


## QUESTIONS À L'AUTEUR

 ELSEVIER MASSON	<b>Revue : CND</b>  <b>Numéro d'article : 320</b>	<b>Merci de retourner vos réponses par e-mail à:</b>  <b>E-mail : <a href="mailto:corrections.esme@elsevier.thomsondigital.com">corrections.esme@elsevier.thomsondigital.com</a></b>
---	---	--

Cher auteur,

Vous trouverez ci-dessous les éventuelles questions et/ou remarques qui se sont présentées pendant la préparation de votre article. Elles sont également signalées dans l'épreuve par une lettre « Q » suivie d'un numéro. Merci de vérifier soigneusement vos épreuves et de nous retourner vos corrections soit en annotant le PDF ci-joint, soit en les listant séparément.

Pour toute correction ou modification dans les figures, merci de consulter la page <http://www.elsevier.com/artworkinstructions>.

**Articles de numéros spéciaux :** merci d'ajouter (dans la liste et dans le corps du texte) la mention « dans ce numéro » pour toute référence à d'autres articles publiés dans ce numéro spécial.

<b>Références non appelées :</b> Références présentes dans la liste des références mais pas dans le corps de l'article – merci d'appeler chaque référence dans le texte ou de les supprimer de la liste.	
<b>Références manquantes :</b> Les références ci-dessous se trouvaient dans le texte mais ne sont pas présentes dans la liste des références. Merci de compléter la liste ou de les supprimer du texte.	
Emplacement dans l'article	Question / Remarque
<b>Q1</b>	Merci de vérifier que les prénoms et les noms ont été correctement identifiés.
<b>Q2</b>	Merci de vérifier les affiliations.
<b>Q3</b>	Merci d'ajouter les appels des références [15] et [16] dans le corps du texte.
<b>Q4</b>	Merci de mettre la référence [11] à jour en nous transmettant au minimum le DOI.

### Utilisation des fichiers électroniques

Si nous n'avons pas pu exploiter le fichier de votre article et/ou de vos figures, nous avons utilisé la méthode suivante :

Scan de (ou de parties de) votre article     
  Ressaisie de (ou de parties de) votre article     
  Scan des figures

Merci de votre collaboration.

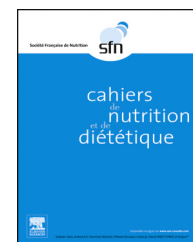


Disponible en ligne sur

**ScienceDirect**  
www.sciencedirect.com

Elsevier Masson France

**EM|consulte**  
www.em-consulte.com



## SANTÉ PUBLIQUE

# Réponses des consommateurs à trois systèmes d'étiquetage nutritionnel face avant

## Consumer responses to three nutrition "Front-of-Pack" labeling systems

Paolo Crosetto<sup>a</sup>, Laurent Muller<sup>a</sup>, Bernard Ruffieux<sup>b,\*</sup>

<sup>a</sup> Inra, UMR 1215 GAEL, 38000 Grenoble, France

<sup>b</sup> Institut polytechnique de Grenoble, 46, avenue Félix-Viallet, 38031 Grenoble cedex 1, France

Reçu le 20 mars 2016 ; accepté le 14 avril 2016

### MOTS CLÉS

Systèmes d'étiquetages nutritionnels ; Label ; Politique nutritionnelle ; Comportement des consommateurs ; France

**Résumé** Dans un magasin expérimental, nous étudions la réponse comportementale en termes d'achats et son impact sur la qualité nutritionnelle des caddies alimentaires des ménages de l'apposition de trois systèmes d'étiquetages nutritionnels différents apposés en face avant des produits. Ces systèmes sont les apports de référence, le Traffic Lights multiples et le « 5C ». On utilise un protocole d'économie expérimentale permettant d'observer pour chaque participant à une cohorte l'écart entre un caddie élaboré sans système d'étiquetage et un caddie élaboré avec l'un des trois systèmes. Nous observons alors, en comparant des cohortes, l'efficacité relative des systèmes. Les trois systèmes améliorent la qualité nutritionnelle, mais avec plus ou moins d'efficacité. Les systèmes d'étiquetage 5C et Traffic Lights ont un impact significativement meilleur que les apports de référence.

© 2016 Société française de nutrition. Publié par Elsevier Masson SAS. Tous droits réservés.

### KEYWORDS

Food rating icons; Label; Nutrition policy;

**Summary** In an experimental store, we study the behavioral responses of consumers food purchasing and their impact on the nutritional quality of shopping carts for three different front-of-pack nutritional labeling systems, i.e. the Reference Intakes, the multiple Traffic Lights and the new French system 5C. We use an experimental economics framework for observing

\* Auteur correspondant.

Adresse e-mail : [Bernard.Ruffieux@grenoble-inp.fr](mailto:Bernard.Ruffieux@grenoble-inp.fr) (B. Ruffieux).

Consumer behavior;  
French

the change for each participant between a cart done without label and a cart done with one of the three systems. By then comparing cohorts, we measure their relative effectiveness. All three systems improve the nutritional quality, but at a lesser or greater extent. 5C and Traffic lights have a significantly greater nutritional impact than the Reference Intakes.

© 2016 Société française de nutrition. Published by Elsevier Masson SAS. All rights reserved.

## Introduction

L'étiquetage nutritionnel en face avant des emballages est considéré sur le plan national (ANSES) et internationale (OMS) comme un instrument digne d'intérêt de politique nutritionnelle pour infléchir les comportements d'achat [1] et dès lors contribuer à prévenir l'obésité et des maladies liées à une mauvaise alimentation. De nombreuses variantes de systèmes d'étiquetages ont été créées à la suite d'initiatives privées ou publiques [2,3]. Face à cette diversité, les pouvoirs publics cherchent à imposer un système unique et normalisé d'étiquetage nutritionnel face avant, comme cela a déjà été fait pour les tables nutritionnelles en face arrière des emballages. Cette volonté d'harmonisation pose la question du choix du système d'étiquetage retenu. Les pouvoirs publics français ont récemment proposé le système d'étiquetage nutritionnel qui utilise un codage à 5 couleurs (5C) [4]. Nous examinons ici l'efficacité relative de ce système par rapport aux Traffic Lights anglo-saxons et aux apports de référence.

