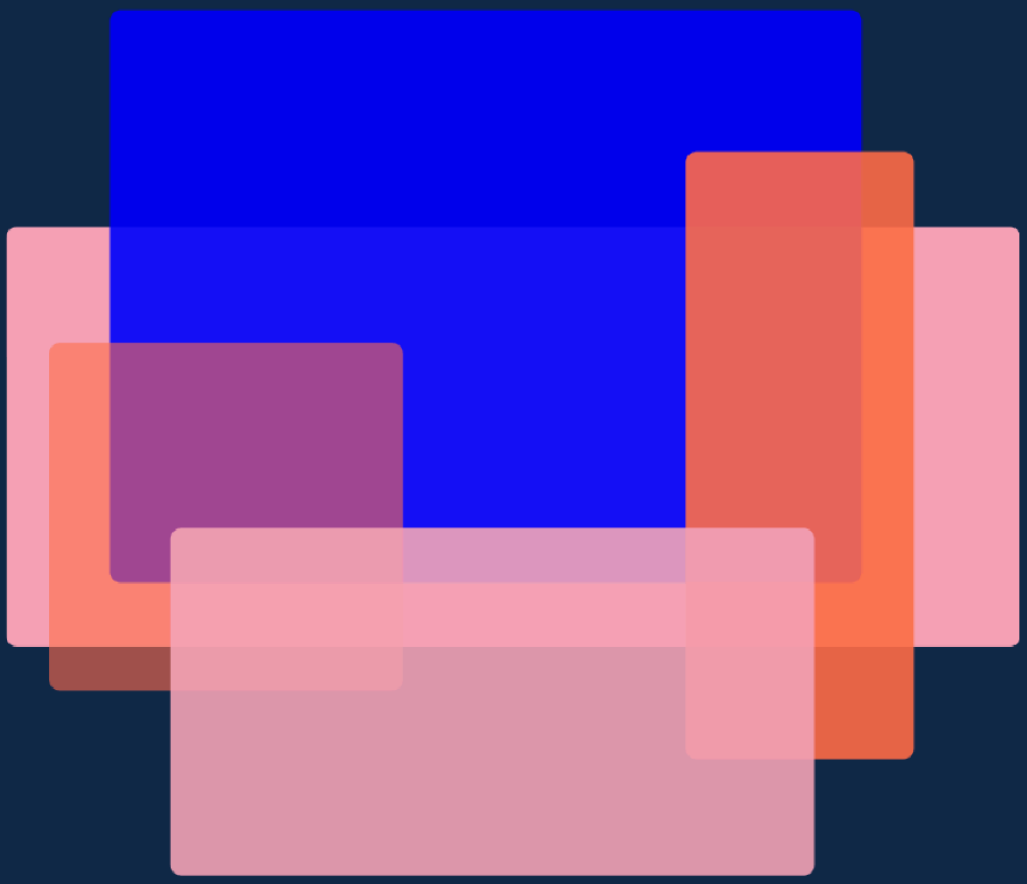
Blueprint
by nesta

Blueprint for halving obesity: rapid review

Front-of-pack labelling (FOPL) interventions for reducing obesity



Author: Anish Chacko

Contents

Summary table

Title	Front-of-pack nutritional labelling schemes: A systematic review and meta-analysis of recent evidence relating to objectively measured consumption and purchasing			Impact of color-coded and warning nutrition labelling schemes: A systematic review and network meta-analysis	Impact of food labelling systems on food choices and eating behaviours: a systematic review and meta-analysis of randomized studies
Author and year	Croker et al. (2020)			Song et al. (2021)	Cecchini & Warin (2016)
Type of study	Systematic review and meta-analysis			Systematic review and meta-analysis	Systematic review and meta-analysis of randomised studies
Outcome variable	Effect in kcal/100g of food or beverages purchased			Calories purchased	Share of people buying healthier options
Treatment	FOPL vs no label	High-in	Multiple Traffic Light	Nutri-Score	Traffic light system (TLS), guideline daily amount (GDA) and other types of food labelling (eg, front-of-pack logos)

Control	No label	No label	No label	No label	No label or other types of FOPL
Magnitude of effect (Adults)	-1.95kcal/100g (*p*>0.05)	-4.43kcal/100g (*p*<0.05)	-4.217kcal/100g (*p*>0.05)	0-54kcals (RMD = -0.06)	Overall, 17.95% increase in share of people buying healthier options, but with high heterogeneity. TLS = 29.36% increase in healthier share; 'other food labels' = 14.69%; GDA at 11.85%
Magnitude of effect (Children)	Not available	Not available	Not available	Not available	Not available
Notes	For modelling the impact of this policy, the review highlighted in the green column was used.				

