NIHR | Policy Research Unit in Obesity

Briefing Paper: Preliminary results from an experimental study examining the impact of front of pack labelling skills on knowledge

Authors

Jessica Packer, Simon Russell, Deborah Ridout, Steven Hope, Anne Conolly, Curtis Jessop, Russell Viner*, Helen Croker*

(*co-principal investigators)

22nd January 2021

Executive Summary

Question

How much do front of pack labels (FOPLs) help participants to understand and rate the healthiness of food and drink items compared to having no label?

Background

FOPLs provide easily accessible nutritional information on the front of food packaging to help guide consumer choices. Research to date has shown that FOPLs improve nutrition knowledge and influence purchasing decisions. There are several different FOPLs in use, from more complex scoring systems (e.g. Nutri-Score or Multiple Traffic Lights) to pack warnings (Warning Labels) to simple markers of healthiness (e.g. Positive Tick). In the UK, little is known about how different FOPLs inform people to help make choices based on the healthiness of food and drink. We used a robust experimental design to test how well specific FOPLs enable accurate judgements to be made about the healthiness of food and drinks.

Methods

We ran an online randomised experimental using a large nationally representative sample of adults in Great Britain (run by NatCen). We selected six common food/drink categories with differences in their nutritional content (pizza, instant hot chocolate, cake, crisps, yoghurt, cereal). Nutritional profiles were created for three realistic products in each category to include a most healthy, middle and least healthy product option; mock images of packaging were created to reflect the different nutritional profiles. Product information which could influence participant responses (such as nutritional claims) was avoided. Our main outcome measure was participants' ability to rank the three products within each food category. Each ranking task required participants to view images of the three products within a category and then rank them as 'most healthy', 'middle/in between' and 'least healthy'.

Participants were randomly allocated to one of five conditions, either to see one of four FOPLs with each food/drink image (Multi Traffic Light; Nutri-Score; Warning Label; Positive Tick) or to a no-FOPL control group. All participants were asked to complete two ranking tasks for each food/drink group: first a baseline ranking (with no FOPL shown) and then an experimental ranking with a FOPL or control. We assessed three outcomes for each food/drink group: 1) ranking of healthiness of products (considered correct if all three products were ranked correctly and incorrect if any were wrong); 2) change in ranking from baseline (no FOPL) to follow-up (FOPL shown as per allocated group) (scored +1 if participants improved, 0 didn't improve or -1 worsened); and 3) change in a global food score, created by combining the changes in the rankings given for each of the five food products (global score range is -5, +5). The global score gives an indication of participants' overall ability to do the ranking tasks according to the FOPL they saw. We also asked participants if they had enough information to do each ranking task.

Analyses were conducted to examine if participants' ability to correctly order the healthiness of products was different between FOPL conditions and the control group. Comparison between Nutri-Score and Multiple Traffic Light was also conducted to see if there was a difference in the accuracy of ranking. The results set out in points 2-4 in the findings below were adjusted to take into account a range of key demographic factors including education, ethnicity, household composition, plus food shopping responsibility and current FOPL use. This is important as it allows us to be more confident that

the differences we are seeing are a result of labels themselves, rather than factors relating to participants.

Findings

4,863 participants responded to the survey (a response rate of 67%) and 4,530 participants with complete data were included in this analysis.

Our preliminary results clearly show that FOPLs are effective at improving the participants' ability to identify the healthiness of foods. Compared to the control group, the Multiple Traffic Light, Nutri-Score and Warning Label conditions significantly improved the proportion of participants correctly carrying out the ranking tasks across all food categories, whereas those in Positive Tick condition only improved for certain products. The results indicated that the Nutri-Score label performed the best, closely followed by the Multiple Traffic Light and then Warning Label, with Positive Tick performing worst.

We found:

- Overall, Multiple Traffic Light had the highest proportion of participants who reported having enough information to do the ranking tasks, followed in order by Nutri-Score, Warning Label, Positive Tick and control. It should be noted that since the Positive Tick is a binary label, it was not possible to have distinct labels across the three products within food categories, this is likely to have made the ranking more difficult.
- Groups which viewed the Nutri-Score, Multiple Traffic Light or Warning Label FOPL had a significantly higher likelihood of participants correctly ranking all products correctly compared to groups who did not see a FOPL (the control condition). This was only the case for the Positive Tick FOPL for some product categories (drink, yoghurt and cereal). These analyses were additionally adjusted for baseline ranking score.
- Those who viewed the Nutri-Score, Multiple Traffic Light and Warning Label FOPLs were more likely to improve their ability to rank products correctly from baseline (no label) to follow-up (with labels), compared to the control condition (no labels for all presentations). For the Positive Tick FOPL, this was only the case for drinks.
- When we looked across the change in rankings across all food categories combined, making a global food score, the Multiple Traffic Light, Nutri-Score and Warning Label FOPLs all showed improved ability to rank with an increased global score, compared to the control group; participants in the Positive Tick group did not show such improvements. The Nutri-Score FOPL resulted in a significantly greater improvement in global score compared to the Multiple Traffic Light FOPL.

Background

What are front of pack labels?

Front of pack labels (FOPLs) provide a simple pictorial representation of the detailed nutrient declaration on the back. FOPLs are useful for comparing products at a glance. They aim to provide clearly visible information to guide consumer choice, quickly and easily, in an accessible format. In the UK, FOPLs are voluntary, whereas nutrition labelling requirements for the back of pack are mandatory.

A range of different formats have been used for FOPLs. The current UK FOPL scheme was introduced and recommended for use by the UK Government and administrations in Scotland, Wales and Northern Ireland in 2013 on a voluntary basis, one of the first to be introduced worldwide. It uses traffic lights to indicate the amount of fat, saturated fat, sugar and salt in a food (Multiple Traffic Light label, MTL). This provides easily accessible information about the energy, fat, saturated fat, sugar, and salt content of a food or drink. It has since been widely adopted by many retailers and other food businesses. The MTL label, like many others, is interpretive. This means that it provides a judgement of the nutritional composition of the food, rather than simply providing nutritional information without any comparison to nutritional requirements (non-interpretive).

Policy context

The UK Government made a commitment to review FOPLs as part of the Childhood Obesity Plans published since 2016.[1] The policy paper 'Tackling obesity: empowering adults and children to live healthier lives' was published in July 2020.[2] This outlined the success of the existing front of pack labelling scheme but highlighted the importance of ensuring that, going forwards, the country's labelling scheme be based on the latest evidence and that Brexit offers an opportunity to 'make decisions on labelling which are best for Britain'. This work, which has been carried out alongside a 4-nation public consultation, aims to provide up to date evidence about the impact of FOPLs in a representative GB sample.

