and in average basket prices were also included in the regressions. Negative coefficients represent an improvement in nutritional quality (i.e., lower FSA scores).

Table 4. Effect sizes and *t* values from pairwise coefficient comparisons

	Nutri-Couleurs	Nutri-Repère	SENS	Nutri-Score
Nutri-Couleurs	**$2.28*10^{-5}$**	0.36	0.81	1.58
Nutri-Repère	0.96	**$7.86*10^{-6}$**	0.52	2.29^{**}
SENS	0.59	0.63	**$4.03*10^{-7}$**	2.52^{**}
Nutri-Score	0.27	1.51	1.12	**$7.22*10^{-5}$**

Note: Numbers in the diagonal are the effect sizes (Cohen's f^2) from the basket analyses. Numbers in the off diagonals are the t-values of pairwise comparisons of the regression coefficients of each label for the basket analyses (lower triangle) or for the purchase of high nutrition products (upper triangle). ** $p<.05$.

Table 5. Survey respondents correctly ranking three foods by nutritional quality (%)

	Control stores	Nutri-Couleur stores	Nutri-Repère stores	SENS stores	Nutri-Score stores
Wave 1 (pre-intervention)	55	61	61	65	56
Wave 2 (during the intervention)	52	28	21	79	89
First difference (Wave 2 - Wave 1)	-3	-33	-40	15	33
Double difference (FD in test - FD in control stores)		-30**	-37**	18**	36**

Note: Respondents intercepted in control stores and in the test-stores during the first wave were encouraged to use back-of-package nutrition information. Respondents intercepted in test stores during the second wave were encouraged to use front-of-pack nutrition labels. ** $p<.05$.