Rapid umbrella review

Background

Front-of-pack nutritional labelling (FOPL) is a common and prominent policy tool being used by governments to promote healthier eating. Many governments in most high-income countries have already mandated the display of nutritional information on pre-packaged foods. This policy/intervention is likely to work in two ways. It may work as an educational tool to inform customers and producers about the nutritional content in the food and thereby its healthiness. This information is expected to lead to healthier food purchasing and consumption by customers, which motivates reformulation of foods by industry. This eventually leads to healthier diets and reduction in comorbidities such as obesity and other diet-related diseases. The logic model for this intervention is explained in Crockett et al. (2018) and shown in Figure 1 below:

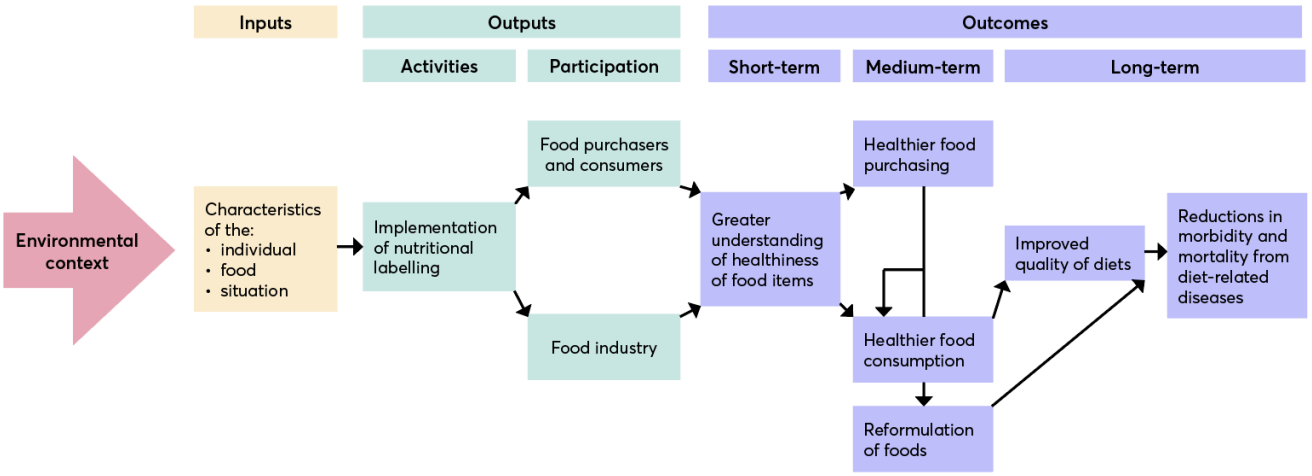


Figure 1: Logic model for intervention based on Crockett et al. (2018)

There are a variety of different front-of-pack nutritional labels that are used in countries around the world. There are many classifications of front-of-pack labels but the two major types are:

- Interpretive: Indicates a product's healthfulness in a graphical manner and simplifies nutritional information in an easy-to-understand manner. These can be either traffic lights, scores or a health logo that is easy to recognise. Some examples are – traffic light labelling, Nutri-Score, guidelines daily advice or reference intakes, nutrition/health warning such as 'High in salt/ sugar', health star ratings, etc.
- Non-interpretive labels: This is numeric in nature and requires customers to interpret by themselves. Adequate levels of literacy should be present for these to be effective and they are typically viewed as being less helpful.

Objective

To summarise the best available evidence on the impact of FOPL on energy intake or body weight.

Methods

We aimed to identify reviews that included quantitative research synthesis (ie, meta-analysis or a government impact assessment) of the effectiveness of FOPL interventions on outcomes relevant to calorie consumption, energy intake, weight loss or obesity. If more than one review was identified that answered our research question, we aimed to identify the review that was reflective of the best evidence, based on (a) year published and (b) quality of review (judged by JBI checklist).

