Methodological quality of front-of-pack labeling studies: A review and identification of research challenges

Ellis L Vyth
Ingrid HM Steenhuis
Hella E Brandt
Annet JC Roodenburg
Johannes Brug
Jacob C Seidell

Nutrition Reviews 2012, In Press
Abstract

Introduction: Evaluations of the effects of front-of-pack (FOP) labeling vary in their methodologies and rigor. This is the first review that evaluates the methodological quality of current FOP labeling research.

Methods: Peer-reviewed articles were identified using a computerized search of the databases PUBMED and the Web of Science (ISI) from 1990 to February 2011; reference lists from key published articles were additionally used to identify published studies. Study quality of the 31 included studies was assessed. Studies were subdivided in self-reported consumer studies (n=11), observational consumer studies (n=9), sales (n=3), reformulation (n=3), and health outcomes (n=5).

Results: Observational consumer studies had a higher quality than self-reported consumer studies. Sales studies make use of large databases in real life settings, but lack control groups. There is a lack of a validated methodology for measuring FOP label use in real life settings. Regarding reformulation, the challenge is to collect large samples of chemically analyzed food composition data. Modeling studies can provide interesting insights in the potential population health effects of FOP labels, but are based on theoretical assumptions. Only one study used biomarkers to assess the health effects of a FOP label in a real life setting.

Conclusions: Few methodologically sound FOP labeling studies are presently available. The highest methodological quality and public health relevance is achieved by measuring health effects of FOP labels by using biomarkers in a longitudinal randomized controlled design in a real life setting. Future research challenges are discussed.
Introduction
Front-of-pack (FOP) nutrition labels are present on food products worldwide (1-4). Most labels are assigned to products which have reduced levels of saturated fat, sugar, salt and calories. The World Health Organization recommends limiting the intake of these nutrients to reduce the prevalence of diet-related chronic diseases (5). Although the present FOP labels have different designs, different product criteria and different developers, they generally have the same two aims: to assist consumers in making healthier food choices and to stimulate food manufacturers to produce healthier products.

In recent years there has been an international debate about the preferred format and potential impact of front-of-pack nutrition labeling. Regulatory changes are currently being considered by the European Parliament (6, 7) and regulatory bodies in Australia and New Zealand (8, 9). The Institute of Medicine (IOM) and the Food and Drug Administration (FDA) in the United States are also currently conducting research in this area (1, 10). In this highly political debate, policy makers, scientists, industry groups and consumer organizations are looking for evidence-based information on the effects of FOP labels to support their policy (7, 11, 12). As a result, more and more studies testing the effectiveness of the FOP labels are published. Researchers study different aspects, such as (self-reported) consumer understanding, liking and use of FOP labels (13-23), observational label use (24-32), and effects on reformulation (33-35), sales, (36-38) and health outcomes (39-43). In this way, they aim to evaluate the impact of the labels on the health of our society. Until now, there is no consensus about the actual (health) effects of different FOP labels, neither is there consensus about the optimal format to guide consumers’ food purchases.

Thus, many studies have investigated whether or not FOP labels are effective in changing consumers’ food choices and improving public health. However, to the best of our knowledge, no overview has been published regarding the methodological quality of these studies. Studying methodological issues is important because good scientific studies can provide reliable evidence about the effectiveness of FOP labels. Some studies already did identify some methodological limitations, for example that self-reported data do not accurately reflect actual FOP label use (23, 26, 28). Consequently, more studies started to make use of observational data, such as in-store observations (25, 26, 29, 32), eye-tracking (27, 31), or collecting supermarket sales data (36, 37). However, do these data accurately reflect real food purchases resulting from FOP labels, or do other factors bias the results? How do we study the effectiveness of FOP labels in the best way, and which outcome measures are most relevant for public health? In considering these questions, this study aimed to provide a review of the methodological quality of current FOP labeling research. We discuss the strengths and limitations of the current studies and propose future research challenges.

Methods
Structure of this overview
“Effectiveness” was defined as the measure of impact of FOP labels on consumer behavior, reformulation, and health outcomes. Consumer behavior was subdivided in effects on consumers’ self reported understanding and use of FOP labels, effects
on consumers’ observational use and effects on sales. See Figure 1 for a schematic representation of this subdivision, which was based on the designs and main outcomes of the studies.

Figure 1. Structure of this review: schematic representation of the subdivision based on the designs and main outcomes of the studies.

Search strategy
Peer-reviewed articles were located using a computerized search of the databases PUBMED and the Web of Science (ISI) from 1990 to February 2011. We used the following keywords: “front of pack,” “nutrition logo,” “nutrition label,” “nutrition symbol,” “on package nutrition information,” and “health logo.” Also, we used the names of all current existing FOP labels as key words, for example: Traffic Light, Guiding Stars, Canada’s Health Check, AHA Heart Check, NuVal, Green Keyhole, Choices logo, Guideline Daily Amount (GDA), Finnish Heart Symbol, and Pick the Tick (1, 3). In addition, we reviewed the reference lists from key published articles and nutrition reviews for relevant articles.