Dans une étude antérieure réalisée en 2011 et financée par le ministère de la Santé [5], l'équipe de recherche qui a conduit la présente étude a testé sept systèmes possibles contrastés. Ces systèmes ont été construits à partir de trois critères pertinents de différenciation : (1) la granularité du système – système soit « synthétique » où le produit est jugé globalement, soit « analytique » où le produit est jugé nutriment par nutriment<sup>1</sup> – ; (2) le référentiel d'évaluation – système soit « référentiel du rayon » ou le produit est jugé face aux autres produits du rayon, soit « référentiel d'ensemble » ou le produit est jugé face à toute l'offre alimentaire – ; (3) l'étendue de l'étiquetage – soit système « bons produits » où seuls les bons produits sont étiquetés, soit « intégral » où tous les produits sont étiquetés. Sept systèmes contrastant ces trois critères ont été mis en compétition. Les résultats montrent d'abord que le choix du système importe : les réponses comportementales des acheteurs sont contrastées d'un système à l'autre. Les systèmes utilisant la granularité synthétique sont plus efficace que les analytiques, les systèmes à étendue intégrale sont plus efficaces que les systèmes « bons produits » seuls. Le référentiel d'évaluation modifie les comportements mais conduit à des performances similaires. Enfin, le système apports de référence se situe dans la moyenne des performances.

Dans le présent article nous proposons de mettre en compétition, selon une méthode semblable à l'étude [5] citée ci-dessus, trois des systèmes de logos aujourd'hui privilégiés par les parties prenantes : Les apports de référence

(les anciens repères nutritionnels journaliers) ; les Traffic Lights multiples [6] ; le 5C [7]. Chacun sera détaillé plus bas.

La suite de cet article répond à la question suivante. Peut-on classer les trois systèmes testés selon leurs performances à améliorer la qualité nutritionnelle et quelle est l'ampleur des éventuelles différences de performance ? Nous présentons les trois logos retenus, puis nous décrivons le protocole de l'expérience et les caractéristiques des participants, enfin nous présentons les résultats et leur discussion.

## Les systèmes d'étiquetage testés

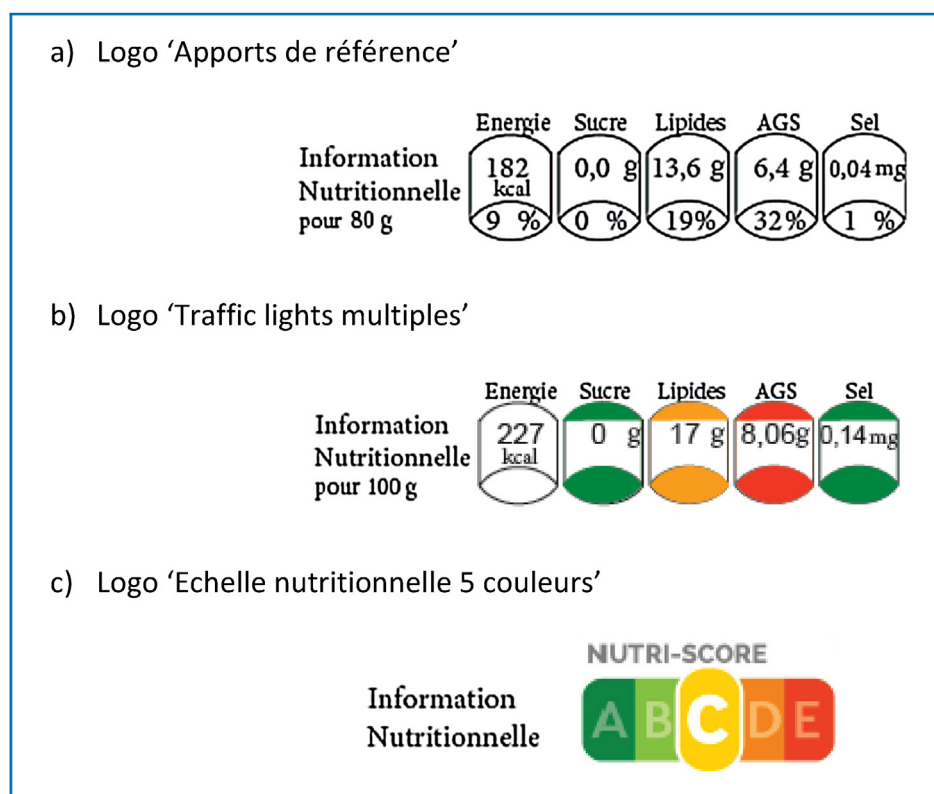
### Apports de référence (AR)

Proposé d'abord par le Royaume-Uni en 1996, le système apports de référence (AR) a d'abord été appelé *guideline daily amount* puis *reference intake* depuis 2013. Il a été adopté par l'Australie en 2006 (*daily intake guide*), par l'Union européenne en 2009 (standard industriel) et enfin par les États-Unis en 2012 (*Facts up fronts*). Les AR, très répandus, ont été adaptés par de grandes entreprises telles que Kraft, Kellogg's ou Danone. Ce système, fondé sur des faits (*fact-based* [8]), reprend de manière concise les informations quantitatives des tables nutritionnelles de face arrière des emballages. Les AR décrivent, sous un format numérique et monochrome, le contenu par portion de produit en énergie, sucres, lipides, acides gras saturés et sel, exprimé en pourcentage des recommandations journalières. La Fig. 1a présente un exemple (entrecôte de bœuf) du système AR utilisé dans notre expérience.

### Traffic Lights multiples (TL)

Les Traffic Lights multiples (TL) ont été développés en 2009 par l'agence britannique de la sécurité alimentaire (FSA) afin de faciliter des choix de produits sains. Le TL attribue quatre pastilles colorées : une aux sucres, une aux lipides, une aux acides gras saturés et une au sel, en fonction de la proportion par 100g. Chaque pastille peut être rouge, jaune ou verte selon que le niveau du critère est élevé, intermédiaire ou faible. Les seuils sont arbitrairement prédéfinis (*criteria-based* [8]). Notons que les TL peuvent parfois être couplés aux apports de référence. Soutenu au Royaume-Uni et en Allemagne par des organismes de santé publique comme la British Medical Association et par des associations de consommateurs, le TL génère de vifs débats. Les industriels de l'alimentation y sont souvent opposés, craignant l'ostracisme de certains produits. La Commission européenne a ouvert en 2014 une procédure en contentieux contre le Royaume-Uni sur la base d'une

<sup>1</sup> Dans l'étude, les nutriments concernés étaient les sucres libres, les acides gras saturés et le sel.