Eligibility criteria

Types of review. To be eligible for inclusion, articles were required to use systematic review methodology (ie, use of systematic search and inclusion strategy to identify all available studies) with randomised control trials and include quantitative data synthesis (ie, meta-analysis) of multiple studies that examined the effect of FOPL on outcomes relevant to calorie consumption, energy intake, weight loss or obesity.

If the search did not identify any studies where a meta-analysis had been conducted due to heterogeneity of outcomes of interest, we intended to include reviews with narrative synthesis or impact assessments from government

departments. We did not set inclusion criteria on the number or type of databases searched in the reviews.

We selected a single review that best represented our research question. If more than one review was identified, we assessed the quality and selected the one with the highest rating (taking into account year of publication). In case of the absence of a single review with a meta-analysis, we included an impact assessment.

Participants. To be eligible for inclusion, articles could examine the effect of FOPL interventions on adults or children.

Intervention. Reviews were required to synthesise FOPL of food products in a retail setting.

Comparator. The comparators were studies with no labels or other labels on food products.

Outcomes. To be eligible for inclusion, reviews needed to include either clinical (eg, weight, BMI, % fat change) or behavioural outcomes (including, but not limited to: purchasing behaviour, consumption behaviour, food diaries). Reviews that only included measures of intentions/plans for future behaviour were excluded due to evidence of the gap between intended and actual eating behaviour.

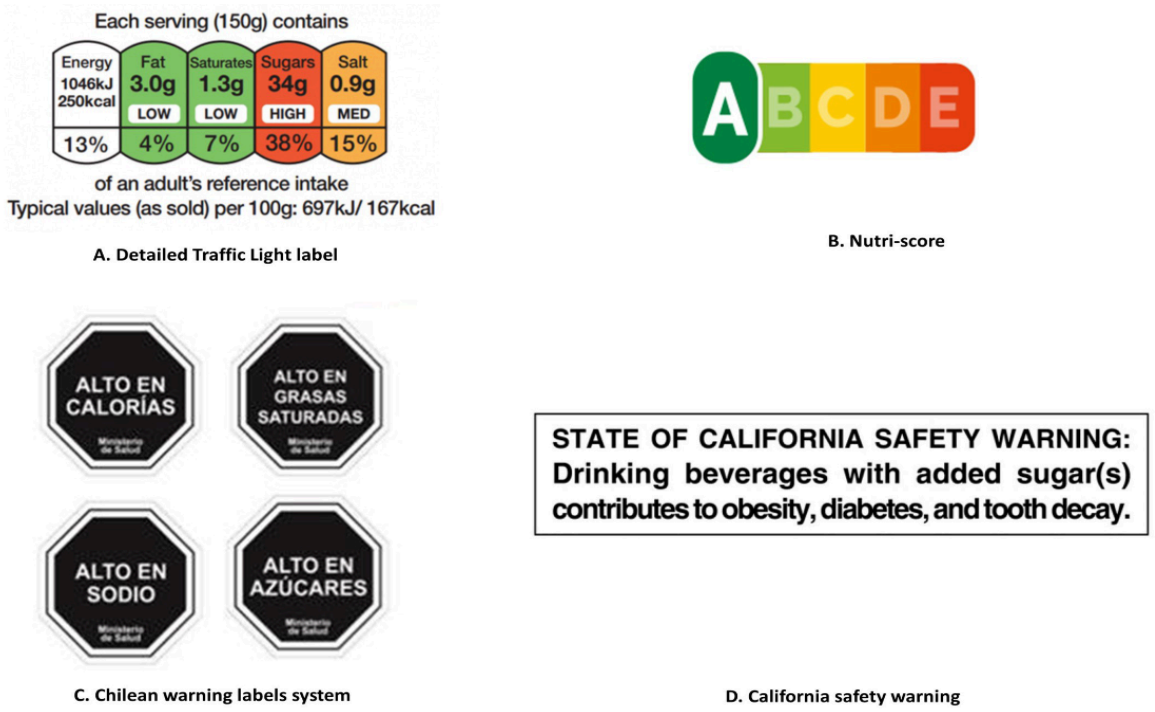


Figure 2: Figure taken from Song et al. (2021) indicating different types of front-of-pack labelling (FOPL)

Information sources and article selection

The search strategy was designed to identify syntheses of research evidence such as systematic reviews between the year 2010 and the date of search. Initial keywords were identified via a scoping review of relevant papers and reports as well as via MEDLINE using the MeSH function. A search was performed in MEDLINE and the Cochrane Database of Systematic Reviews. We searched grey literature using Google Scholar and Google to identify relevant reports. The search was run in March 2023.