How are FOPLs proposed to work?

Two models underpin the current work. The first is the logic model proposed in the Cochrane review on nutritional labelling **(Figure 1)**.[3] This describes FOPLs as improving dietary quality through two main pathways. The first is through improving consumer understanding, therefore encouraging healthier purchasing. The second is via reformulation of food products by the food industry, leading to a healthier diet.

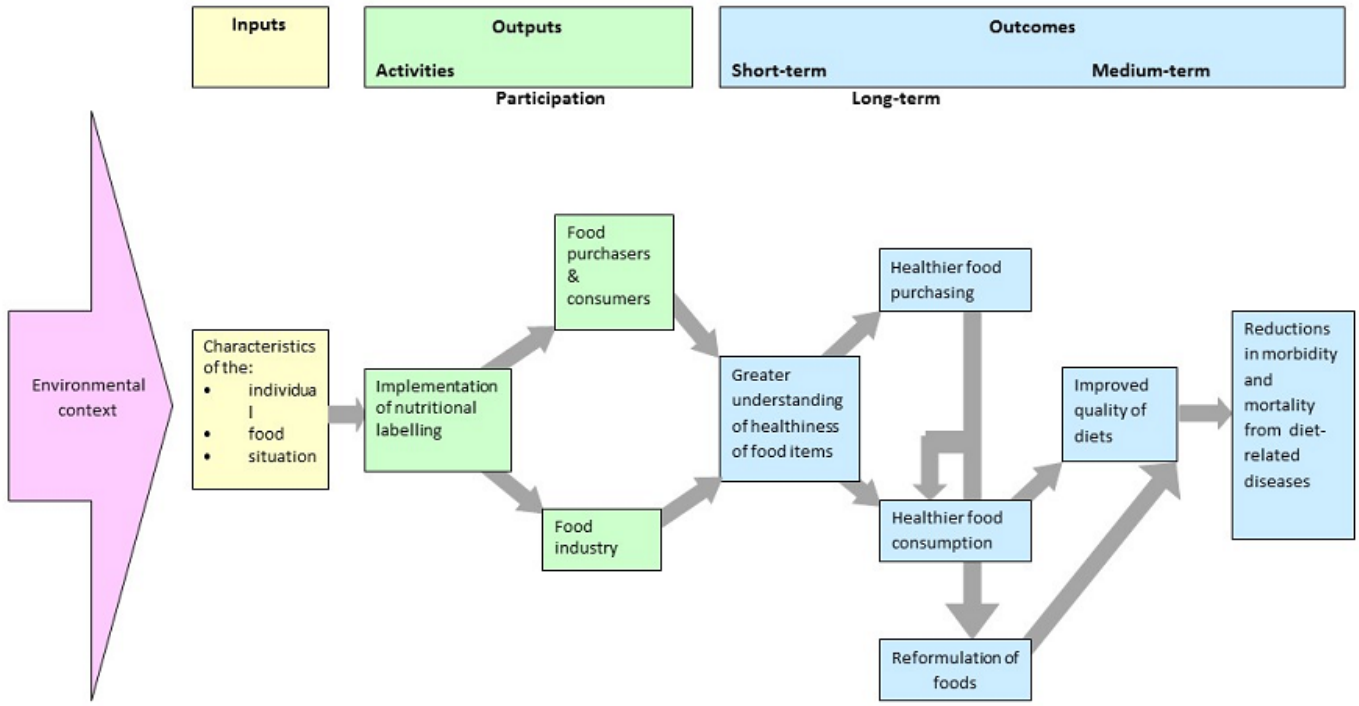
---

[1] Department of Health and Social Care (2018) Childhood obesity: a plan for action, chapter 2 https://www.gov.uk/government/publications/childhood-obesity-a-plan-for-action-chapter-2

[2] Department of Health and Social Care (2020) Tackling obesity: empowering adults and children to live healthier lives. https://www.gov.uk/government/publications/tackling-obesity-government-strategy/tackling-obesity-empowering-adults-and-children-to-live-healthier-lives

[3] Crockett, R. A., et al. (2018) Nutritional labelling for healthier food or non-alcoholic drink purchasing and consumption. https://doi.org//10.1002/14651858.CD009315.pub2

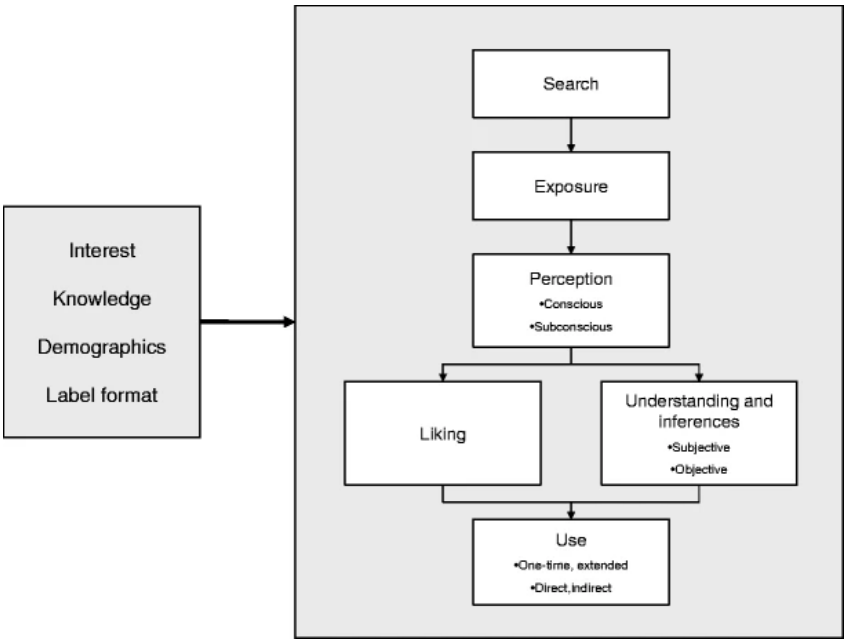
Figure 1. Logic model of the process by which nutritional labelling may have an impact on diets and health (taken from Crockett et al, 2018)



The second model is a theoretical framework based on consumer decision-making **(Figure 2)**.[4] This describes in more detail the pathways to increased understanding and label use. In this model, the outcome is label use (both over the short and longer-term), with the expectation that label use will improve the healthiness of food purchasing. For a label to be 'effective', it must be noticed, liked and used by the consumer. Interest and knowledge (in nutrition), demographics and the format of the label are identified as modifying factors, meaning that these factors are likely to influence how the label is used and its impact.