**Figure 1.** Reproduction des logos (exemple pour une entrecôte de bœuf.). a: logo «apports de référence»; b: logo «Traffic lights multiples»; c: logo «Échelle nutritionnelle 5 couleurs».

possible incompatibilité du TL avec le marché intérieur de l'Union européen (Fig. 1b).

### Échelle nutritionnelle 5 couleurs (5C)

Échelle nutritionnelle 5 couleurs (ici 5C) est un système synthétique basé sur une échelle de qualité nutritionnelle représentée par 5 couleurs couplées à 5 lettres. Ce système se distingue des TL par un score nutritionnel unique. Il est basé sur le modèle de profilage nutritionnel UK Ofcom [9] qui évalue la qualité globale d'un aliment. Le calcul est basé sur la teneur au 100g de 4 éléments négatifs (énergie, sucres, acides gras saturés et sel) et d'éléments positifs (fruits, légumes, noix, fibres et protéines) (Fig. 1c).

### Description de l'expérience

La méthodologie générale retenue pour la présente étude reprend celle de l'étude [5] réalisée en 2010–2011 pour le ministère de la Santé. Comme celle-ci, elle a été réalisée sur la plateforme expérimentale Inra à Grenoble INP–Génie industriel. Les données recueillies sont issues d'une méthodologie spécifique de l'économie expérimentale qui s'appelle Framed field experiment [10] appliqué au cas des achats alimentaires tels que décrite dans Muller et al. [11]. Elle se définit de la façon suivante : les membres d'un échantillon de participants (humains) sont aléatoirement affectés à plusieurs cohortes. Les comportements de chaque participant sont individuellement observés dans un univers expérimental contrôlé, reproductible et contextualisé (ici, suivant [11], un mini magasin alimentaire). Les comportements observés sont réels, les décisions prises ont des conséquences effectives (ici, suivant [11], des achats de

produits), en contraste avec les données déclaratives des méthodes d'enquêtes. L'observation en laboratoire permet le contrôle des variables explicatives sans bruit (*ceteris paribus*). Elle permet également la reproductibilité de l'étude et donc des tests de robustesse. Dans ce contexte, l'expérimentateur introduit un choc exogène artificiel qui permet d'observer les «réponses comportementales» des participants. Chaque cohorte est exposée à un choc différent (ici les systèmes d'étiquetage, [11] étudiaient des chocs de prix suite à des taxes ou subventions nutritionnelles).

Dès lors on est à même de mesurer l'efficacité relative des trois systèmes d'étiquetage retenus – TL (Fig. 1a), AR, (Fig. 1b) et 5C (Fig. 1c) – sur les comportements d'achats alimentaires et sur la qualité nutritionnelle des paniers. Dans la présente étude, suivant [5], les participants à l'expérience composent d'abord un premier caddie alimentaire sans la présence d'un système, puis composent un second caddie alternatif dans un contexte identique au premier, à ceci prêt que les produits sont maintenant présentés avec l'apposition (systématique) d'un des trois systèmes d'étiquetage nutritionnel testés. Les comportements d'achat observés sont effectifs : chaque participant achète une partie des produits qu'il a choisis, aux prix du marché tels que relevés au moment de l'expérience. L'illustration de la Fig. 2 synthétise le protocole expérimental.

Magasin expérimental – le magasin expérimental est construit sur le format «e-shopping». Au début de l'expérience chaque participant reçoit un catalogue papier en couleur présentant rayon par rayon mes produits disponibles à la vente. Ce premier catalogue ne contient pas de logo nutritionnel face avant. Il contient 290 produits répartis en 37 rayons ou catégories (un par page ou double page). Pour chaque produit, le participant dispose de la

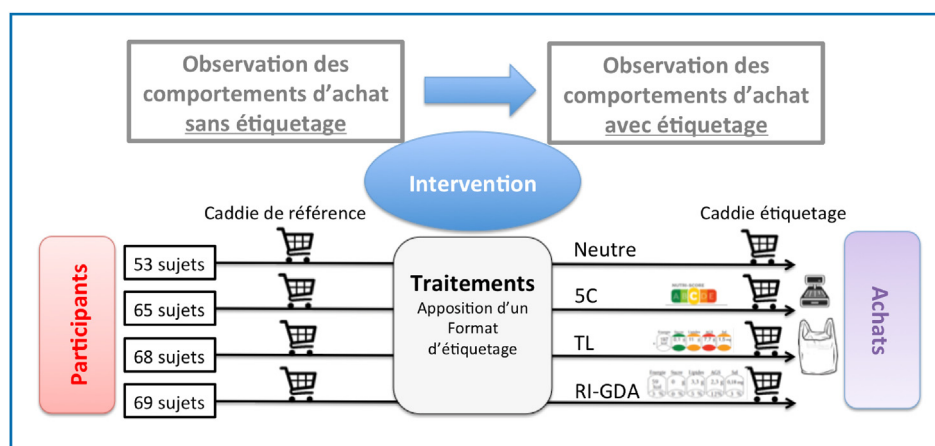


Figure 2. Protocole expérimental.

189 photographie couleur de sa face avant, de son nom, de  
190 sa marque, de son prix observé en grande surface à proxi-  
191 mité au moment de l'étude, et de son poids. Parallèlement,  
192 chaque participant dispose d'un écran numérique, d'un cla-  
193 vier et d'une douchette lectrice de code-barres. À l'aide  
194 de la douchette, chaque participant peut scanner un pro-  
195 duit du catalogue pour le faire apparaître sur son écran.  
196 À partir de là, le participant peut, d'un clic, connaître  
197 la composition en ingrédients du produit ou sa compo-  
198 sition nutritionnelle, reprenant les données disponibles en  
199 face arrière des emballages des produits offerts. Par cette  
200 interface informatique, le participant ajoute (ou retire) des  
201 produits dans son caddie. L'ensemble des produits choisis  
202 apparaît à l'écran. Une fois le premier caddie élaboré, la  
203 première phase de l'expérience est terminée.