Screening

Due to the rapid nature of the reviews, a single reviewer screened titles and abstracts and discussed any uncertainty with a second reviewer. For relevant titles/abstracts, the full text was retrieved for full text review. One reviewer reviewed the full texts and discussed uncertainties with a second reviewer.

Assessment of methodological quality

All relevant reviews were critically appraised by two reviewers individually using the JBI Critical Appraisal Checklist for Systematic Reviews and Research Syntheses. We selected the highest quality and up-to-date review for data extraction. Suitability to our research question was also taken into account when selecting the final review for extraction.

Data extraction

The JBI Data Extraction Form for Review for Systematic Reviews and Research Syntheses was used for data extraction for the final included review. Extracted characteristics included:

- Review characteristics: author/year, objectives, participants (characteristics, total number), setting/context, interventions of interest, date range of included studies, detailed description of the included studies (number/type/country of origin), appraisal instrument and rating, type of review/method of analyses and outcomes.
- Results: findings of the review and comments.

Results

Multiple systematic reviews and meta-analyses have been conducted to assess different types of FOPL. Croker et al. (2020) analysed the effects of FOPLs such as 'high-in' and multiple traffic lights (MTL). Song et al. (2021) assessed the effectiveness of colour-coded and warning nutrition labelling schemes which included traffic lights systems, Nutri-Score, Nutrition Warning and Health Warning. Similarly, Cecchini & Warin (2016) analysed categories such as traffic lights system, guideline daily amount (GDA) and 'other' food labelling schemes (including Nutri-Score, warning labels etc.) This was one of the first systematic reviews to include a meta-analysis of this intervention.

What studies did the review include?

The studies included in the reviews are detailed below:

- Croker et al. (2020) included 14 studies, of which 11 were experimental and three were 'real-world' interrupted times series (ITS) to analyse the effect of national policy. Five of the experimental studies were used in the meta-analysis. The studies were all conducted in UK comparable high-income economies. These studies were carried out in shopping centres, online/using smartphones, in food stores or in a laboratory. The participants were primarily adults, although some had a mix of both adults and children (see table 1 for an overview of findings).

- Song et al. (2021) included 101 experimental and 55 quasi-experimental studies on general population on all colour-coded labels and warning FOPL such as traffic light system (TLS), Nutri-Score, nutrition and health warning. 134 studies were eligible for a meta-analysis. The comparison group was either no-labels or those having a nutrition facts table. 95% of the studies were conducted in laboratory settings with the remaining conducted in out-of-home settings and retail outlets. All the studies in out-of-home settings also used the TLS system on the package.

- Cecchini & Warin (2016) included nine studies that analysed TLS, guideline daily amount (GDA) and other types of food labelling (eg, front-of-pack logos). All the studies were randomised and conducted in high-income economies similar to the UK. The comparison groups were either no-label or other types of FOPL.

What were the systematic review methods?

The systematic reviews used the following methods:

Croker et al. (2020):

- ASSIA (ProQuest), ABI/Inform Global (ProQuest), CINAHL (EBSCO), Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Embase (Ovid), HMIC (Ovid), MEDLINE (OVID), PsycINFO, Sociological Abstracts (ProQuest), Scopus, Trials Register of Promoting Health Interventions (TRoPHI) and Web of Science (Science Citation Index, Social Sciences Citation Index, Emerging Sources Citation Index). The risk of bias was

identified as being low except for the ones examining consumption. Results were reported as per PRISMA.

- STATA/SE, version 15.1 (StataCorp, College Station, TX, USA) was used for the meta-analyses. Purchasing data from experimental studies were meta-analysed. Two series of meta-analyses were undertaken; combined FOPL versus no-FOPL and specific FOPL scheme versus no-FOPL.

