

QUESTIONS À L'AUTEUR

 ELSEVIER MASSON	Revue : CND Numéro d'article : 320	Merci de retourner vos réponses par e-mail à: E-mail : corrections.esme@elsevier.thomsondigital.com
---	---	--

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.

Références non appelées : 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.	
Références manquantes : 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
Q1 Q2 Q3 Q4	Merci d'insérer votre réponse ou votre correction à la ligne correspondante dans l'épreuve Merci de vérifier que les prénoms et les noms ont été correctement identifiés. Merci de vérifier les affiliations. Merci d'ajouter les appels des références [15] et [16] dans le corps du texte. 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



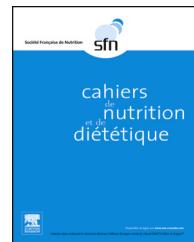
Scan des figures

Merci de votre collaboration.



Disponible en ligne sur
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com



2 SANTÉ PUBLIQUE

3 **Réponses des consommateurs à trois
4 systèmes d'étiquetage nutritionnel face
5 avant**6 *Consumer responses to three nutrition "Front-of-Pack" labeling systems*7Q1 **Paolo Crosetto^a, Laurent Muller^a, Bernard Ruffieux^{b,*}**8 Q2 ^a *Inra, UMR 1215 GAEL, 38000 Grenoble, France*9 ^b *Institut polytechnique de Grenoble, 46, avenue Félix-Viallet, 38031 Grenoble cedex 1,
10 France*11 *Reçu le 20 mars 2016 ; accepté le 14 avril 2016***MOTS CLÉS**

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

KEYWORDS

Food rating icons;
Label;
Nutrition policy;

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.

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

23 * Auteur correspondant.

24 *Adresse e-mail :* Bernard.Ruffieux@grenoble-inp.fr (B. Ruffieux).

27
28
29
30
31Consumer 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.

32

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 nutriment¹ – ; (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. Souvent 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