Web Appendix

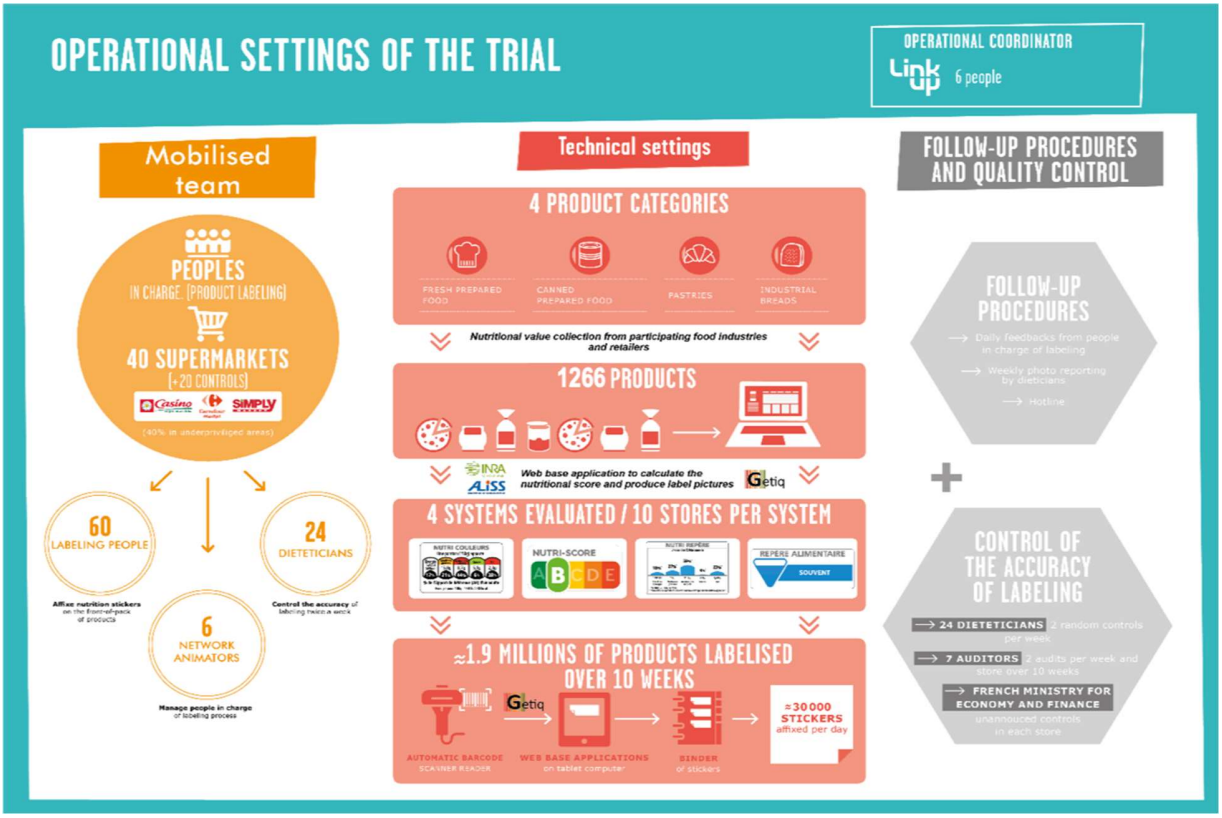

Figure A1. Summary of the operational settings

Figure A2. Example of leaflet (Nutri-Score)

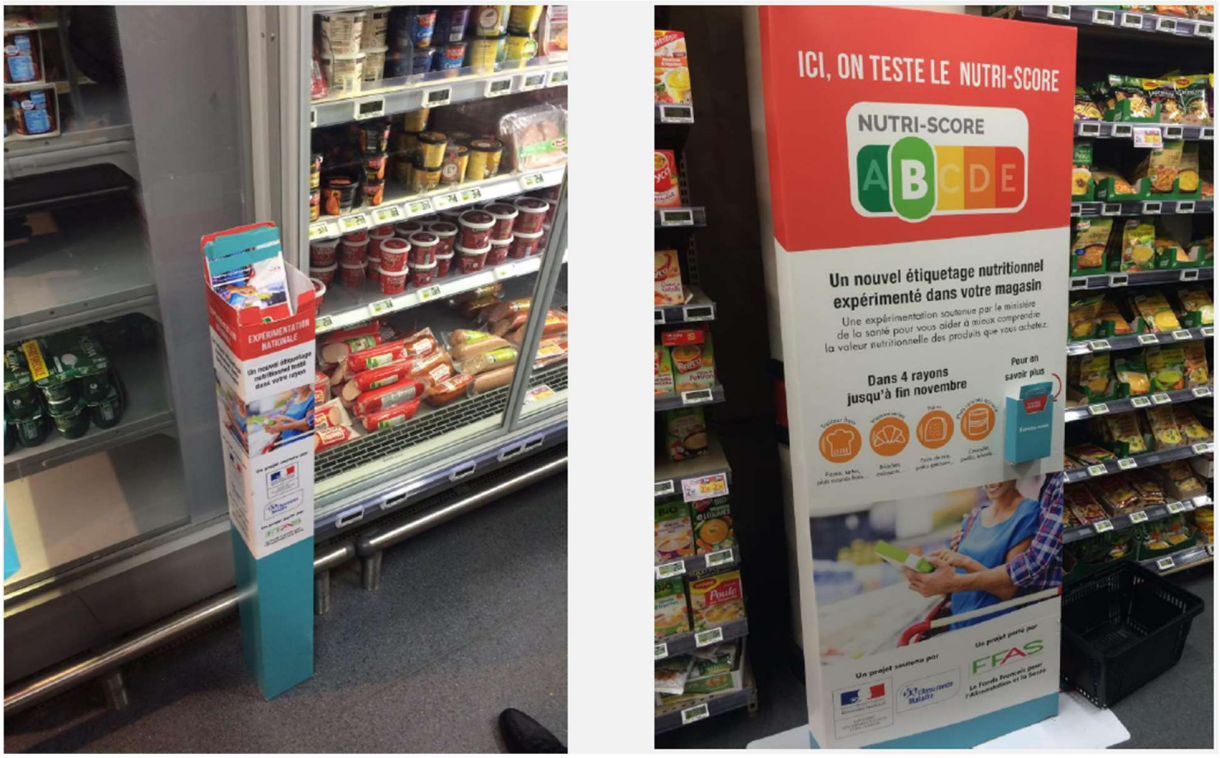

Figure A3. Totems in stores

Figure A4. Nutri-Score stickers in two product categories (top: bread; bottom: canned prepared food)