Inclusion and exclusion criteria
We included studies that evaluated the effectiveness of existing FOP labels actually in use. Studies investigating the nutrition facts panel and other back-of-pack information, health claims, calorie labeling, general on-package nutrition statements (e.g. “low salt” or “healthy food”), the general term “FOP labels” without mentioning the name of any specific FOP label, or self-developed health logos (e.g. a tick with the text ‘healthy choice’ developed by researchers and tested in a specific study design) were excluded. Further, FOP labeling studies that had no clear effectiveness measure on consumer behavior or product development, and studies belonging to larger studies or that did not contain original research were excluded, such as overview articles, review articles, reports of penetration of FOP labels and editorials.
Quality score
The quality of the included studies was assessed by two independent researchers (ELV and HEB) by using the quality assessment tool developed by Sirriyeh et al. (44). This tool is applicable to diverse research designs, including quantitative, qualitative and mixed designs and enables comparison among a diverse range of studies. The tool consists of 16 criteria. Each research paper was awarded a score on a scale from 0 to 3 for each of the criteria. Discussion following the independent scoring of papers resolved any differences in agreement by the two researchers. The sum of the scores provided a quality score per paper, and this score was expressed as a percentage of the maximum score possible (range 0-100%).

Results
The initial search generated a total of 622 citations, of which 122 titles appeared to meet the inclusion criteria and were reviewed. After reading the abstracts and/or full text articles, 31 studies met the inclusion criteria and were included. Table 1 lists the characteristics of these 31 studies. We found 11 self-reported consumer studies, 9 objective consumer studies, 3 sales studies, 3 studies focusing on reformulation, and 5 studies focusing on health outcomes. The last column of Table 1 shows the quality scores (%) per paper. Table 2 lists the main items identified to contribute to a high and a low quality scoring, which are discussed per subheading below.

Self reported consumer studies
The earliest studies evaluating effects of FOP labels, of which we found 11 studies (13-23), use consumer surveys. The outcome measures of these studies are based on self-reported data. Consumer studies using questionnaires generally aim to provide insight toward the understanding and use of FOP labels, and to explore any differences in perception between consumer groups (high/low educated, normal weight/obese, men/women etc.). Most of the studies compare different FOP labeling formats and try to identify which format guides consumers best in making healthier food choices (13, 15-18, 21, 22). The mean quality score of these studies was 48.8% (range 35.7% - 62.5%). A criterion on which the studies scored high in general was a “representative population sample of considerable size”: 9 of the 11 studies used large, well-balanced consumer panels consisting of 400 up to 2200 consumers (13-15, 17-20, 23). Criteria on which most studies received low scores were the “explanation for choice of data collection tools” and the “fit between research question and method of data collection”: many different tools were used to measure “use” and “understanding,” such as labeling tasks with photos, choice cards, computer tests, and comparing mock packages (13-16, 18). How well do these experimental self-reported data reflect actual understanding and use in real life shopping environments? A critical reflection on the chosen tools in the limitations sections was generally lacking. Also, the studies scored poorly with regard to validity of the measurement tools: although one study referred to some pretested scales (13), none of the used questionnaires were validated.

Only two studies used focus group interviews (22, 23). These qualitative group interviews were used to provide more in-depth insights as to how consumers understand and use FOP labels while, for example, showing some product packages.
### Table 1. Study characteristics of front-of-pack (FOP) labeling studies included in this review (n=31) and their quality scores.