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

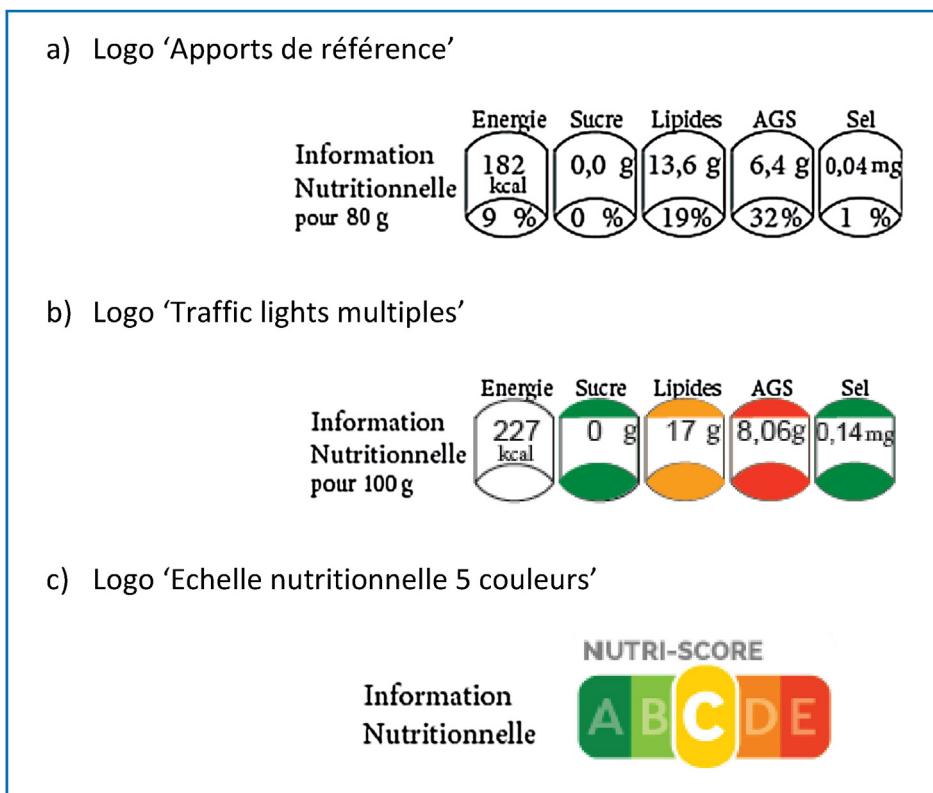


Figure 1. Reproduction des logos (exemple pour une entrecôte de bœuf.). a: logo « apports de référence » ; b: logo « Traffic lights multiples » ; c: logo « Échelle nutritionnelle 5 couleurs ».

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 100 g 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 fied experiment [10] appliquée 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

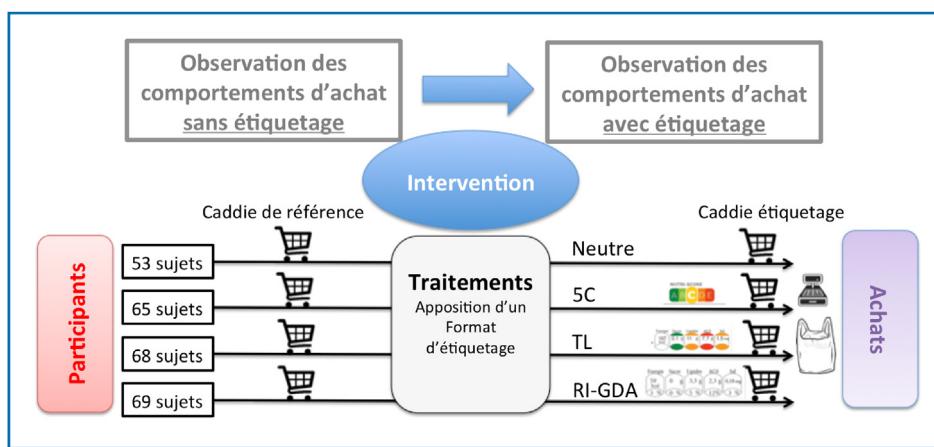


Figure 2. Protocole expérimental.

photographie couleur de sa face avant, de son nom, de sa marque, de son prix observé en grande surface à proximité au moment de l'étude, et de son poids. Parallèlement, chaque participant dispose d'un écran numérique, d'un clavier et d'une douchette lectrice de code-barres. À l'aide de la douchette, chaque participant peut scanner un produit du catalogue pour le faire apparaître sur son écran. À partir de là, le participant peut, d'un clic, connaître la composition en ingrédients du produit ou sa composition nutritionnelle, reprenant les données disponibles en face arrière des emballages des produits offerts. Par cette interface informatique, le participant ajoute (ou retire) des produits dans son caddie. L'ensemble des produits choisis apparaît à l'écran. Une fois le premier caddie élaboré, la première phase de l'expérience est terminée.

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

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

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

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

Analyse et calcul des impacts – l'architecture de l'expérience permet l'observation individuelle, pour chaque sujet, du changement de comportement induit par l'introduction d'un système d'étiquetage (méthode dite *within*). Nous mesurons la distance entre la qualité nutritionnelle avant (caddie sans système d'étiquetage) et la qualité nutritionnelle après l'apposition d'un des systèmes d'étiquetage (caddie avec étiquetage). Le traitement de contrôle permet de mesurer un éventuel impact artéfactuel lié à la répétition de la tâche. Cette mesure d'impact du changement individuel traitement par traitement est effectuée dans un contexte strictement commun. Nous calculons la variation individuelle de qualité nutritionnelle du premier caddie au second caddie. Les moyennes par traitement de ces variations individuelles sont ensuite comparées pour obtenir les changements relatifs induits par chacun des systèmes (méthode *between*). Au total, la double mesure de changement *within subject* et de changement *between groups of subjects* permet de répondre à la question posée.