204 Une nouvelle version du même catalogue de produits est  
205 alors distribuée à chaque participant. Ce second catalogue  
206 est identique au précédent à ceci près qu'il incorpore main-  
207 tenant, selon le traitement de l'expérience, soit l'un des  
208 trois systèmes testé – TL, AR et 5C – soit aucun système  
209 pour le quatrième traitement « de contrôle ». Pour ce trai-  
210 tement de contrôle, les participants conservent le catalogue  
211 utilisé pour le premier caddie. Les annexes A et B donnent  
212 des exemples de pages des catalogues papiers et des saisies  
213 d'écran de l'interface informatique.

214 Tâches – chaque participant compose successivement  
215 deux caddies, chacun correspondant aux achats alimen-  
216 taires de son foyer pour deux jours. Ces deux caddies sont  
217 des substituts, les achats effectifs en fin d'expérience ne  
218 portent que sur l'un des deux, tiré au sort. L'architecture  
219 d'expérience est double. Les sujets composent un pre-  
220 mier caddie hors étiquetage dans des conditions strictement  
221 identiques dans tous les traitements. Dans un second temps,  
222 les sujets d'un traitement donné composent un caddie  
223 alternatif en présence soit de l'un des trois systèmes  
224 d'étiquetage testés (AR, TL ou 5C) soit en absence de  
225 système pour le traitement neutre de contrôle. Chaque  
226 système d'étiquetage utilisé est décrit aux sujets avant  
227 le démarrage de cette seconde phase. Les participants ne  
228 connaissaient pas l'existence de cette seconde phase quand  
229 ils réalisent la première. Avant le démarrage de la seconde  
230 phase ils sont informés du fait que, pour les achats effectifs,  
231 l'un des deux caddies sera tiré au sort.

232 Sujets et traitements – l'expérience a fait appel à  
233 255 participants (voir annexe C pour une description de  
234 l'échantillon). Les 255 participants ont été répartis en  
235 quatre groupes, un par traitement.

236 Incitations – chaque sujet est indemnisé d'un montant  
237 fixe de 30€ pour sa participation. Ce montant, annoncé  
238 au moment du recrutement, indemnise le déplacement  
239 et le temps passé. Conformément aux standards de la  
240 communauté scientifique, la tâche doit être compatible  
241 en incitations [12] afin de s'assurer de l'absence de biais  
242 hypothétiques, biais qui risque d'entacher les études décla-  
243 ratives [13]. Ainsi, les participants achètent réellement un  
244 sous-ensemble des produits de l'un ou de l'autre des caddies  
245 composés durant l'expérience. Les deux caddies sont substi-  
246 tuables. L'un des deux est tiré au sort en fin d'expérience  
247 et sert de base pour le tirage au sort des achats réels. Un  
248 quart des produits du magasin expérimental est physiqu-  
249 ment disponible sur la plateforme expérimentale au moment  
250 des sessions. Ce sous-ensemble est bien sûr inconnu des par-  
251 ticipants. Pour un sujet, tous les produits disponibles à la  
252 vente qui sont présents dans son caddie tiré au sort doivent  
253 impérativement être achetés et payés par lui à la fin de  
254 l'expérience. Il repart avec les produits. Cette obligation  
255 d'achat est connue des participants qui sont ainsi incités  
256 à prendre au sérieux chaque décision prise. On dit que les  
257 décisions sont « conséquentielles ».

258 Analyse et calcul des impacts – l'architecture de  
259 l'expérience permet l'observation individuelle, pour chaque  
260 sujet, du changement de comportement induit par  
261 l'introduction d'un système d'étiquetage (méthode dite  
262 *within*). Nous mesurons la distance entre la qualité nutri-  
263 tionnelle avant (caddie sans système d'étiquetage) et la  
264 qualité nutritionnelle après l'apposition d'un des systèmes  
265 d'étiquetage (caddie avec étiquetage). Le traitement de  
266 contrôle permet de mesurer un éventuel impact artéfac-  
267 tuel lié à la répétition de la tâche. Cette mesure d'impact  
268 du changement individuel traitement par traitement est  
269 effectuée dans un contexte strictement commun. Nous cal-  
270 culons la variation individuelle de qualité nutritionnelle du  
271 premier caddie au second caddie. Les moyennes par trai-  
272 tement de ces variations individuelles sont ensuite comparées  
273 pour obtenir les changements relatifs induits par chacun des  
274 systèmes (méthode *between*). Au total, la double mesure  
275 de changement *within subject* et de changement *between*  
276 *groups of subjects* permet de répondre à la question posée.

277 Nous avons retenu un indicateur principal de perfor-  
278 mance nutritionnelle des systèmes : le score issu du modèle  
279 de profilage nutritionnel (Nutrient Profiling Model), déve-  
280 loppé par l'agence de sécurité alimentaire britannique  
281 (Food Standard Agency [FSA]). Ce score est rapporté à  
282 2000 kCal [9]. Rappelons que, dans l'étude précédente, nous

**Tableau 1** Variation en pourcentage entre le premier caddie (« référence ») et le deuxième caddie (« logo ») de la qualité nutritionnelle globale du caddie (indicateur LIM et score FSA), de l'énergie, du contenu nutritionnel (AGS, sel, sucres, protéines et fibres) pour 100 g du caddie pour les 5 traitements (neutre, AR, TL, 5C et 5CPart).