<table>
<thead>
<tr>
<th>Type of study</th>
<th>First author, year (reference number)</th>
<th>Type of existing FOP label</th>
<th>Subjects (n)</th>
<th>Setting</th>
<th>Study design/control group</th>
<th>Main outcome measure(s)</th>
<th>Type of data collected</th>
<th>Authors’ conclusions</th>
<th>Quality score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported consumer studies</td>
<td>Andrews, 2011 (13)</td>
<td>Smart Choices, traffic lights – GDA</td>
<td>US consumers (520)</td>
<td>Online survey, mock packages in questionnaire</td>
<td>Cross sectional</td>
<td>Healthiness perception, nutrient use, purchase intentions</td>
<td>Self-reported questionnaire data</td>
<td>Smart Choices can lead to positive and potentially misleading nutrient and healthiness evaluations when compared to traffic lights-GDA or no FOP.</td>
<td>54.8%</td>
</tr>
<tr>
<td></td>
<td>Balcombe, 2010 (14)</td>
<td>Traffic lights</td>
<td>UK consumers (477)</td>
<td>Experimental setting with choice cards in questionnaire</td>
<td>Cross sectional</td>
<td>Understanding of traffic lights</td>
<td>Questionnaire data</td>
<td>Consumers prefer to reduce any nutrient with red light; different consumer groups respond differently to traffic lights.</td>
<td>52.4%</td>
</tr>
<tr>
<td></td>
<td>Borgmeijer, 2009 (15)</td>
<td>Traffic lights, GDA</td>
<td>Consumers (420)</td>
<td>Experimental setting, tasks with photos with labeled foods</td>
<td>Randomized experimental design: 4 labeling groups and 1 control group</td>
<td>Identification of healthier food in pairwise comparisons and daily food selection</td>
<td>Photo selections by interviewer</td>
<td>Most correct identifications of healthier foods with traffic lights; envisaged daily food consumption did not differ among conditions.</td>
<td>45.2%</td>
</tr>
<tr>
<td></td>
<td>Feunekes, 2008 (16)</td>
<td>Traffic lights, Wheel of Health (+ fake labels)</td>
<td>Consumers from 4 European countries (1630)</td>
<td>Online questionnaire</td>
<td>Cross sectional</td>
<td>Consumer friendliness and usage intention</td>
<td>Self-reported questionnaire data</td>
<td>Simpler labeling formats more appropriate to make quick purchase decisions.</td>
<td>40.5%</td>
</tr>
<tr>
<td>Type of study</td>
<td>First author, year (reference number)</td>
<td>Type of existing FOP label</td>
<td>Subjects (n)</td>
<td>Setting</td>
<td>Study design/control group</td>
<td>Main outcome measure(s)</td>
<td>Type of data collected</td>
<td>Authors' conclusions</td>
<td>Quality score</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------------</td>
<td>---------------------------</td>
<td>--------------</td>
<td>---------</td>
<td>---------------------------</td>
<td>------------------------</td>
<td>----------------------</td>
<td>---------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Self-reported consumer studies (continued)</td>
<td>Gorton, 2009 (17)</td>
<td>Traffic lights, GDA</td>
<td>New Zealand shoppers (1525)</td>
<td>25 supermarkets</td>
<td>Cross sectional</td>
<td>Understanding and use of different FOP labels</td>
<td>Questionnaire data</td>
<td>High reported use of labels; traffic lights showed high understanding across ethnic and income groups.</td>
<td>59.5%</td>
</tr>
<tr>
<td></td>
<td>Kelly, 2009 (18)</td>
<td>Traffic lights, GDA</td>
<td>Australian consumers (790)</td>
<td>Experimental setting, choice tasks with mock packages</td>
<td>Cross sectional</td>
<td>Consumer ability to compare the healthiness of labeled products</td>
<td>Questionnaire data + interviews about the choice tasks</td>
<td>Traffic lights were most effective in assisting consumers to identify healthier foods.</td>
<td>47.6%</td>
</tr>
<tr>
<td></td>
<td>Larsson, 1996 (19)</td>
<td>Keyhole</td>
<td>Swedish women (616)</td>
<td>Experimental setting</td>
<td>Cross sectional</td>
<td>Knowledge of Keyhole and relation to dietary fibre and fat intake</td>
<td>Questionnaire data + 24h recall</td>
<td>Most women understood meaning of Keyhole, but no difference in total fat or fiber intake between less and more knowledge.</td>
<td>50.0%</td>
</tr>
<tr>
<td></td>
<td>Larsson, 1999 (20)</td>
<td>Keyhole</td>
<td>Swedish participants (1591)</td>
<td>Experimental setting</td>
<td>Cross sectional</td>
<td>Knowledge of Keyhole and relation to dietary behavior</td>
<td>Questionnaire data + food frequency questionnaire</td>
<td>Higher intake of Keyhole low-fat foods in people with knowledge about Keyhole, but this counted not for less educated ones.</td>
<td>50.0%</td>
</tr>
</tbody>
</table>
### 8. Methodological review