Nous avons retenu un indicateur principal de performance nutritionnelle des systèmes : le score issu du modèle de profilage nutritionnel (Nutrient Profiling Model), développé par l'agence de sécurité alimentaire britannique (Food Standard Agency [FSA]). Ce score est rapporté à 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^a (12,9)	–6,6^{a,b} (10,6)	–3,7^b (11,8)	+2,8^c (13,9)
LIM par 100 g	–14,5^a (16,3)	–13,7^a (15,5)	–7,0^b (15,9)	+5,8^c (24,3)
Energie (kCal par 100 g)	–5,2^{a,b} (14,1)	–8,9^a (14,2)	–1,1^b (17,2)	+4,2^c (20,0)
AGS (g par 100 g)	–23,4^a (36,4)	–23,2^a (38,9)	–12,8^b (35,7)	+12,2^c (55,9)
Sel (mg par 100 g)	–5,4^a (32,0)	–2,0^a (19,9)	–1,3^a (23,3)	+4,7^a (32,2)
Sucres (g par 100 g)	–5,9^{a,b} (19,9)	–8,1^a (22,5)	–1,7^b (14,5)	+11,2^c (35,9)
Protéines (g par 100 g)	+6,4^a (37,6)	+7,0^a (42,2)	+3,5^a (20,6)	+11,8^a (49,6)
Fibres (g par 100 g)	+16,6^a (40,4)	+0,0^b (17,1)	+6,4^a (22,7)	–3,0^b (23,0)
Prix (€ par 100 g)	–1,5^a (14,3)	–1,5^a (11,9)	+3,7^b (15,9)	+3,8^{a,b} (24,3)

Les valeurs sont des moyennes par individu (déviation 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.

utilisions 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 pas 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 devancent 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 déparent 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 significatif (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 au 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 déparent pas statistiquement pour aucun nutriment sauf pour les fibres où TL, peu performant, est devancé par