Variation en % entre les 2 caddies par individu	5C	TL	AR	Neutre
Score FSA par 100 g	−10,6 <sup>a</sup> (12,9)	−6,6 <sup>a,b</sup> (10,6)	−3,7 <sup>b</sup> (11,8)	+2,8 <sup>c</sup> (13,9)
LIM par 100 g	−14,5 <sup>a</sup> (16,3)	−13,7 <sup>a</sup> (15,5)	−7,0 <sup>b</sup> (15,9)	+5,8 <sup>c</sup> (24,3)
Energie (kCal par 100 g)	−5,2 <sup>a,b</sup> (14,1)	−8,9 <sup>a</sup> (14,2)	−1,1 <sup>b</sup> (17,2)	+4,2 <sup>c</sup> (20,0)
AGS (g par 100 g)	−23,4 <sup>a</sup> (36,4)	−23,2 <sup>a</sup> (38,9)	−12,8 <sup>b</sup> (35,7)	+12,2 <sup>c</sup> (55,9)
Sel (mg par 100 g)	−5,4 <sup>a</sup> (32,0)	−2,0 <sup>a</sup> (19,9)	−1,3 <sup>a</sup> (23,3)	+4,7 <sup>a</sup> (32,2)
Sucres (g par 100 g)	−5,9 <sup>a,b</sup> (19,9)	−8,1 <sup>a</sup> (22,5)	−1,7 <sup>b</sup> (14,5)	+11,2 <sup>c</sup> (35,9)
Protéines (g par 100 g)	+6,4 <sup>a</sup> (37,6)	+7,0 <sup>a</sup> (42,2)	+3,5 <sup>a</sup> (20,6)	+11,8 <sup>a</sup> (49,6)
Fibres (g par 100 g)	+16,6 <sup>a</sup> (40,4)	+0,0 <sup>b</sup> (17,1)	+6,4 <sup>a</sup> (22,7)	−3,0 <sup>b</sup> (23,0)
Prix (€ par 100 g)	−1,5 <sup>a</sup> (14,3)	−1,5 <sup>a</sup> (11,9)	+3,7 <sup>b</sup> (15,9)	+3,8 <sup>a,b</sup> (24,3)

Les valeurs sont des moyennes par individu (déviations standard). Si il existe une différence significative (à un niveau de 1% selon le test signé des rangs de Wilcoxon) entre le premier caddie de référence et le second la variation moyenne est en gras. Si deux moyennes ont une lettre commune à leur exposant (a, b ou c), alors ces deux moyennes ne sont pas significativement différentes à un degré de 1% selon le test de Mann–Whitney.

utilisons un indicateur LIM, issu du SAIN-LIM [14]. Cet indicateur est repris ici afin de permettre la comparaison avec les résultats de l'étude 2011.

Plus l'indicateur FSA est petit (il en va de même du LIM), plus la qualité nutritionnelle du panier de consommation est grande. Ainsi, une baisse des scores traduit une amélioration nutritionnelle. En complément de ces deux indicateurs synthétiques, nous utilisons des indicateurs analytiques de qualité nutritionnelle des caddies : apport calorique, densité en sucres, en sel, en matières grasses, en fibres et en protéines. Enfin, nous examinons l'impact des systèmes sur le prix des caddies.

Les scores FSA de notre échantillon (pas plus que les autres critères) ne suivant une loi normale, des tests non-paramétriques ont été utilisés pour tester la significativité statistique des résultats. Un test de rangs de Wilcoxon (TSRW) est utilisé pour tester les variations du premier au second caddie. Un test de Wilcoxon–Mann–Whitney (TWMW) est utilisé pour comparer les variations relatives entre les traitements. Enfin, un test exact de Fisher (FE) est utilisé pour comparer les proportions d'individus.

## Résultats

Tous traitements confondus, les participants ont sélectionné en moyenne 15,3 items dans leur premier caddie de référence et 15,1 dans le second. La valeur moyenne des caddies s'élève respectivement à 26,8€ et à 26,6€. Entre les deux caddies des traitements AR, TL et 5C, en moyenne 4,1 produits ont été retirés du premier caddie et 4,0 produits ont été rajoutés dans le second. Dans le traitement de contrôle, le nombre de changements est supérieur (6,3 sorties et 6,5 entrées), mais la qualité nutritionnelle demeure inchangée d'un caddie à l'autre. Pour aucun indicateur nutritionnel du Tableau 1, il n'existe de différence significative entre les deux caddies pour le traitement de contrôle (TSRW : la valeur *p* n'excède pas 0,45).

Les trois systèmes nutritionnels testés améliorent la qualité nutritionnelle des caddies. Le score FSA s'améliore en moyenne de 6,9% pour les trois systèmes, alors qu'il se dégrade en moyenne de 2,8% pour le traitement de contrôle. Les systèmes 5C et TL améliorent de façon significative le score FSA : une amélioration de 10,6% pour

5C (TSRW, *p* < 0,01) et de 6,6% pour TL (TSRW, *p* = 0,02). L'amélioration observée de 3,7% du score FSA pour le système AR n'est pas significative (TSRW, *p* = 0,27). En revanche, l'ampleur de l'amélioration obtenue pour ce système AR est significativement supérieure à celle du traitement neutre (TWMW, *p* < 0,01). Ce premier résultat est confirmé avec le score LIM : la baisse du LIM d'un caddie à l'autre pour les systèmes 5C (−14,5%), TL (−13,7%) et AR (−7,0%) est à chaque fois significative (TSRW, *p* < 0,01 pour 5C et TL et *p* = 0,08 pour AR) avec une ampleur d'amélioration supérieure à celle du traitement de contrôle (TWMW, *p* < 0,01 pour les 3 logos).

Les systèmes 5C et TL sont significativement plus efficaces que le système AR. Selon le score FSA, le système 5C ressort comme le plus efficace en moyenne (−10,6%) suivi par TL (−6,6%) et AR (−3,7%). Mais si l'écart est statistiquement significatif entre 5C et AR (TWMW, *p* = 0,01), les écarts ne sont pas significatifs, ni entre 5C et TL (TWMW, *p* = 0,21) ni entre TL et AR (TWMW, *p* = 0,18). Selon l'index LIM, les systèmes 5C et TL devaient significativement les apports de références avec des baisses respectives de 14,5% et 13,7% contre 7,0% (TWMW, *p* = 0,04 pour 5C vs AR et *p* = 0,03 pour TL vs AR). Comme avec le score FSA, les systèmes 5C et TL ne se départagent pas statistiquement (TWMW, *p* = 0,93).

Si on retient les seuls 50% des acheteurs pour lesquels l'amélioration de la qualité nutritionnelle de leur caddie est la plus importante dans chaque système, le 5C a des effets bénéfiques de plus grande ampleur que le TL (TWMW, *p* = 0,01). Pour les autres 50% l'écart n'est pas significative (TWMW, *p* = 0,36) (Fig. 3).