[4] Grunert, K.G. & Wills, J.M. (2007) A review of European research on consumer response to nutrition information on food labels. https://doi.org/10.1007/s10389-007-0101-9

Figure 2. Theoretical model for response to nutritional information on food labels (taken from Grunert & Wills, 2007)



What do we know about impact?

As indicated by the above models, there are several outcomes which can be studied. We briefly outline below the current evidence for how FOPLs influence knowledge and behaviour (consumption and purchasing).

Knowledge

There is consistent evidence from systematic reviews, which include some UK studies, that FOPLs improve knowledge.[5,6] Recent work has provided additional information about the comparative effects of individual labelling schemes. A study carried out across 12 countries worldwide (including the UK) also found that FOPLs improved knowledge.[7] In this study, the following labels were compared: Health Star Rating (HSR), Multiple Traffic Light (MTL), Nutri-Score (N-S), Reference Intake, Warning Label (WL). Knowledge was assessed by ranking three food types for healthiness (pizza, cake, breakfast cereal). Across all countries and products, N-S performed the best, followed by the MTL, HSR, WL, and Reference Intake. A subsequent study provided data on an additional six European countries and produced similar

[5] Talati, Z., et al. (2017) Consumers' responses to health claims in the context of other on-pack nutrition information: a systematic review. https://doi.org/10.1093/nutrit/nuw070
[6] Hersey, J.C., et al.(2013) Effects of front-of-package and shelf nutrition labeling systems on consumers. https://doi.org/10.1111/nure.12000
[7] Egnell, M., et al. (2018) Objective Understanding of Front-of-Package Nutrition Labels: An International Comparative Experimental Study across 12 Countries. https://dx.doi.org/10.3390%2Fnu10101542

results.[8] Another recent study in a Canadian sample, also conducted online, compared MTL, HSR and WL.[9] In this study, all labels resulted in participants being able to more accurately identify unhealthy products and the MTL and HSR labels both enabled more accurate identification of healthy products. It is therefore clear that FOPLs improve knowledge, with signs that N-S has the greatest impact on knowledge. However, MTL labels also improve knowledge and, in the studies discussed here, have performed better than some other labels. However, evidence for effectiveness in improving knowledge in UK populations is limited, especially in relation to the comparative effect of individual FOPL schemes.

Behaviour

We recently undertook a systematic review on the impact of FOPLs on consumption and purchasing behaviours. The review was a partial update of the Cochrane review published in 2018, with a focus on the impact of FOPLs on pre-packaged foods.[10] The Cochrane review identified six experimental studies measuring food consumption in response to exposure to different FOPLs on pre-packaged foods; no effect on consumption was observed in this meta-analysis but the studies were deemed to be of poor quality. [10] In our systematic review, we identified eight studies reporting purchasing outcomes, three studies (from two articles) reporting consumption outcomes and three 'interrupted time series' studies.[11] The studies which reported consumption outcomes were of poor quality and were not able to be meta-analysed. However, the experimental studies reporting purchasing outcomes were either good or moderate quality. Meta-analysis of these studies indicated a significant reduction in the sugar content of purchased food with a FOPL compared to no FOPL (-0.40g sugar/ 100g food, $p<0.01$) and a trend for reduced energy content (-2.03kcals/ 100g food, $p=0.08$). In terms of the impact of specific FOPLs, we found that compared to no FOPL, the 'high in' FOPL significantly reduced the sugar content of purchases (-0.67g/ 100g, $p<0.01$) and the energy content of purchases (-4.43kcals/ 100g, $p<0.05$). However, the studies did not allow us to make direct comparisons between individual FOPL schemes. Interrupted time series studies take advantage of natural experiments in real-life settings using data before and after implementation of an 'intervention' (in this case a labelling scheme), but without a control condition. Narrative synthesis of the studies in our review also indicated a positive effect on purchasing after the implementation of a labelling scheme. These were all high quality studies. However, only one of the experimental studies and one of the interrupted time series studies took place in the UK, which limits the application of these findings.

Engagement with FOPL

In order for FOPLs to be effective, consumers need to engage with them when making food choices. Evidence suggests that engagement with FOPLs is influenced by comprehension and motivation to use

[8] Egnell, M., et al. (2020) Objective understanding of the Nutri-score front-of-pack label by European consumers and its effect on food choices: an online experimental study. https://doi.org/10.1186/s12966-020-01053-z
[9] Vanderlee, L., et al. (2020) The efficacy of 'high in' warning labels, health star and traffic light front-of-package labelling: an online randomised control trial. https://doi.org/10.1017/s1368980020003213
[10] Crockett, R. A., et al. (2018) Nutritional labelling for healthier food or non-alcoholic drink purchasing and consumption. https://doi.org//10.1002/14651858.CD009315.pub2
[11] Croker, H., et al. (2020) Front of pack nutritional labelling schemes: a systematic review and meta-analysis of recent evidence relating to objectively measured consumption and purchasing. https://doi.org/10.1111/jhn.12758

the relevant nutritional information.[12,13] There are numerous other factors which might influence engagement with FOPLs including factors relating to the individual (e.g. socio-economic status, age, sex, ethnicity) and contextual factors (e.g. food packaging, size, and other information included on the front of pack).

In summary, the current literature focuses on the impact of FOPLs on knowledge and purchasing behaviours. There is evidence to support FOPLs improving both knowledge and the healthiness of purchased foods. Some studies have indicated that N-S brings the greatest improvements in knowledge, but there is very limited evidence on this from UK populations. It is unclear which label schemes perform the best in terms of influencing purchasing and little information on whether FOPLs change behaviour in UK populations specifically. There is also limited information on whether socio-demographic or contextual factors influence any effects. Therefore, research is needed to inform public policy around regulatory actions which could encourage and enable healthier food choices in the UK.

Study aims

The current study focuses on knowledge, as the precursor of other behaviours. We operationalised knowledge as the ability to determine the healthiness (based on nutritional composition) of specific food items. This study therefore examined whether FOPLs impacted ability to identify the healthiness of foods and drinks in a large population-based UK sample (including England, Scotland, Wales), and explored whether this was influenced by socio-demographic and contextual factors.

The current briefing paper reports preliminary findings relating to the below aims. Future papers will report on the influence of socio-demographic and contextual factors on outcomes.

Primary aims

a) To identify if specific front of pack labels (MTL, N-S, WL, PT) are effective at improving participants' ability to determine the healthiness of food items.
b) To identify if for each label there is a difference in the ability to determine the healthiness of food items, compared to a no label control condition.