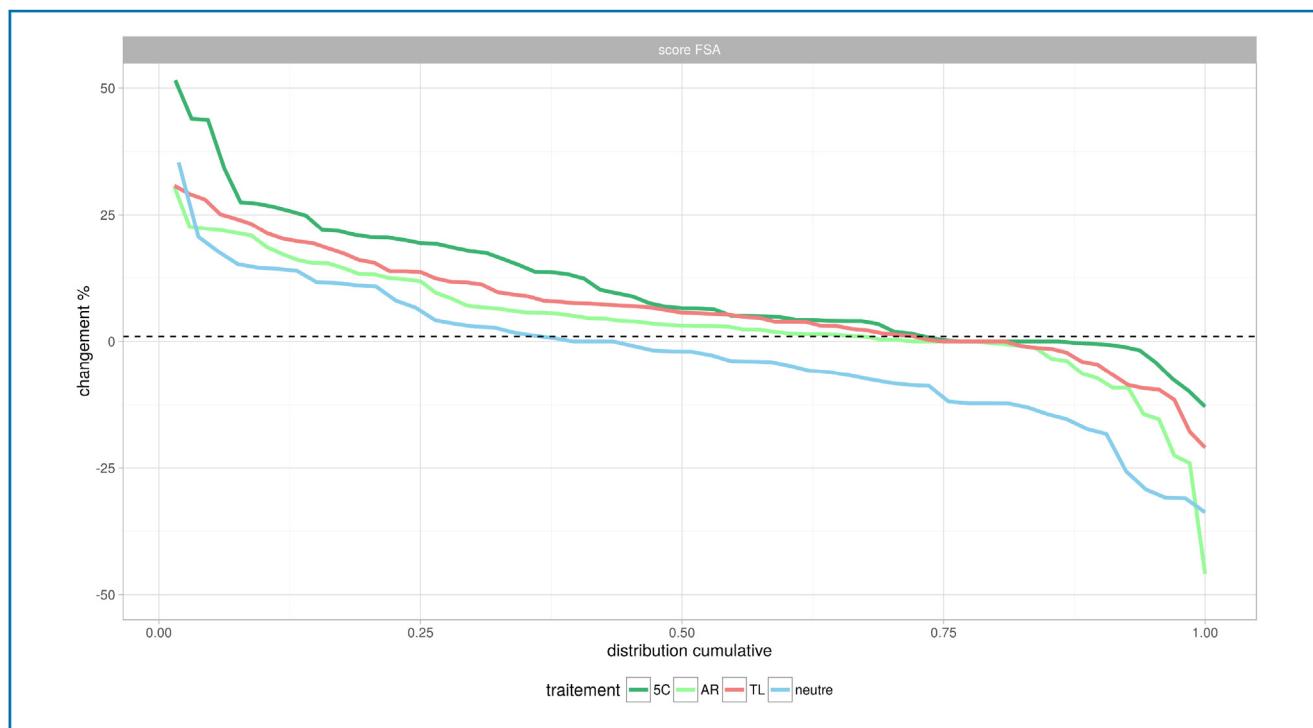


Figure 3. Distribution des sujets selon la variation du score FSA entre les deux caddies.

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

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

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

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

Discussion

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

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

Sur le plan de l'interprétation des résultats, ce traitement de contrôle nous conduit à réviser les résultats présentés dans [5] sous le terme « d'effets pervers ». La causalité de la dégradation de la qualité nutritionnelle d'un nombre important de caddie est remise en cause. En comparant les changements de comportements induits par les systèmes avec ceux du traitement de contrôle, nous observons qu'une part de ce que nous qualifions alors d'effet pervers est en fait artéfactuel. Seuls les écarts relatifs d'effets pervers d'un logo à l'autre peuvent être retenus comme mesurables et non, comme nous l'avons fait dans [5], la part des consommateurs exposés à de tels effets.

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

Conclusion

L'enjeu principal de cette étude était de mettre en compétition trois systèmes d'étiquetage nutritionnel. TL se montre

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

446 Le système 5C fait aussi bien et peut-être mieux que TL. Une
 447 expérience avec deux fois plus de sujets aurait vraisemblablement conduit à une différence significative entre
 448 les deux. Les impacts nutritionnels moyens sont d'un ordre
 449 de grandeurs comparable, mais 5C permet une amélioration
 450 d'une plus grande ampleur pour une proportion d'individus
 451 élevée (dans les résultats présentés ici : la moitié). L'analyse
 452 fine par nutriment maintient le classement des systèmes
 453 par les indicateurs globaux : 5C et TL sont plus efficaces
 454 qu'AR pour diminuer le contenu des caddies en acides gras
 455 saturés et en sucres. S'écartent du classement principal les
 456 fibres où TL fait moins bien que 5C et même que AR. Enfin,
 457 tous les logos se révèlent décevants pour améliorer le sel
 458 et les protéines. Ce constat sur le sel était déjà fait dans
 459 notre étude [5], en cohérence avec les résultats de [19] et
 460 [20]. Notons pour finir que le prix par 2000 kCal augmente
 461 avec l'apposition des systèmes étudiés, les consommateurs
 462 optant pour des produits à prix aux 100 grammes équivalents,
 463 mais de densité calorique moindre.

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

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

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

495 résultats *in vivo* en termes de performances relatives des
 496 systèmes d'étiquetage, il s'agit d'un biais d'amplification,
 497 c'est-à-dire de surestimation des écarts. Les réponses rela-
 498 tives et leurs effets sur la qualité nutritionnelle telles que
 499 mesurées en laboratoires constituent des limites maximales
 500 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
 504 générale de la santé et du programme ANR OCAD (Offrir
 505 et Consommer une Alimentation Durable). Nous remercions
 506 notre équipe locale du laboratoire GAEI, notamment Sabine
 507 Pique, Marie Cronfalt, Mariane Damois, Anne Lacroix et
 508 Jean-Loup Dupuis. Nous remercions également Chantal Julia
 509 pour la révision du catalogue de produits et pour la mise en
 510 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
 515 en ligne de cet article est disponible sur
 516 <http://dx.doi.org/10.1016/j.cnd.2016.04.002>.