Les trois systèmes testés réduisent l'apport calorique par 100 g des caddies ainsi que les densités en acides gras saturés et sucre. À l'inverse, les densités en sel et en protéines ne sont significativement impactées par l'apposition d'aucun des trois systèmes. Concernant les fibres, les systèmes AR et 5C améliorent la performance nutritionnelle face traitement de contrôle. En comparant les systèmes deux à deux, on voit que TL fait mieux que AR pour l'énergie (−8,9% vs −1,1%, TWMW : *p* = 0,03) et pour les sucres (−8,1% vs −1,7%, TWMW : *p* = 0,01). On voit aussi que TL et 5C font mieux que AR pour les acides gras saturés (−23,2% et −23,4% vs −12,8%, TWMW : *p* = 0,04 et *p* = 0,06). 5C et TL ne se départagent significativement pour aucun nutriment sauf pour les fibres où TL, peu performant, est devancé par

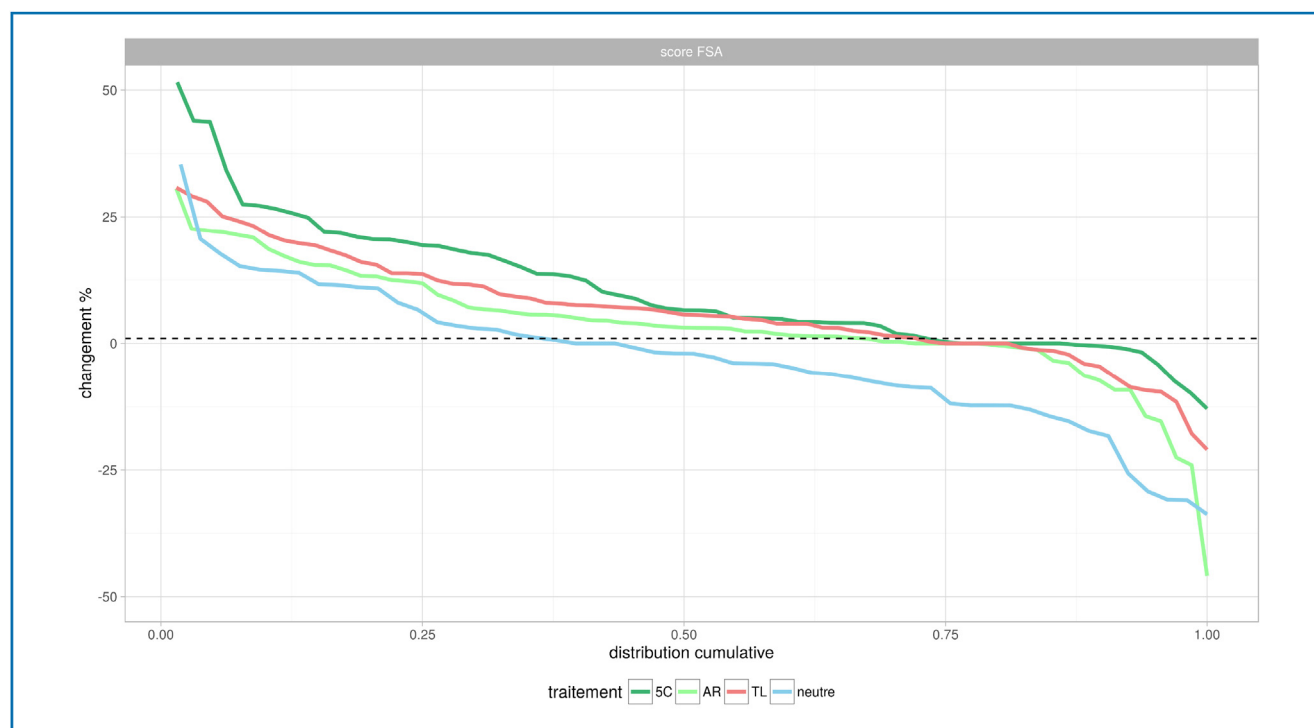


Figure 3. Distribution des sujets selon la variation du score FSA entre les deux caddies.

368 5C (+0,0% vs -16,6%, TWMW :  $p=0,01$ ). Enfin, alors que le  
369 prix par 100 g du contenu du caddie n'évolue pas significati-  
370 vement du premier au second caddie, notons que les prix par  
371 2000 kCal augmentent significativement pour les 3 systèmes  
372 (+5,2% pour 5C, +10,1% pour TL, +10,0% pour AR et +1,1%  
373 pour le traitement neutre ; selon TSRW, les valeurs  $p$  respec-  
374 tives :  $p=0,011$ ,  $p<0,001$ ,  $p=0,003$ ,  $p=0,429$ ).

375 L'apposition des logos, tous systèmes fusionnés, entraîne  
376 une hausse du volume d'achats de fruits (+11% contre -3%  
377 pour le traitement de contrôle), de légumes (+11% vs +8,8%)  
378 et de féculents (+7% vs +2,5%). Elle entraîne aussi une baisse  
379 du volume des plats composés (-7,4% vs -2,1%), des pro-  
380 duits laitier (-3,7% vs -0,7%), des viandes-poissons-œufs  
381 (-3,3% vs +2,5%) et des en-cas salés et sucrés (-14,2% vs  
382 -1,3%).

383 TL se distingue de 5C et AR en entraînant une hausse des  
384 achats de viandes-poissons-œufs (+4,1%) et des produits  
385 laitiers (+0,6%). AR génère une hausse limitée des légumes  
386 (+1,7%) comparée à TL (+16,3%) et 5C (+15,1%) et une  
387 baisse limitée de plats composés (-1,9%, contre -10,4%  
388 avec TL et -10,6% avec 5C).

389 La grande majorité des changements induits par les trois  
390 systèmes se fait au sein des produits du même rayon (intra-  
391 catégorie), même si des changements inter-catégories sont  
392 induits pas chaque système. Les changements extra-famille  
393 sont de 27,4% pour TL et 5C, de 13,6% avec AR et de 6,4%  
394 dans le traitement de contrôle.