Secondary aims

a) To identify if N-S label is more or less effective than MTL at enabling accurate product ranking.

Methods

Participants

Participants were recruited using NatCen's panel, a nationally representative, probability-based sample of adults in Great Britain. Participants were required to be 18+, able to complete the survey online and able to read and write in English. An email inviting all active panel members was sent by NatCen and

[12] 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. https://doi.org/10.1016/j.appet.2011.04.011
[13] Vyth, E.L., et al. (2010) Actual use of a front-of-pack nutrition logo in the supermarket: consumers' motives in food choice. https://doi.org/10.1017/S1368980010000637

participants were offered a small financial incentive on completion. Recruitment took place from the 28th October to the 15th November 2020.

Procedure

Experimental design

The experiment was carried out entirely online and on one occasion. We measured 'ability to determine the healthiness of food items' by presenting participants with three images of a particular food or drink product and asking them to rank these in terms of how healthy they considered each to be (from most healthy to least healthy). Participants were required to rank all three products in the correct order to be classified as 'correct'. The correct ranking of the foods was determined by PHE using the foods' 2004/5 Nutrient Profiling Model scores. We selected six food/ drink categories, in consultation with the Steering Group (see **Acknowledgements** for members). These were required to have adequate variability in nutritional profile to allow us to create three distinct products within each category, and also cover a range of different types of food and drinks.

Food categories selected: pizza, instant hot chocolate, cake, crisps, yoghurt, breakfast cereal

(we refer to instant hot chocolate as 'drink' and breakfast cereal as 'cereal' for brevity in this paper)

We worked with a graphic designer to develop 'mock packaging' images for each product. In total, 18 images were created (three products for each of the six food categories). These were designed so that individuals with basic nutritional knowledge would be able to rank the products within a particular category in order of healthiness. Nutritional claims, allergen information and other information (such as suitable for vegetarians/ organic) which could influence participants' responses were excluded. Product information was only included when required to distinguish the products (e.g. use of word 'baked' to indicate 'healthier' crisps; 'four cheese margherita' to indicate a less healthy pizza). Product weights were consistent across products within a category.

How did we manipulate the images to make them appear more or less healthy?

- Colour - lighter for most healthy products (e.g beige, light pink, white) and darker for least healthy products (e.g. dark brown, bright red); this is in line with current available 'healthier' products often being a lighter colour.
- Product characteristics - plain for most healthy products (e.g. thin and crispy, baked, no icing, grainy) and richer for least healthy products (e.g. chocolate icing, cheese and oil, chocolate swirls)

Each of the 18 products was assigned a set of FOPLs based on their nutritional profile, done in consultation with PHE. Criteria set out in technical guidance documents were used to produce the MTL[14] and N-S labels.[15] Warning labels for sugar, total fat, salt, and saturated fat were assigned where a food

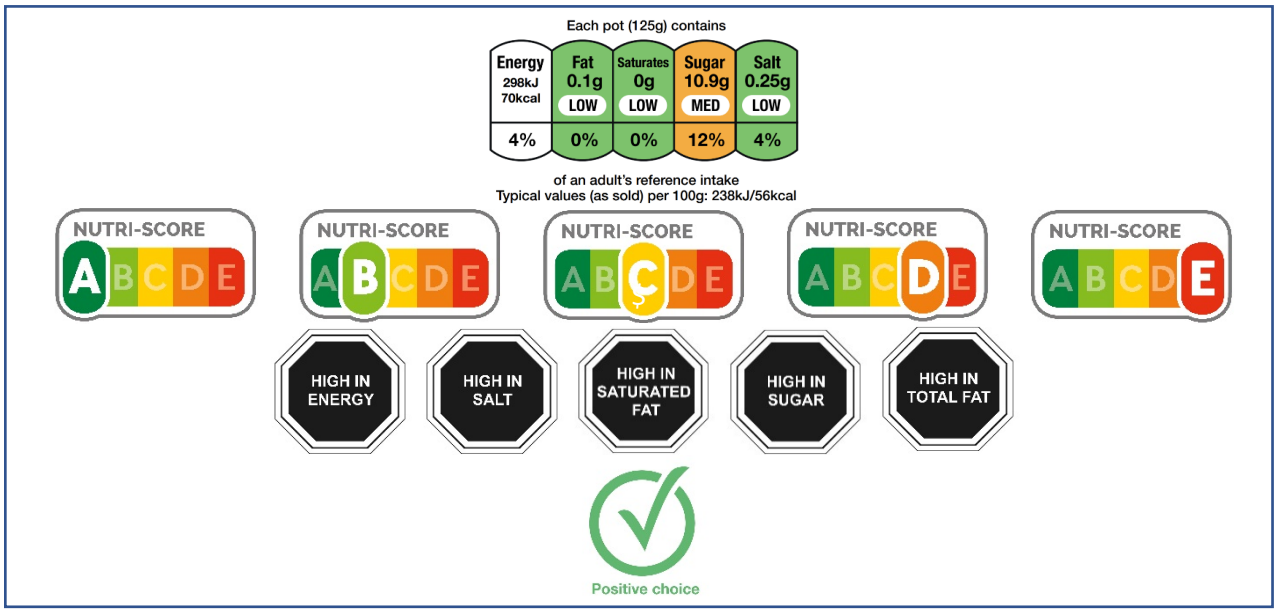
[14] UK Gov (2016) Guide to creating a front of pack (FoP) nutrition label for pre-packed products sold through retail outlets https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/566251/FoP_Nutrition_labelling_UK_guidance.pdf

[15] Santé Publique France (2020) https://partage.santepubliquefrance.fr/public/folder/MMGg-LaV6UiJ0BAm1Bb5PA/Pack_logo_hzt

qualified for a 'red' MTL label for that nutrient. An energy WL was assigned where calories were greater than >275kcal/100g for food products and >70kcal/100ml for drinks (thresholds at 36 months), except for products that crossed the threshold but without added sugars or saturated fat, as per Chilean guidance.[16] The PT label was assigned where the food did not have any 'red' MTL labels and did not breach energy thresholds. We aimed for the three foods within a category to have distinct label profiles, although this was not possible for the PT label. See **Appendix 1** for the images and FOPLs of the 18 food and drink products used in the experiment.