<table>
<thead>
<tr>
<th>Type of study</th>
<th>First author, year (reference number)</th>
<th>Type of existing FOP label</th>
<th>Subjects (n)</th>
<th>Setting</th>
<th>Main outcome measure(s)</th>
<th>Type of data collected</th>
<th>Authors’ conclusions</th>
<th>Quality score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported consumer studies (continued)</td>
<td>Moser, 2010 (21)</td>
<td>GDA, traffic lights</td>
<td>Consumers from Germany (147) and Belgium (128)</td>
<td>Consumer surveys (face to face interviews + self-administered questionnaires) Cross sectional</td>
<td>Understanding and preference</td>
<td>Self-reported survey data</td>
<td>German consumers prefer traffic lights, while Belgium consumers prefer GDA. Socio-demographics also play an important role.</td>
<td>35.7%</td>
</tr>
<tr>
<td></td>
<td>Signal, 2008 (22)</td>
<td>Pick the Tick, traffic lights</td>
<td>Maori, Pacific and low-income New Zealanders (158)</td>
<td>Focus groups interviews Cross sectional</td>
<td>Understanding and use of different labeling systems Qualitative focus groups</td>
<td></td>
<td>Lack of use of nutrition labels due to lack of time, lack of understanding, shopping habits and absence on low-cost foods.</td>
<td>38.1%</td>
</tr>
<tr>
<td>Vyth, 2009 (23)</td>
<td>Dutch consumers (2200)</td>
<td>Choices</td>
<td>Online questionnaires + focus group interviews Cross sectional design on two different times</td>
<td>Exposure, and reported logo use</td>
<td>Self-reported questionnaire data + qualitative focus groups</td>
<td></td>
<td>Exposure to logo increased over time; health-interested consumers use Choices logo.</td>
<td>62.5%</td>
</tr>
<tr>
<td>Observational consumer studies</td>
<td>Bialkova, 2010 (24)</td>
<td>Choices, GDA</td>
<td>Dutch young adults (24)</td>
<td>Experimental setting, visual search tasks on computer Cross sectional</td>
<td>Determinants of consumer attention to labels Reaction time, accurate answers on tasks</td>
<td></td>
<td>Display size, color, familiarity, and location are key determinants of consumer attention.</td>
<td>64.3%</td>
</tr>
<tr>
<td>Type of study</td>
<td>First author, year (reference number)</td>
<td>Type of existing FOP label</td>
<td>Subjects (n)</td>
<td>Setting</td>
<td>Study design/control group</td>
<td>Main outcome measure(s)</td>
<td>Type of data collected</td>
<td>Authors’ conclusions</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>---------------------------------------</td>
<td>---------------------------</td>
<td>--------------</td>
<td>---------</td>
<td>---------------------------</td>
<td>-------------------------</td>
<td>------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Observational consumer studies (continued)</td>
<td>Grunert, 2010 (25)</td>
<td>GDA</td>
<td>Shoppers from 6 European countries (11781)</td>
<td>14 major retailers in 6 countries</td>
<td>Cross sectional</td>
<td>Use and understanding of nutrition information</td>
<td>In-store observations, in-store interviews and questionnaires</td>
<td>Understanding higher than use, possibly due to lack of motivation. Considerable national differences in both understanding and use.</td>
</tr>
<tr>
<td></td>
<td>Grunert, 2010 (26)</td>
<td>GDA, traffic lights, traffic lights + color coded GDA</td>
<td>UK shoppers (2019)</td>
<td>3 major UK retailers</td>
<td>Cross sectional</td>
<td>Use and understanding of FOP nutrition information</td>
<td>In-store observations, in-store interviews and questionnaires</td>
<td>Traffic lights helped to pay attention to important nutrients and improved accuracy of healthiness ratings compared to standard label.</td>
</tr>
<tr>
<td></td>
<td>Jones, 2007 (27)</td>
<td>Traffic lights</td>
<td>UK participants from university setting (92)</td>
<td>Experimental setting, labeling tasks on computer</td>
<td>Cross sectional</td>
<td>Perceived healthiness of nutrition label and the areas of label examined</td>
<td>Eye movements + healthiness ratings</td>
<td>Only Tesco shoppers rarely used FOP label. Shoppers claim to use FOP labels, although thinking aloud revealed hardly actual use.</td>
</tr>
<tr>
<td></td>
<td>Rayner, 2001 (28)</td>
<td>Logos from Tesco, Sainsbury and Pick the Tick</td>
<td>Shoppers from UK and Australia (44)</td>
<td>Supermarket</td>
<td>Cross sectional, thinking aloud 2 times: shopping normally and shopping healthily</td>
<td>Reported use and actual use of FOP</td>
<td>Qualitative data from thinking aloud + quantitative interviews + cash receipts</td>
<td>Only Tesco shoppers rarely used FOP label. Shoppers claim to use FOP labels, although thinking aloud revealed hardly actual use.</td>
</tr>
<tr>
<td>Type of study</td>
<td>First author, year (reference number)</td>
<td>Type of existing FOP label</td>
<td>Subjects (n)</td>
<td>Setting</td>
<td>Study design / control group</td>
<td>Main outcome measure(s)</td>
<td>Type of data collected</td>
<td>Authors’ conclusions</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------</td>
<td>---------------------------</td>
<td>--------------</td>
<td>--------------------------</td>
<td>-----------------------------</td>
<td>--------------------------</td>
<td>------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Observational consumer studies (continued)</td>
<td>Reid, 2004 (29)</td>
<td>Canada’s Health Check symbol</td>
<td>Canadian shoppers (200)</td>
<td>Supermarket</td>
<td>Cross-sectional</td>
<td>Number of Health Check products purchased by shoppers</td>
<td>Grocery receipts + questionnaire data</td>
<td>Shoppers purchasing Health Check products had lower fat intake; logo awareness related to use and to interest in healthy foods.</td>
</tr>
<tr>
<td></td>
<td>Steenhuis, 2010 (30)</td>
<td>Choices</td>
<td>Female consumers from university setting (36)</td>
<td>Experimental lab-setting, taste experiment</td>
<td>Cross-over design with 2 conditions: logo and no-logo</td>
<td>Weighed consumption of chocolate cake</td>
<td>Consumption data + questionnaire data</td>
<td>Use of logo did not increase consumption and no effect on taste rating.</td>
</tr>
<tr>
<td></td>
<td>Visschers, 2010 (31)</td>
<td>General FOPs on breakfast cereals in Swiss</td>
<td>Swiss students (32)</td>
<td>Experimental setting, food choice task on computer</td>
<td>Between-subjects design, respondents randomized over 2 conditions: health or taste motivation</td>
<td>Eye movements + questionnaire</td>
<td>Eye movements</td>
<td>66% perceived nutrition label and/or FOP. Health motivation and package design direct consumer’s attention towards on package nutrition information.</td>
</tr>
<tr>
<td></td>
<td>Vyth, 2010 (32)</td>
<td>Choices</td>
<td>Dutch consumers (404)</td>
<td>Nine supermarkets</td>
<td>Cross-sectional</td>
<td>Proportion of purchased logo products</td>
<td>Product observations + questionnaire data</td>
<td>Health-interested consumers purchase most logo products; hedonists purchase least logo products.</td>
</tr>
<tr>
<td>Type of study</td>
<td>First author, year (reference number)</td>
<td>Type of existing FOP label</td>
<td>Subjects (n)</td>
<td>Setting</td>
<td>Study design/control group</td>
<td>Main outcome measure(s)</td>
<td>Type of data collected</td>
<td>Authors' conclusions</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------</td>
<td>---------------------------</td>
<td>--------------</td>
<td>---------</td>
<td>----------------------------</td>
<td>------------------------</td>
<td>----------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Sales</td>
<td>Sacks, 2009 (36)</td>
<td>Traffic lights</td>
<td>Supermarket store chains of 1 retailer (over 1000) in UK</td>
<td>One retailer</td>
<td>Natural experiment with measurements during 8 weeks</td>
<td>Sales of 6 ready meals and 12 sandwiches</td>
<td>Sales data</td>
<td>Introduction of traffic lights had no effect on relative healthiness of consumer purchases</td>
</tr>
<tr>
<td></td>
<td>Sutherland, 2010 (37)</td>
<td>Guiding Stars</td>
<td>Supermarket chain stores in US (168)</td>
<td>One store chain</td>
<td>Natural experiment with 3 measurements in 2 years</td>
<td>Proportion of food items with star-rating</td>
<td>Sales data</td>
<td>Proportion of purchased star-items increased at 1- and 2-y follow-up.</td>
</tr>
<tr>
<td></td>
<td>Vyth, 2011 (38)</td>
<td>Choices</td>
<td>Dutch worksites (25) and employees (368)</td>
<td>Worksite cafeterias</td>
<td>Randomized controlled trial with 1 labeling and 1 control condition for 9 weeks</td>
<td>Proportion of Choices sandwiches and soups</td>
<td>Sales data + questionnaire data</td>
<td>No effect on sales when comparing logo to non-logo cafeterias; health-interested employees report to use logo.</td>
</tr>
<tr>
<td>Reformulation</td>
<td>Vyth, 2010 (33)</td>
<td>Choices</td>
<td>Dutch Choices products (821)</td>
<td>47 Food manufacturers</td>
<td>Cross sectional</td>
<td>Reason of logo assignment and nutrient composition data</td>
<td>Data provided by food manufacturers</td>
<td>Choices had stimulated healthier product development, especially regarding sodium and dietary fiber.</td>
</tr>
</tbody>
</table>