Annexe C. Description de l'échantillon

Description de l'échantillon – Moyenne (Déviation standard)
 519 de l'âge, de la taille du foyer et des revenus par unité de
 520 consommation et fréquence en pourcentage des femmes par
 521 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 ambients 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 controversies ». *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-
535 nible sur : <http://www.ladocumentationfrancaise.fr/rapports->
536 [publics/144000068/](http://144000068/).
- 537 [5] Muller L, Ruffieux B. « Modification des achats en réponse
538 à l'apposition de différents logos d'évaluation nutritionnelle
539 sur la face avant des emballages ». Cah Nutr Diététique
540 2012;47(4):171–82.
- 541 [6] « Front of Pack nutrition labelling guidance – publications –
542 GOV.UK ». [En ligne]. Disponible sur : <https://www.gov.uk/government/publications/front-of-pack-nutrition-labelling-guidance>.
- 543 [7] S. Hercberg, « Système d'information "Echelle nutritionnelle" –
544 analyse d'un système d'étiquetage volontaire nutritionnel
545 complémentaire de la déclaration nutritionnelle obligatoire »,
546 [En ligne]. Disponible sur : http://alimentation-sante.org/wp-content/uploads/2015/01/Fiche_EchelleNutritionnelle_1501.pdf.
- 547 [8] Pereira R. « Understanding the maze of nutrition symbols and nutrition labeling systems ». Diabetes Spectr 548 2010;23(1):57–63.
- 549 [9] Rayner M, Scarborough T, Lobstein. « The UK Ofcom nutrient 550 profiling model »; 2009.
- 551 [10] Harrison GW, List JA. « Field Experiments ». J Econ Lit 552 2004;42(4):1009–55.
- 553 Q4 [11] Muller L, Lacroix A, Lusk JL, Ruffieux B. « Distributional impacts 554 of fat taxes and thin subsidies ». Econ J 2016.
- 555 [12] Harrison GW. « Making choice studies incentive compatible ». 556 In: Kanninen BJ, editor. Valuing environmental amenities using
557 stated choice studies. Éd. Springer Netherlands; 2006. p. 558 67–110.
- 559 [13] List JA, Gallet CA. « What experimental protocol influence 560 disparities between actual and hypothetical stated values? ». 561 Environ Resour Econ 2001;20(3):241–54.
- 562 [14] Darmon N, Vieux F, Maillot M, Volatier J-L, Martin A. « Nutrient 563 profiles discriminate between foods according to their contribu- 564 tion to nutritionally adequate diets: a validation study using 565 linear programming and the SAIN, LIM system ». Am J Clin Nutr 566 2009;89(4):1227–36.
- 567 [15] Kelly B, Hughes C, Chapman K, Louie JC-Y, Dixon H, Crawford 568 J, et al. « Consumer testing of the acceptability and effectiveness 569 of front-of-pack food labelling systems for the Australian 570 grocery market ». Health Promot Int 2009;24(2):120–9.
- 571 [16] Balcombe K, Fraser I, Falco SD. « Traffic lights and food choice: 572 a choice experiment examining the relationship between nutri- 573 tional food labels and price ». Food Policy 2010;35(3):211–20.
- 574 [17] van Herpen E, van Trijp HCM. « Front-of-pack nutrition 575 labels. Their effect on attention and choices when consumers 576 have varying goals and time constraints ». Appetite 2011;57(1):148–60.
- 577 [18] Muller L, Ruffieux B. « Do price-tags influence consumers' 578 willingness to pay? On the external validity of using auctions 579 for measuring value ». Exp Econ 2010;14(2):181–202.
- 580 [19] Higginson CS, Rayner MJ, Draper S, Kirk TR. « The nutri- 581 tion label – which information is looked at? ». Nutr Food Sci 582 2002;32(3):92–9.
- 583 [20] Drewnowski A, Moskowitz H, Reisner M, Krieger B. « Testing 584 consumer perception of nutrient content claims using conjoint 585 analysis ». Public Health Nutr 2010;13(05):688–94.
- 586
- 587
- 588
- 589
- 590