Participants were randomised to one of five FOPL conditions: MTL, N-S, WL, PT or control (no label), examples of FOPLs are shown in **Figure 3** below. Randomisation was stratified on key variables: year of recruitment to panel, sex, age, government office region, and household income. The primary outcome measures were based on the ranking tasks, where participants viewed images of three products and had to rank them in order of healthiness, from most healthy to least healthy. Participants completed the ranking tasks once with no FOPLs (baseline ranking tasks) and again with the images displaying the FOPL as per their assigned condition (follow-up ranking tasks). A total of 12 ranking tasks were completed by each participant (6 product categories x 2 ranking conditions- baseline and follow-up). Participants could not rank two products the same (an error message appeared, asking them to go back and rank one product per category). The order of the product categories and order of products within categories were randomised, to limit presentation order effects. We created a 'change in global food score', each food category could score +1 (if their ranking improved), -1 (if their ranking got worse) or 0 (if their ranking stayed the same), giving an overall score in the range of +5 (if all five food categories improved) to -5 (if all five food categories got worse).

Figure 3. Examples of all front of pack labels: Multiple Traffic Lights, Nutri-Score, Warning Labels, Positive Tick, from top to bottom



[16] Republica de Chile Ministerio de Salud (2019) Reglamento sanitario de los alimentos Article 120, pg 49 https://dipol.minsal.cl/wp-content/uploads/2019/06/DECRETO_977_96_2019-2.pdf

To minimise missing data, participants were not initially given a 'don't know' option to the ranking questions, but this option was presented if they tried to move onto the next page without completing the ranking. After each ranking task, participants were asked if they had enough information to do the ranking. After completing each set of ranking tasks (i.e. at both baseline and follow-up), participants were asked how confident they were in their rankings. After the follow-up ranking tasks, participants were asked if they had seen the FOPLs and if they had used the FOPLs to complete the ranking task, excluding those in the control condition. They were then asked how easy/difficult they found the labels to understand, if they wanted the labels on packaging in the UK, how helpful in choosing what to buy they'd be, and views on how long it takes to use the labels.

Current food shopping and eating habits were assessed by asking food shopping responsibility, label use when shopping, influence of shopping, healthy eating knowledge, healthy eating interest, if currently trying to lose weight. We collected other participant information such as height and weight, pregnancy status, physical or mental health conditions that affect vision/learning, understanding or concentration/ diet, English as second language, and household composition (children under 16 years). As NatCen panel members, certain demographic information was already known (sex, age), and other key demographic information were only asked if it had been more than six months since last updated (income, educational level). It is important to note that, apart from the ranking tasks, all other measures were self-reported.

The outcomes of interest for the current briefing paper are outlined below (other outcomes will be reported in future papers):

- Participants' ability to rank the products within each category in order of healthiness (correct if all products ranked correctly, incorrect if any wrong)
- Change in ability to correctly rank the products within each category from baseline (no FOPL) to follow-up (FOPL shown) (participants classified as improved, didn't improve, worsened)
- Change in global food score, aggregating change in rankings for the five food products (score range is -5, 5).
- We also report descriptively whether participants reported seeing the FOPLs and whether they reported having enough information to do the rankings.

Statistical analysis

Logistic regression analysis was used to compare the proportion of correct order of healthiness rankings at follow-up, adjusted for baseline rankings between each of the FOPL conditions and control condition. Ordinal logistic regression was used to compare the change in rankings from baseline to follow-up and linear regression analysis was used to compare the change in the global food score from baseline to follow-up, between the FOPL conditions and control. All models were adjusted for the five stratification factors and the following pre-specified covariates: ethnicity, highest education level, household composition, food shopping responsibility and current FOPL use. Participants who self-reported not buying or consuming products from a product category in the last 12 months were excluded from analyses specific to that product, to ensure lack of familiarity with products did not impact results. For assessment of the change in global food score only participants who reported buying or consuming all five food products were included. All models were weighted to account for non-response and to ensure

findings are representative of the British adult population. Associations between participants who reported seeing the labels following the ranking tasks and ranking outcomes are described. Stata v16 was used for all analysis.

Ethical approval and study documentation

Full details of the study protocol and data analysis plan are available on Open Science Framework (https://osf.io/k9v2p/) (NB/ this is not currently publicly accessible, but we will amend this once the academic paper is submitted for publication). Both of these developed in advance of conducting the experiment. Ethical approval was granted by NatCen Research Ethics Committee (application reference: P15640).

Results

4,863 participants responded to the online survey, a response rate of 67% (7,218 participants were approached, as indicated by NatCen). 4,530 participants were included here as the analysis sample, as they had complete stratification and covariate information available (see **Appendix 2**). Participants who reported buying or consuming each product category are also presented in **Appendix 2**, a requirement for being included in the analysis of each food category. Demographic characteristics, along with other individual characteristics, are provided in **Table 1**. Overall, 57.1% of participants were female, 93% were white British or white other, 48.5% had a degree or higher, 30.4% had children in the household, 95.8% were responsible or partly responsible for food shopping, and 55.0% reported currently using food labels very or quite often. The demographic characteristics of the sample were broadly comparable to the full NatCen panel and to the UK population.

The number of participants who reported seeing the FOPLs is shown below in **Table 2**. Overall, 74.4% reported seeing the FOPLs that they were randomised to, this differed by FOPL condition. N-S (87.3%) had the highest recall rate followed by WL (78.0%), MTL (76.9%) and PT (55.0%). The number of participants who reported needing more information to rank each of the products at follow-up is presented below in **Table 3**. Overall, across each of the products, the PT and control groups had the lowest proportion of participants reporting having enough information, followed by WL, N-S and MTL the highest.

Table 1. Individual characteristics of the analysis sample (*n* = 4,530; unweighted)

	***n* (%)**
Sex	
Female	2,586 (57.1)
Male	1,944 (42.9)
Age	
18-29	279 (6.2)
30-39	642 (14.2)
40-49	881 (19.5)
50-59	969 (21.4)
60-69	933 (20.6)
70+	823 (18.2)
Ethnicity	
White British	3,974 (87.3)
White other	259 (5.7)
Mixed or multiple ethnic groups	54 (1.2)
Asian or Asian British	155 (3.4)
Black or Black British	72 (1.6)
Other	16 (0.4)
Education	
Degree or equivalent +	2,196 (48.5)
A levels or vocational level 3 or equivalent	878 (19.4)
Other qualifications below A levels or equivalent	784 (17.3)
Other qualification	271 (6.0)
No qualifications	401 (8.9)
Children in household	
Yes	1,377 (30.4)
No	3,153 (69.6)
Shopping responsibility	
Yes – some or all	4,340 (95.8)
No – someone else does	190 (4.2)
Current label use	
Very often	945 (20.9)
Quite often	1,546 (34.1)
Occasionally	1,318 (29.1)
Rarely	559 (12.3)
Never	162 (3.6)