Song et al. (2021):

- Searched four databases (PubMed, Embase via Ovid, Cochrane Central Register of Controlled Trials, and Scopus). Overall the risk of bias was assessed as low. Results were reported as per PRISMA.

- The authors used the frequentist NMA method to synthesise studies and make both direct (observed) and indirect (unobserved) comparisons of multiple interventions in between themselves. Random effect models were fitted in the NMA as it was assumed that the heterogeneity in the network model was high. Cochran's Q-statistic and Higgins and Thompson's I^2 were applied to assess the degree of total heterogeneity in the network model, which was further divided into within (conventional between-study heterogeneity) and between-design (overall inconsistency) variations, respectively. To test the transitivity and consistency assumptions underlying NMA, we further calculated the Q-statistic in a full design-by-treatment intervention random-effect inconsistency model.

Cecchini & Warin (2016):

- The search was conducted using the following databases: PubMed, Biomed, ScienceDirect, Sage Database, Google Scholar, EBSCOhost Database Academic. It's not clear if the PRISMA was used to report findings. All studies have been observed to have low risk of bias.

- Data were analysed with Stata 13. Forest plots were generated, and overall estimates of the pooled relation and 95% confidence intervals (CI) were calculated with the use of fixed-effect and random effect models. Heterogeneity across studies was tested with the I^2 statistics. In case of low

heterogeneity, they carried out a graphical assessment of the potential publication bias through a funnel plot.

Further details on the methodology can be accessed in the respective reviews (links provided).

What did the review find?

Croker et al. (2020) reported findings as purchasing behaviour and consumption behaviour. 'High-in' warning labels (ie, those which indicate a food is high in a certain thing such as salt or sugar) showed a statistically significant effect on energy (kcal), sugar(g) and salt(mg) per 100g purchased. MTL was found to have a statistically significant effect only on sodium per 100g purchased. In terms of consumption behaviour, only one study was available and it was not meta-analysed.

Combined FOPL vs No-FOPL has been found to lead to a statistically significant reduction in purchases of sugar by -0.4g/100g (p<0.01) and salt by -24.48mg/100g (p=0.012). There was no statistically significant effect on energy consumption (-2.03kcal/100g) or saturated fat (-0.154g/100g).

However, FOPL ('high-in') was found to lower purchase of calories by 4.43kcal/100g (p<0.05). Similarly, there was a statistically significant reduction in sugar by 0.67g/100g (p≤0.01) and sodium purchases by 33.78mg/100g (p=0.01).

Additionally, FOPL (MTL) was found to only have a statistically significant effect on reducing sodium purchases by 34.94mg/100g (p<0.01). All other outcome measures were not statistically significant: calories: -4.217kcal/100g, sugar: -0.272 (p=0.162), fat: -0.207g/100g (p=0.065).

Song et al. (2021) found that all colour-coded and warning labels were significantly associated with changes in purchasing behaviour. In real-world settings (eight studies), nutrition warning is linked to reduced probability of purchasing/selecting of unhealthy foods (OR = 0.5, 95%) and TLS is linked to increased likelihood of purchasing/selecting healthy foods (OR = 1.32, 95%).

In comparisons between colour-coded vs warning labels, NWS performed better than TLS in discouraging the purchase of unhealthy food (0R = 0.81, 95%) and

lowering total energy purchased (RMD = -0.064, 95%). In addition, TLS in comparison to control was also found to statistically reduce the energy purchased (RMD = -0.065, 95%). NW is better than NS in improving overall healthfulness (RMD = 0.127, 95%); reducing energy purchased (RMD = -0.07, 95%) and reducing fat purchased (RMD = -0.156, 95%). And, NS is better than NW in encouraging purchase of healthier foods (OR = 1.51, 95%).

Cecchini & Warin (2016) found that food labelling increased the share of people purchasing healthier options by 17.95% but with high heterogeneity.

In terms of effect on the share of people buying healthier options, traffic light labelling seems to be the most effective intervention with an increase at 29.36%, followed by other food labels leading to 14.69% increase in share of people purchasing healthier options, although the confidence interval (CI) is wide. This is followed by GDA at 11.85%.

None of the labelling systems were found to have a statistically significant effect on calories purchased. However, single studies report large CIs, suggesting that different individuals respond to the introduction of food labels with a wide range of behaviours, indifference included.