## 395 Discussion

396 Cette étude est basée sur un protocole expérimental  
397 éprouvé, identique à celui de [5] qui comparait l'impact de  
398 différents systèmes d'étiquetages nutritionnels sur la qua-  
399 lité nutritionnelle de paniers alimentaire de ménages. La  
400 présente étude compare les systèmes Traffic Lights multiples  
401 (TL), les apports de référence (AR), et l'échelle nutrition-  
402 nelle 5 couleurs (5C) récemment proposée en France. La  
403 qualité méthodologie est renforcé par rapport à [5] par

404 l'introduction d'un traitement de contrôle sans logo, qui  
405 traque un éventuel effet artéfactuel du protocole, un chang-  
406 ement induit par la répétition de la tâche. Cet artéfactuel  
407 est faible et maintenant sous contrôle. En faisant les courses  
408 alimentaires avant et après l'apposition des logos nutri-  
409 tionnels au cours d'une même session expérimentale, un  
410 participant peut modifier son caddie par simple envie de  
411 changement plutôt que par réaction au système de logo. Les  
412 résultats du traitement de contrôle soutiennent cette idée :  
413 ce traitement génère plus de changements d'un caddie à  
414 l'autre que les traitements avec système.

415 Sur le plan de l'interprétation des résultats, ce trai-  
416 tement de contrôle nous conduit à réviser les résultats  
417 présentés dans [5] sous le terme « d'effets pervers ». La cau-  
418 salité de la dégradation de la qualité nutritionnelle d'un  
419 nombre important de caddie est remise en cause. En compa-  
420 rant les changements de comportements induits par les  
421 systèmes avec ceux du traitement de contrôle, nous obser-  
422 vons qu'une part de ce que nous qualifions alors d'effet  
423 pervers est en fait artéfactuel. Seuls les écarts relatifs  
424 d'effets pervers d'un logo à l'autre peuvent être retenus  
425 comme mesurables et non, comme nous l'avons fait dans  
426 [5], la part des consommateurs exposés à de tels effets.

427 Le traitement de référence vient conforter en revanche  
428 nos résultats sur les moyennes. Du premier au second cad-  
429 die, le traitement de contrôle produit un résultat neutre :  
430 la qualité nutritionnelle moyenne ne s'améliore ni ne se  
431 dégrade du premier au second caddie de ce traitement.  
432 Ainsi, les effets positifs moyens sur la qualité nutrition-  
433 nelle de tous les systèmes d'étiquetages mesurée par notre  
434 méthodologie ne sont pas des artéfacts, ce qui n'était pas  
435 un résultat acquis dans [5].

## 436 Conclusion

437 L'enjeu principal de cette étude était de mettre en compéti-  
438 tion trois systèmes d'étiquetage nutritionnel. TL se montre

clairement plus efficace que AR. Nos résultats confirment donc les conclusions des travaux [13] et [14]. van Herpen et van Trijp [17] suggèrent que le système TL attire plus l'attention et induit dès lors plus de changements vers les produits sains face au système AR, monochrome et sophistiqué. AR resterait un outil efficace pour les acheteurs qui prennent le temps d'assimiler les informations [18].

Le système 5C fait aussi bien et peut-être mieux que TL. Une expérience avec deux fois plus de sujets aurait vraisemblablement conduit à une différence significative entre les deux. Les impacts nutritionnels moyens sont d'un ordre de grandeurs comparable, mais 5C permet une amélioration d'une plus grande ampleur pour une proportion d'individus élevée (dans les résultats présentés ici : la moitié). L'analyse fine par nutriment maintient le classement des systèmes par les indicateurs globaux : 5C et TL sont plus efficaces qu'AR pour diminuer le contenu des caddies en acides gras saturés et en sucres. S'écartent du classement principal les fibres où TL fait moins bien que 5C et même que AR. Enfin, tous les logos se révèlent décevants pour améliorer le sel et les protéines. Ce constat sur le sel était déjà fait dans notre étude [5], en cohérence avec les résultats de [19] et [20]. Notons pour finir que le prix par 2000 kCal augmente avec l'apposition des systèmes étudiés, les consommateurs optant pour des produits à prix aux 100 grammes équivalents, mais de densité calorique moindre.

Dans quelle mesure les résultats de cette étude peuvent-ils être transférés in vivo en grandeur nature ? C'est la question de la « validité externe » de nos résultats qui se pose.

En renfort de cette validité externe, nous avançons un argument très fort : les décisions prises par les participants sont des décisions réelles puisqu'elles sont consécutives : ils achètent effectivement une partie des produits choisis et ils le savent. On échappe ainsi au biais déclaratif qui dévalorise des enquêtes.

Les arguments qui limitent cette validité externe ne doivent pas pour autant être négligés. Parmi ceux-ci le caractère inhabituel dans lequel sont plongés les participants en laboratoire : un contexte expérimental jugé parfois surprenant, un mini magasin à offre limitée, l'effet « blouse blanche » qui résulte du fait de se savoir observé et étudié. La mesure du seul effet de court terme de l'étiquetage est

résultats in vivo en termes de performances relatives des systèmes d'étiquetage, il s'agit d'un biais d'amplification, c'est-à-dire de surestimation des écarts. Les réponses relatives et leurs effets sur la qualité nutritionnelle telles que mesurées en laboratoires constituent des limites maximales de ce qui pourra advenir dans la vie réelle.

## Références non citées

[15,16].

## Remerciements

Cette étude a bénéficié d'un soutien financier de la Direction générale de la santé et du programme ANR OCAD (Offrir et Consommer une Alimentation Durable). Nous remercions notre équipe locale du laboratoire GAEL, notamment Sabine Pique, Marie Cronfalt, Mariane Damois, Anne Lacroix et Jean-Loup Dupuis. Nous remercions également Chantal Julia pour la révision du catalogue de produits et pour la mise en place du système 5C dans un des catalogues.

## Déclaration de liens d'intérêts

Les auteurs déclarent ne pas avoir de liens d'intérêts.

## Annexes A et B. Matériel complémentaire

Le matériel complémentaire accompagnant la version en ligne de cet article est disponible sur <http://dx.doi.org/10.1016/j.cnd.2016.04.002>.

## Annexe C. Description de l'échantillon

Description de l'échantillon – Moyenne (Déviation standard) de l'âge, de la taille du foyer et des revenus par unité de consommation et fréquence en pourcentage des femmes par traitement. Test de différence (Kruskal–Wallis).