Table 2. Results of FOPL recall across the study sample, excluding control (*n* = 3,617)

Reported seeing FOPLs during the survey*, overall	**n (%)**
Yes	2,692 (74.4)
No or Not sure	925 (25.6)
Reported seeing FOPLs during the survey*, by allocated FOPL condition (Yes response)	
MTL	697 (76.9)
N-S	807 (87.3)
WL	698 (78.0)
PT	490 (55.0)

*Control condition not applicable

Table 3. Number and proportion of participants who reported having enough information to rank the products at baseline, by product and FOPL condition (*n* = 4,530)

	Reported having enough information to rank (at follow-up) (n=4,530)					
	Pizza n (%)	**Drink n (%)**	**Cake n (%)**	**Crisps n (%)**	**Yoghurt n (%)**	**Cereal n (%)**
Control	86 (12.8)	22 (7.0)	88 (14.0)	122 (16.2)	86 (11.3)	89 (11.4)
MTL	550 (85.0)	276 (82.1)	546 (83.9)	647 (87.6)	647 (85.7)	649 (85.7)
N-S	405 (59.0)	180 (55.6)	379 (59.9)	472 (61.1)	444 (57.5)	455 (60.3)
WL	243 (35.2)	111 (33.0)	279 (43.9)	317 (42.1)	231 (30.9)	253 (33.5)
PT	67 (10.0)	22 (6.9)	70 (11.4)	93 (13.2)	86 (11.6)	87 (11.5)

The proportions correctly ranking products at baseline were similar across the FOPL conditions at baseline. The number of participants who correctly ranked the three products by food category and FOPL condition is presented in **Figure 4** below. It is important to note that none of the crisps or cakes qualified for the PT FOPL, so these foods had no FOPL to guide participants' ranking.

Figure 4. Correct ranking at baseline and follow-up, by FOPL condition and product category

† Cake and crisps categories had no products qualify for Positive Tick

The associations between FOPLs and correct product healthiness ranking at follow-up, adjusted for covariates, are shown below in **Table 4**. We found that the likelihood of participants correctly ranking the products at follow-up was significantly greater when N-S, MTL or WL FOPL were shown, compared with the control condition (who had no FOPL). N-S was associated with the greatest odds followed by MTL, WL and PT. For example, for pizza, the OR is 9.20 (95%CI 5.70, 14.86, p-value < 0.001), meaning there is a nine times higher likelihood of participants in the N-S condition correctly ranking the product, compared with participants in the control condition. This was highly statistically significant. For the PT condition, in comparison with the control condition, there were significant differences in the correct rank rate for drink, yoghurt, cereal categories; there was no difference for pizza. Furthermore, there was no difference between control and PT for either cake or crisps, where none of the three products qualified (essentially no label condition). Larger differences in the OR between control and FOPL condition at baseline were seen for yoghurt and, to a lesser extent, cereal categories, showing differential impact by food type. The comparison between N-S and MTL showed that participants in the N-S group were significantly more likely than the MTL group to correctly rank the pizzas and drinks, but the associations were not significant for cake, crisps, yoghurt and cereal.

Table 4. Logistic regression results - follow-up (FOPL condition) correct (yes /no) and adjusted for baseline rank (also adjusted for design factors and covariates as planned)

	MTL v Control OR (95%CI)	**N-S v Control OR (95%CI)**	**WL v Control OR (95%CI)**	**PT v Control OR (95%CI)**	**N-S v MTL OR (95%CI)**
Pizza	3.94 (2.37, 6.54) **P < 0.001**	9.20 (5.70, 14.86) **P < 0.001**	2.69 (1.82, 3.99) **P < 0.001**	1.23 (0.87, 1.74) 0.24	2.34 (1.30, 4.21) **P < 0.01**
Drink	5.38 (2.85, 10.14) **P < 0.001**	23.60 (11.83, 47.10) **P < 0.001**	3.38 (1.85, 6.17) **P < 0.001**	2.09 (1.29, 3.38) **P < 0.01**	4.39 (2.06, 9.35) **P < 0.001**
Cake	10.31 (5.75, 18.46) **P < 0.001**	12.47 (6.73, 23.12) **P < 0.001**	7.27 (4.30, 12.30) **P < 0.001**	1.27† (0.87, 1.85) 0.21	1.21 (0.63, 2.32) P = 0.57
Crisps	9.71 (6.09, 15.47) **P < 0.001**	15.21 (8.73, 26.49) **P < 0.001**	4.96 (3.34, 7.35) **P < 0.001**	1.28† (0.92, 1.79) 0.14	1.57 (0.86, 2.84) P = 0.14
Yoghurt	57.53 (37.74, 87.71) **P < 0.001**	64.22 (42.51, 97.01) **P < 0.001**	17.39 (11.80, 25.62) **P < 0.001**	1.63 (1.11, 2.41) **P = 0.01**	1.12 (0.75, 1.67) P = 0.59
Cereal	17.96 (12.57, 25.67) **P < 0.001**	23.58 (15.89, 35.01) **P < 0.001**	9.12 (6.29, 13.24) **P < 0.001**	1.55 (1.15, 2.10) **P < 0.01**	1.31 (0.87, 1.99) P = 0.20

† *Cake and crisps categories had no products qualify for Positive Tick; all analyses were adjusted for the five stratification factors (year of recruitment to panel, sex, age, government office region, household income) and the following pre-specified covariates: ethnicity, highest education level, household composition, food shopping responsibility, and current FOPL use.*

The associations between FOPLs and an improved change score, adjusted for covariates, are shown in **Table 5**. We found that N-S, MTL and WL all significantly increased odds of participants improving their score from baseline to follow-up compared to control. The odds of improved change were similar across N-S, MTL and WL, with N-S slightly outperforming followed by MTL and then WL. PT was associated with greater odds of improved score in drinks and yoghurts only (again, cake and crisps had no products with PT). The comparison between N-S and MTL showed that N-S significantly increased the odds of improving in score compared to MTL for drinks but none of the food categories.