8. Methodological review
<table>
<thead>
<tr>
<th>Type of study</th>
<th>First author, year (reference number)</th>
<th>Type of existing FOP label</th>
<th>Subjects (n)</th>
<th>Setting</th>
<th>Study design/control group</th>
<th>Main outcome measure(s)</th>
<th>Type of data collected</th>
<th>Authors’ conclusions</th>
<th>Quality score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reformulation (continued)</td>
<td>Williams, 2003 (34)</td>
<td>Pick the Tick</td>
<td>Tick products from Australia (12)</td>
<td>1 Food manufacturer</td>
<td>Cross sectional</td>
<td>Sodium content before and after (re) formulation</td>
<td>Chemical analyses + data provided by food manufacturer</td>
<td>Tick influenced food manufacturer to remove 235 tonnes of salt from breakfast cereals.</td>
<td>26.2%</td>
</tr>
<tr>
<td></td>
<td>Young, 2002 (35)</td>
<td>Pick the Tick</td>
<td>Tick products from New Zealand (23)</td>
<td>Food manufacturers</td>
<td>Cross sectional</td>
<td>Sodium content before and after (re) formulation</td>
<td>Chemical analyses + data provided by food manufacturers</td>
<td>Tick influenced food companies to exclude 33 tonnes of salt through (re)formulation of breads, breakfast cereals and margarines.</td>
<td>50.0%</td>
</tr>
<tr>
<td>Modeling and health outcomes</td>
<td>Ireland, 2010 (39)</td>
<td>Pick the Tick</td>
<td>Adult Australian consumers (49)</td>
<td>Real life setting, free living</td>
<td>One group randomized to Tick products and others to low salt foods for 8 weeks</td>
<td>Sodium excretion</td>
<td>24h urine samples</td>
<td>Simple dietary education was effective in reducing sodium intake in free-living individuals in a short time study.</td>
<td>88.1%</td>
</tr>
<tr>
<td></td>
<td>Roodenburg 2011 (40)</td>
<td>Choices</td>
<td>Intake data from 7 countries</td>
<td>Virtual setting, modeling</td>
<td>Comparisons of typical daily menus with Choices daily menus</td>
<td>Nutrient intakes</td>
<td>Food composition + food consumption data combined</td>
<td>Replacing typical daily menus by Choices menus can potentially lead to improved nutrient intakes towards recommendations.</td>
<td>69.0%</td>
</tr>
<tr>
<td>Type of study</td>
<td>First author, year (reference number)</td>
<td>Type of existing FOP label</td>
<td>Subjects (n)</td>
<td>Setting</td>
<td>Study design/control group</td>
<td>Main outcome measure(s)</td>
<td>Type of data collected</td>
<td>Authors' conclusions</td>
<td>Quality score</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------------------------</td>
<td>---------------------------</td>
<td>--------------</td>
<td>---------</td>
<td>-----------------------------</td>
<td>-------------------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Modeling and health outcomes (continued)</td>
<td>Roodenburg, 2009 (41)</td>
<td>Choices</td>
<td>Dutch young adults (750)</td>
<td>Virtual setting, modeling</td>
<td>Evaluation of different Choices compliant scenarios</td>
<td>Nutrient intakes</td>
<td>Food composition + food consumption data combined</td>
<td>Replacing foods by Choices compliant foods can potentially lead to improved nutrient intakes.</td>
<td>59.5%</td>
</tr>
<tr>
<td></td>
<td>Sacks, 2010 (42)</td>
<td>Traffic lights</td>
<td>Australian adult population</td>
<td>Virtual setting, modeling</td>
<td>Evaluation of traffic lights and junk food tax scenario</td>
<td>Weight reduction, DALY's averted and costs</td>
<td>Data from national nutrition survey + costing data</td>
<td>Traffic lights labelling and junk food tax are highly cost-effective as obesity prevention measures.</td>
<td>69.0%</td>
</tr>
<tr>
<td></td>
<td>Temme, 2010 (43)</td>
<td>Choices</td>
<td>Dutch young adults (750)</td>
<td>Virtual setting, modeling</td>
<td>Evaluation of different Choices scenarios by using market shares</td>
<td>Nutrient intakes</td>
<td>Food composition + food consumption data combined</td>
<td>With Choices foods available in 2007, SAFA and sugar intake can be slightly reduced.</td>
<td>69.0%</td>
</tr>
</tbody>
</table>