	Neutre	AR	TL	5C	Valeur-p
Age	38,4 (13,6)	37,8 (12,4)	36,7 (12,0)	38,4 (12,2)	0,460
Femme	70% (0,5)	59% (0,5)	53% (0,5)	55% (0,5)	0,170
Taille du foyer	2,0 (0,9)	2,1 (1,2)	2,2 (1,1)	2,0 (1,0)	0,090
Revenus par unité de consommation*	1631 € (781)	1356 € (643)	1477 € (977)	1611 € (858)	0,250

\* Une unité de consommation est égale à 1 pour le premier adulte du foyer, égale à 0,5 pour les personnes suivantes de 14 ans ou plus et égale à 0,4 pour les enfants de moins de 14 ans.

sans aucun doute une autre faiblesse importante de notre méthodologie.

Mais la question centrale de validité est la suivante : les écarts de performances d'un système à l'autre sont-ils biaisés par notre protocole ? La réponse est claire. Le protocole agit comme un amplificateur des réponses comportementales, en particulier du fait que, en laboratoire, nous capturons l'attention des participants. La très forte validité interne de notre protocole joue aussi comme amplificateur de résultats : en supprimons les bruits ambiants et en isolant le choc externe de toute autre différence, nous accroissons les effets mesurés. Au total, si différence il y a entre les résultats de laboratoires et les

## Références

- [1] Lusk JL, Roosen J. The Oxford Handbook of the economics of food consumption and policy. OUP Oxford; 2011.
- [2] Drichoutis AC, Nayga Jr RM, Lazaridis P. « Nutritional labeling »; 2011. p. 520–45.
- [3] Kleef EV, Dagevos H. « The growing role of front-of-pack nutrition profile labeling: a consumer perspective on key issues and controverses ». Crit Rev Food Sci Nutr 2015;55(3):291–303.
- [4] L. D. française, « Propositions pour un nouvel élan de la politique nutritionnelle française de santé publique dans le cadre de la Stratégie nationale de santé – 1ère partie : mesures



- 534 concernant la prévention nutritionnelle». [En ligne]. Dispo- 563  
 535 nible sur : [http://www.ladocumentationfrancaise.fr/rapports- 565  
 537 \[5\] Muller L, Ruffieux B. «Modification des achats en réponse 566  
 538 à l'apposition de différents logos d'évaluation nutritionnelle 567  
 539 sur la face avant des emballages». Cah Nutr Diététique 568  
 540 2012;47\(4\):171–82. 569  
 541 \[6\] «Front of Pack nutrition labelling guidance – publications – 570  
 542 GOV.UK». \[En ligne\]. Disponible sur : \[https://www.gov.uk/ 571  
 543 government/publications/front-of-pack-nutrition-labelling- 572  
 544 guidance\]\(https://www.gov.uk/government/publications/front-of-pack-nutrition-labelling-guidance\). 573  
 545 \[7\] S. Hercberg, «Système d'information "Echelle nutritionnelle" 574  
 546 – analyse d'un système d'étiquetage volontaire nutritionnel 575  
 547 complémentaire de la déclaration nutritionnelle obligatoire». 576  
 548 \[En ligne\]. Disponible sur : \[http://alimentation-sante.org/ 577  
 549 wp-content/uploads/2015/01/Fiche.EchelleNutritionnelle\\\_ 578  
 550 1501.pdf\]\(http://alimentation-sante.org/wp-content/uploads/2015/01/Fiche.EchelleNutritionnelle\_1501.pdf\). 579  
 551 \[8\] Pereira R. «Understanding the maze of nutrition sym- 580  
 552 bols and nutrition labeling systems». Diabetes Spectr 581  
 553 2010;23\(1\):57–63. 582  
 554 \[9\] Rayner M, Scarborough T, Lobstein. «The UK Ofcom nutrient 583  
 555 profiling model»; 2009. 584  
 556 \[10\] Harrison GW, List JA. «Field Experiments». J Econ Lit 585  
 557 2004;42\(4\):1009–55. 586  
 558 Q4 \[11\] Muller L, Lacroix A, Lusk JL, Ruffieux B. «Distributional impacts 587  
 559 of fat taxes and thin subsidies». Econ J 2016. 588  
 560 \[12\] Harrison GW. «Making choice studies incentive compatible». 589  
 561 In: Kanninen BJ, editor. Valuing environmental amenities using 590  
 562 stated choice studies. Éd. Springer Netherlands; 2006. p. 591  
 67–110.](http://www.ladocumentationfrancaise.fr/rapports- 564<br/>
    536 publics/144000068/)
- [13] List JA, Gallet CA. «What experimental protocol influence 563  
 564 disparities between actual and hypothetical stated values?». 565  
 Environ Resour Econ 2001;20(3):241–54.
- [14] Darmon N, Vieux F, Maillot M, Volatier J-L, Martin A. «Nutrient 566  
 567 profiles discriminate between foods according to their contri- 568  
 569 bution to nutritionally adequate diets: a validation study using 570  
 571 linear programming and the SAIN, LIM system ». Am J Clin Nutr 572  
 2009;89(4):1227–36.
- [15] Kelly B, Hughes C, Chapman K, Louie JC-Y, Dixon H, Crawford 573  
 574 J, et al. «Consumer testing of the acceptability and effective- 575  
 576 ness of front-of-pack food labelling systems for the Australian 577  
 578 grocery market ». Health Promot Int 2009;24(2):120–9.
- [16] Balcombe K, Fraser I, Falco SD. «Traffic lights and food choice: 579  
 580 a choice experiment examining the relationship between nutri- 581  
 582 tional food labels and price ». Food Policy 2010;35(3):211–20. 583  
 584
- [17] van Herpen E, van Trijp HCM. «Front-of-pack nutrition 585  
 586 labels. Their effect on attention and choices when consu- 587  
 588 mers have varying goals and time constraints ». Appetite 589  
 2011;57(1):148–60.
- [18] Muller L, Ruffieux B. «Do price-tags influence consumers' 590  
 591 willingness to pay? On the external validity of using auctions 592  
 for measuring value ». Exp Econ 2010;14(2):181–202.
- [19] Higginson CS, Rayner MJ, Draper S, Kirk TR. «The nutri- 593  
 594 tion label – which information is looked at? ». Nutr Food Sci 595  
 2002;32(3):92–9.
- [20] Drewnowski A, Moskowitz H, Reisner M, Krieger B. «Testing 596  
 597 consumer perception of nutrient content claims using conjoint 598  
 599 analysis ». Public Health Nutr 2010;13(05):688–94. 600