Based on the available reviews, Croker et al. (2020) was chosen to provide the most appropriate evidence for this technical report for the following reasons:

- Cecchini & Warin (2016) was not chosen for this review as the review was published some time ago and was unlikely to include studies conducted recently.
- Croker et al. (2020) used objective measures of purchase and consumption at an individual level, such as quantity purchased, nutritional content such as calories, amount of salt/sugar etc. This is in comparison to more recent reviews like Song et al. (2021) that used self-reported intentions of purchase along with objective purchase data to report the effect of interventions. However, it must be noted that both reviews were able to comment only on purchases and not on consumption due to absence of studies.
- In terms of settings, Croker et al. (2020) exclusively focused on FOPL in prepackaged foods in retail settings unlike Song et al. (2021) which also

included out-of-home settings as well which wasn't aligned with the research question. Moreover, the majority (95%) of studies in the latter were set in laboratories.

- The number of studies to estimate effects in Song et al. (2021) is fewer than in Croker et al. (2020). In addition, Croker et al. (2020) also provides estimated effects in kcal per 100 grams while Song et al. (2021) gave estimates in RMD.
- However, Song et al. (2021) has been included because the review compares different labelling types between themselves.

Table 1: Characteristics of meta-analysis from Croker et al. (2020)

Total number of studies	Country	Total sample size	Settings	Age range	Intervention and comparison	Effect	Quality of evidence (GRADE)[1]
17	High-income economies comparable to UK	7,631	Different settings – retail, lab and online	Adults and children	FOPL vs No label	-1.95kcal/100 g (p>0.05)	Moderate to high
4	High-income economies comparable to UK	4,696	Different settings – retail, lab and online	Adults and children	High-in vs no label	-4.43kcal/100 g (p<0.05)	Moderate to high
5	High-income economies comparable to UK	5,599	Different settings – retail, lab and online	Adults and children	Multiple Traffic Lights	-4.217kcal/100 g (p>0.05)	Very low

[1] *GRADE = Grading of Recommendations, Assessment, Development and Evaluations

Table 2a: Characteristics and results from Song et al. (2021) (Outcome = Change in energy content)

Total number of studies	Country	Total sample size	Settings	Age range	Intervention and comparison	Effect – change in energy content	Quality of evidence (GRADE)[2]
3	High-income economies comparable to UK	n/a	Different settings – labs and some real-world including OOH	Adults and children	Nutri-Score vs control	-0.06 (-0.107, -0.012)	Not available in the review
1	High-income economies comparable to UK	n/a	Different settings – labs and some real-world including OOH	Adults and children	Nutri-Score vs nutrient warning	0.07 (0.008, 0.131)	Not available in the review

[2] *GRADE = Grading of Recommendations, Assessment, Development and Evaluations

3	High-income economies comparable to UK	n/a	Different settings – labs and some real-world including OOH	Adults and children	Nutrient warning vs control	-0.129 (-0.179, -0.08)	Not available in the review
1	High-income economies comparable to UK	n/a	Different settings – labs and some real-world including OOH	Adults and children	Nutrient warning vs TLS	-0.064 (-0.125, -0.004)	Not available in the review
4	High-income economies comparable to UK	n/a	Different settings – labs and some real-world including OOH	Adults and children	TLS vs control	-0.065 (-0.111, -0.019)	Not available in the review

Table 2b: Characteristics and results from Song et al. (2021) (Outcome = Energy per 100g/ml)

Total number of studies	Country	Total sample size	Settings	Age range	Intervention and comparison	Effect – change in energy per 100g/ml	Quality of evidence (GRADE)[3]
5	High-income economies comparable to UK	n/a	Different settings – labs and some real-world including OOH	Adults and children	Nutri-Score vs control	-0.035 (-0.054, -0.016)	Not available in the review
2	High-income economies comparable to UK	n/a	Different settings – labs and some real-world including OOH	Adults and children	Nutrient warning vs control	-0.038 (-0.068, -0.009)	Not available in the review
5	High-income economies comparable to UK	n/a	Different settings – labs and some real-world including OOH	Adults and children	TLS vs control	-0.03 (-0.049, -0.01)	Not available in the review

[3] *GRADE = Grading of Recommendations, Assessment, Development and Evaluations