1 The quality of the 31 included studies was assessed by two independent researchers by using the quality assessment tool developed by Sirriyeh et al. (44). The tool consists of 16 criteria. Each research paper was awarded a score on a scale from 0 to 3 for each of the criteria. Discussion following the independent scoring of papers resolved any differences in agreement by the two researchers. The sum of the scores provided a quality score per paper, range: 0-100% (with 100% the maximum score possible).
Observational consumer studies

We found 9 studies which used observational methods to measure FOP label understanding and use (24-32). Self-reported data appear to over-estimate actual FOP label use in real life settings (26, 28, 32), stressing the importance of collecting more objectively assessed behavioral data. The mean quality score of these studies was 68.6% (range 50.0% - 76.2%). A criterion on which the studies in general scored high was a “clear description for the choice of data collection tools”: different observational data were collected, such as product observations in supermarkets (25, 26, 32), collection of grocery receipts (29), reaction time records when doing a computer task (24), tracking eye-movements when doing computer labeling tasks (27, 31), thinking aloud data (shoppers were asked to “think aloud” during their shopping trip, and the conversations were recorded on tape) (28), and food consumption data (30). Five studies were conducted in the real life supermarket setting (25, 26, 28, 29, 32), leading to a higher quality score because actual supermarket observations better reflect actual shopping behavior than studies conducted in experimental settings (26, 32).