Table 5. Ordinal logistic regression results - odds that ranking improved (follow-up vs baseline) between FOPL condition and control adjusted for design factors and covariates as planned

	MTL v Control OR (95%CI)	**N-S v Control OR (95%CI)**	**WL v Control OR (95%CI)**	**PT v Control OR (95%CI)**	**N-S v MTL OR (95%CI)**
Pizza	2.58 (1,69, 3.95) **P < 0.001**	3.19 (2.24, 4.55) **P < 0.001**	2.24 (1.54, 3.27) **P < 0.001**	0.95 (0.67, 1.35) 0.79	1.24 (0.86, 1.79) P = 0.25
Drink	3.15 (1.89, 5.24) **P < 0.001**	6.33 (4.10, 9.77) **P < 0.001**	2.01 (1.21, 3.34) **P < 0.001**	1.93 (1.21, 3.05) **P < 0.01**	2.01 (1.27, 3.18) **P < 0.001**
Cake	6.25 (3.97, 9.84) **P < 0.001**	5.76 (3.74, 8.88) **P < 0.001**	4.86 (3.13, 7.56) **P < 0.001**	1.16† (0.79, 1.71) 0.46	0.92 (0.65, 1.31) P = 0.65
Crisps	4.84 (3.46, 6.77) **P < 0.001**	5.65 (4.01, 7.95) **P < 0.001**	2.98 (2.15, 4.13) **P < 0.001**	1.17† (0.83, 1.65) 0.38	1.17 (0.85, 1.60) P = 0.34
Yoghurt	35.65 (24.50, 57.58) **P < 0.001**	39.81 (27.53, 57.58) **P < 0.001**	14.43 (9.84, 21.15) **P < 0.001**	1.70 (1.24, 2.33) **P < 0.001**	1.12 (0.81, 1.54) P = 0.50
Cereal	8.41 (6.35, 11.13) **P < 0.001**	10.01 (7.42, 13.49) **P < 0.001**	6.51 (4.65, 9.10) **P < 0.001**	1.29 (0.96, 1.73) 0.09	1.19 (0.91, 1.5) P = 0.21

† *Cake and crisps categories had no products qualify for Positive Tick; all analyses were adjusted for the five stratification factors (year of recruitment to panel, sex, age, government office region, household income) and the following pre-specified covariates: ethnicity, highest education level, household composition, food shopping responsibility, and current FOPL use.*

The associations between the global food score (across the five food categories) and the FOPL conditions, adjusted for covariates, are shown below in **Table 6**. We found that MTL, N-S and WL were significantly associated with an increase in global score compared to control. For example, for N-S the co-efficient was 2.1 (95%CI 1.9, 2.2, p-value < 0.001), meaning the N-S condition scored 2.1 more points compared to the control condition, showing greater improvement across the rankings. PT did not result in any significant change in global score. N-S was found to significantly increase the global score by 0.3 points (95%CI 0.2, 0.5; p<0.001), compared to MTL. Note that the global score is a continuous outcome, as opposed to the categorical outcomes presented in Tables 4 and 5. It has a possible range of -5 to +5, a higher score indicates more accurate ranking, but the actual numbers cannot be quantified.

Table 6. Multivariate regression analysis results- association between global food score and FOPL condition (adjusted for design factors and covariates as planned)

	MTL v Control OR (95%CI)	N-S v Control OR (95%CI)	WL v Control OR (95%CI)	PT v Control OR (95%CI)	N-S v MTL OR (95%CI)
Score (-5, +5) Regression (coefficients)	1.7 (1.6, 1.9) **P < 0.001**	2.1 (1.9, 2.2) **P < 0.001**	1.4 (1.3, 1.6) **P < 0.001**	0.1 (-0.02, 0.3) 0.09	0.3 (0.2, 0.5) **P < 0.001**

Global food score was an aggregated score of correct ranking in the five food products, range -5 to +5 (where -5 denotes a worsening in ranking in all five food products and +5 denotes an improvement in all five food products); all analyses were adjusted for the five stratification factors (year of recruitment to panel, sex, age, government office region, household income) and the following pre-specified covariates: ethnicity, highest education level, household composition, food shopping responsibility, and current FOPL use.

Discussion

In this online experimental study, we found that FOPLs significantly improved the ability of participants to rank products based on healthiness. All of the FOPL conditions showed significant improvements from baseline to follow-up for at least one food category, and ranking improved across all food categories for the MTL, N-S and WL conditions. Patterns were similar when making comparisons between specific FOPL conditions and the control condition; all conditions had significantly better ranking compared to the control condition for at least one food category with MTL, N-S and WL being significantly better across all food categories. The magnitude of the ORs indicates the largest effects were in the N-S condition, followed by MTL, WL and then PT. However, to avoid multiple testing, we only statistically tested the N-S and MTL associations; this reduces the risk of type 1 errors (false positives). These findings are consistent with the existing literature, where N-S labels have tended to out-perform other FOPLs.[17,18] The results here suggest that the N-S label was only marginally superior to the MTL, this could be due to the UK population being more familiar with this particular label.

Overall, three quarters of the participants who had been randomised to a FOPL condition recalled seeing the FOPL. The proportion recalling seeing the FOPLs was markedly lower in the PT condition, this is unsurprising since several of the foods did not qualify for a PT label so there would have been no FOPL present.

The proportion of participants reporting having enough information to do the ranking varied considerably across FOPL conditions. Most participants (>80%) in the MTL condition reported having enough information, with just over half of the participants in the N-S condition and approximately a third in the WL having enough information. Those in the PT and control conditions had the lowest

[17] Egnell, M., et al. (2018) Objective Understanding of Front-of-Package Nutrition Labels: An International Comparative Experimental Study across 12 Countries. https://dx.doi.org/10.3390%2Fnu10101542
[18] Egnell, M., et al. (2020) Objective understanding of the Nutri-score front-of-pack label by European consumers and its effect on food choices: an online experimental study. https://doi.org/10.1186/s12966-020-01053-z

proportions reporting having enough information (<20%). It is interesting to note that being able to rank does not appear to be related to perceiving having adequate information.

We have extended the findings of others by including a control condition to strengthen the design. This has allowed us to demonstrate a minimal learning effect between baseline and follow-up ranking tasks, so the observed effects can be confidently attributed to the FOPLs. We conducted a high quality and well powered experimental study, in a representative UK sample, with the protocol and analysis plan specified a priori. The fact that analyses were adjusted for covariates indicates that the effects seen were independent of level of education, age, sex etc. There were however some limitations. The PT label was problematic since it was impossible to provide participants with adequate information to perform the ranking tasks; it is a binary label so it is only able to discriminate between two products, and in the current study, two of the food categories did not qualify for any PT FOPLs. The numbers in our analysis of the global score were reduced to 43.6% (n=1976) as we included only participants who had purchased all five food products, but we will consider this for future analyses. Creating mock images for the food products allowed us to control variability between the foods within each category, but it is possible that this reduced the ecological validity of the study. The proportions of participants correctly ranking products varied between the food products at baseline; it was lowest for yoghurts and cereal (especially striking for yoghurts). This resulted in large increases in correct ranking at follow-up for these products (and the large ORs seen). This suggests that these products were more difficult to rank without a FOPL, despite us designing the images to facilitate correct ranking.