However, most of these studies struggled with the question of how accurately observational data really reflect actual FOP label understanding and use, and identified this in their study limitations. As stated by Grunert et al. (26), even though shoppers...
may have looked at nutrition information in-store, this does not necessarily mean that this information had an impact on their choice. Or, looking at the observational data of Vyeth et al. (32), and Reid et al. (29), neither of the studies are able to conclude whether health-conscious participants purchase FOP labeling products due to the logo or due to another reason. Also, the eye-tracking studies report that their results do not indicate whether respondents understood the information they perceived correctly, hampering the interpretation of the data (27, 31). The studies stress the importance of conducting more longitudinal studies in which causality can be assessed. Further, although most studies used some validated items from prior research, e.g. the food choice questionnaire (32), the dietary restraint scale (30), questions to assess the use of food package information (29), or an instrument measuring nutritional knowledge (25, 26), none of the used measurement tools and questionnaires reported being validated for the specific research purpose.

**Sales studies**

Only two studies have been published which collected supermarket sales data before and after the introduction of a FOP label to study whether the FOP label influences sales. These studies had a quality score of 66.7% (36) and 59.5% (37). Both scored high on “rationale for choice of data collection tools”: both studies collected a large amount of objective longitudinal sales data from real-life supermarket chain stores. The study by Sacks et al. scored lower on “representative sample of products” than Sutherland et al. because Sacks et al. only collected sales data from 6 ready meals and 12 sandwiches, whereas Sutherland et al. looked at all products with the Guiding Stars symbol (36, 37). Further, both studies scored low on “fit between research question and method of data collection”: both studies had no control group. The change in sales could have occurred due to new product and package introductions, possibly in combination with the FOP label or other on-package nutrition information such as “low fat” or “light” statements, effects of price discounts, product group promotions, and/or product life-cycles (45). Sutherland et al. do not discuss this limitation, but Sacks et al. discuss that attributing the observed increase in sales to the introduction of the FOP labels is not completely possible, as the products examined were also reformulated at the time the labels were introduced, and the product packaging and manufacturer was changed (36).

One other study measured changes in sales after the introduction of a FOP label, but in worksite cafeterias, not in supermarkets (38). This randomized controlled study in 25 worksites measured objective sales data, consisted of a sample of reasonable size, and did include a control group in the longitudinal design, which makes it possible to link any change in sales to the label (quality score 83.3%).

**Reformulation**

Only three studies have been published that evaluated the effects of FOP labels on product development (33-35). The mean quality score of these studies was 48.4% (range 26.2% - 69.0%). Young and Swinburn (35), and Vyeth et al. (33) scored high for example on a “clear description of the procedure for data collection” (data provided by food manufacturers) and “clear explanation for choice of data collection tools” (nutrient composition data before and after assignment of the FOP labels). Clear explanations for these items were lacking in the study by Williams et al. (34).
Low scores were assessed for the sample size of Young and Swinburn, and Williams et al.: these were quite small (23 and 12 products respectively), while the sample of Vyth et al. was larger (821), but still not exhaustive. Further, only Vyth et al. had a clear limitations section. An important limitation mentioned was that most data were self-reported by the food manufacturers.

**Health outcomes**

Epidemiological modeling is a way to investigate potential effects of FOP labels on nutrient intakes and health outcomes. We found 4 studies that evaluated the effects of FOP labels by modeling (40-43). The mean quality score of these studies was 66.6% (range 59.5% - 69.0%). These studies scored high for example on “fit between research question and method” and “representative sample size”: they estimate the effects of FOP labels on nutrient intakes and health outcomes based on national databases with population data instead of small consumer groups. Nevertheless, these studies scored low on “validity of the measurement tools”, mainly caused by the fact that modeling studies are based on so many highly selective assumptions. Assumptions in these 4 studies were related for example to compensation behavior, food replacement procedures, scenario development, costs estimations, and the associations between nutrients and health from limited literature sources. Only one studied the actual effects of FOP labels on biomarkers in the real life setting (39). Ireland et al. reached the highest quality score of all studies in this overview (88.1%). This high score is for example due to its longitudinal design with free living individuals in the real life setting, and due to its validated measurement tools: they used the biomarker “24h urinary sodium excretion”, which is considered the most reliable method of assessing sodium intake compared to more subjective measures such as dietary recall methods (46).

**Discussion**

This is the first study that provides an overview of the methodological quality of current FOP labeling research. Based on the quality assessments, we now identify some challenges for future research. Table 3 lists these research challenges which are discussed per subheading below. We start with the least relevant and we end this overview with the most relevant research challenge from a public health perspective. Public health relevance is illustrated by Figure 2. The upper part of this figure is based on the theoretical framework for studying consumer responses to nutrition labeling, developed by Grunert and Wills (47). Figure 2 is further clarified in the subheadings below.