Future analyses will focus on the impact of participant characteristics on ability of participants to correctly identify product healthiness (age, socio-economic status, ethnicity, health conditions, children in household). This will allow us to understand if different FOPLs are more effective in different UK populations. Attitudes towards the labels and impact of nutrition knowledge/food habits and time taken to rank will also be examined for future papers. As already stated, here, we have presented the results from preliminary analyses, we will additionally address missing data and make any adjustments for multiple comparisons in future analyses.

In conclusion, this work demonstrates that FOPLs are effective at improving knowledge in a large representative UK sample. There were signs that N-S performed the best, closely followed by MTL and then WL. PT labels performed the worst, but all labels were superior to the no label control group.

Acknowledgements

We thank our collaborators Dr Oliver Robinson Institute of Cognitive Neuroscience, UCL) and Dr Sandro Stoffell (Research Department of Behavioural Science and Health, UCL), NatCen for all of their assistance designing and setting up the experiment, Huw Jones from the UCL Graphic Design Team, and all of the panel members who participated. We also thank our Steering Group for all of their input and participation in Steering Group meetings.

Members of Steering Group:

External experts:

Professor Ashley Adamson, Professor of Public Health Nutrition, Newcastle University

Dr Chantal Julia, Professeur des Universités - Praticien hospitalier, Université Sorbonne Paris Nord

Government:

Department of Health and Social Care – Officials from the Science, Research and Evidence Directorate, Nutrition and Healthy Weight Branch and Population Health Analysis Branch.

Public Health England – Officials from the Diet, Obesity and Physical Activity Branch.

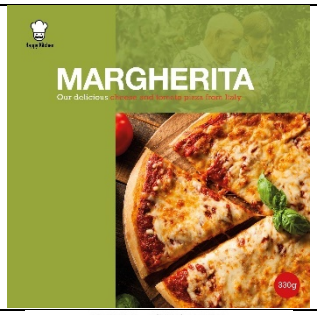
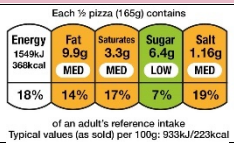
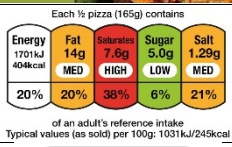
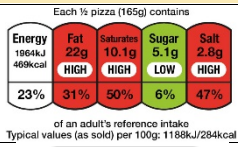
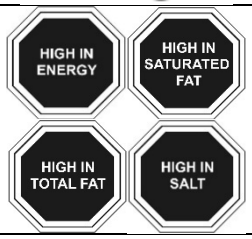
Food Standards Northern Ireland

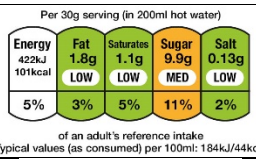
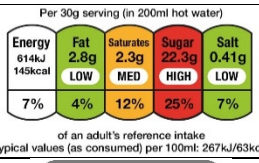
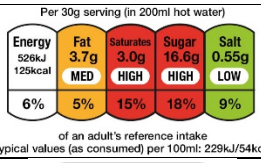
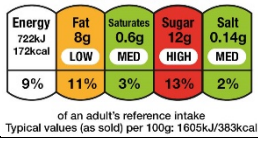
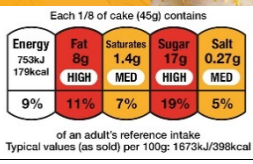
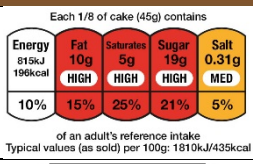
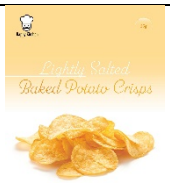
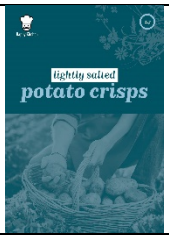
Food Standards Scotland

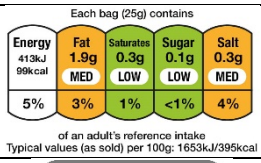
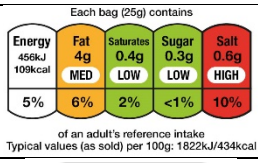
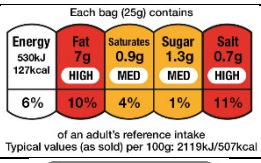
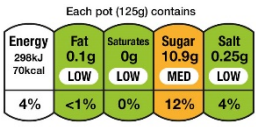
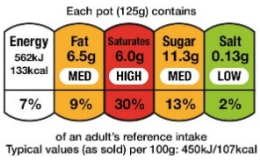
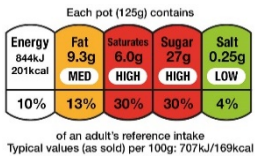
Welsh Government

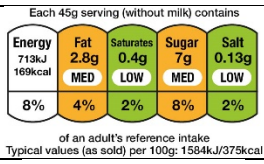
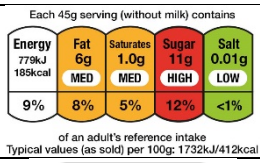
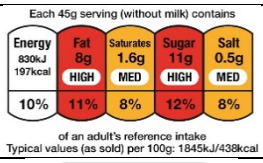
Appendices

Appendix 1. Product images and applicable labels

FOPL condition	Product		
	Pizza		
No label (baseline and control)			
Multiple Traffic Light			
Nutri-Score			
Warning Label			
Positive Tick			
	Drink (instant hot chocolate powder)		
No label (baseline and control)			

Multiple Traffic Light			
Nutri-Score			
Warning Label			
Positive Tick			
	Cake		
No label (baseline and control)			
Multiple Traffic Light			
Nutri-Score			
Warning Label			
Positive Tick			
	Crisps		
No label (baseline and control)			

Multiple Traffic Light			
Nutri-Score			
Warning Label			
Positive Tick			
	Yoghurt		
No label (baseline and control)			
Multiple Traffic Light			
Nutri-Score			
Warning Label			
Positive Tick			
	Cereal		
No label (baseline and control)			

Multiple Traffic Light			
Nutri-Score			
Warning Label			
Positive Tick			

Appendix 2. Flowchart of participants and proportions that consumed each food product