**Self-reported consumer studies**

Self-reported consumer studies provide interesting initial insights into the understanding and intention of FOP label use. Providing a questionnaire to a consumer panel is considered a relatively easy, quick, and inexpensive way to collect research data. However, the relatively low quality scores of these studies and their relatively low public health relevance (see Figure 2) make them scientifically less interesting and the results should be used with caution with regard to policy recommendations. Whether these experimental self-reported data reflect actual use in real life shopping environments is highly questioned (26, 28, 32, 36).
Observational consumer studies
Observational consumer studies have higher quality scores and higher public health relevance than self-reported consumer studies (see Figure 2). Although observational studies lack the ability to assess causality, these studies better reflect actual behavior in real life settings, in which consumers are influenced by many food choice motives, such as price, taste, time and convenience (48). Nevertheless, in both self-reported and observational consumer studies, we identified a lack of a validated methodology and of a validated questionnaire to measure FOP label use, forming a challenge for future FOP labeling research.

Sales
This type of research is considered to be of higher public health relevance than small-scale consumer studies because of its large objective databases. Inclusion of a control group is essential to attribute the sales effect to the introduction of the FOP label alone. Regarding the study of Sutherland et al. (37), it would have been interesting if, for example, the FOP label had been introduced in half of the chain stores and in the other half 6 months later. Then, one could compare the sales of the labeled stores with the sales of the non-labeled stores (while keeping all other factors equal). However, sales databases are quite crude and cannot be used to reflect individual food intake (36, 38), explaining their lower public health relevance compared to studies measuring food intake and health outcomes (Figure 2). This stresses the importance of also collecting individual dietary intake data.

Reformulation
Although the reformulation studies had a relatively low quality score due to methodological weaknesses, they have a relatively high public health relevance: reformulation can increase the availability of healthier products and consequently has a

<table>
<thead>
<tr>
<th>Type of FOP labeling studies</th>
<th>Methodological challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported and observational consumer studies</td>
<td>Develop a validated methodology and a validated questionnaire to measure FOP label usage in real life settings</td>
</tr>
<tr>
<td>Sales</td>
<td>Introduce FOP label in half of the chain stores and compare sales of labeled stores with sales of non-labeled stores Collect individual purchase data and dietary intake data</td>
</tr>
<tr>
<td>Reformulation</td>
<td>Collect (chemically analyzed) food composition data, right from the start of the introduction of a FOP label Also collect data about unhealthy product introductions to evaluate the overall picture of the food supply</td>
</tr>
<tr>
<td>Health Outcomes</td>
<td>Collect updated food consumption and food composition data for modeling studies Measure health effects of FOP labels in real life settings by using biomarkers</td>
</tr>
</tbody>
</table>
large impact on all consumer groups without the necessity of changing behavior (33). The challenge is to collect objective food composition data. Ideally, we consider these data not to be self-reported data by food manufacturers, but chemically analyzed data. It is recommended to collect these reformulation data from the start of the introduction of a FOP label, because afterwards it can be difficult to retrieve. Also, it is likely the case that especially those manufacturers that participated in the reformulation studies are the ones that had significantly improved their products. Therefore, it is recommended to also collect data regarding how many unhealthy products were introduced in the same time frame to be able to evaluate the overall picture of the food supply.

**Health outcomes**

What are the effects of FOP labeling on health outcomes? Modeling studies provide some potential insights. If updated food consumption and food composition data are available, these studies can provide interesting insights into the potential effects of FOP labels on nutrient intakes and health (40, 42, 43). Although, the most relevant
question from a public health perspective remains, as illustrated by Figure 2: what are the actual effects of FOP labels on a population’s health? We consider the most interesting research challenge to be measuring health effects of FOP labels in real life settings by using biomarkers that are good predictors of disease risk.

Study limitations
First, this study is limited by the fact that it is not a systematic review. Nevertheless, it is unlikely that the studies we missed would have had a major impact on our conclusions.

Second, the quality assessments can be discussed. Although many validated quality assessment tools are currently available, they are largely limited to the assessment of studies with a specific research design: 50-60 tools are currently available to assess randomized controlled trial quality, along with a range of other tools for other single research designs (49, 50). Because FOP labeling studies are found to deal with diverse research methods and designs, we choose the tool of Sirriyeh et al. (44), which is able to evaluate overall quality of different designs. This tool is limited in that it relies on the researcher’s knowledge and expertise to enable fair and consistent assessments to be drawn. Nevertheless, we tackled this limitation by assessing the quality scores by two independent researchers. Large-scale validation of the tool is still needed (44).

The research challenge for the coming years
The two studies with the highest quality scores (38, 39) have interesting methodological characteristics in common. Both have a longitudinal randomized design and use observational methods to measure the effects of FOP labels in real life settings. Inclusion of a control group enables the ability to attribute the effect to the FOP label alone. Although there is no single, universally accepted hierarchy of evidence, there is broad agreement that randomized controlled longitudinal research designs in real life settings are providing one of the highest forms of scientific evidence (51). When we further take the public health relevance of FOP labeling studies into account, we consider measuring health effects of FOP labels in real life settings by using biomarkers as the research challenge for the coming years. It would be interesting to develop a randomized controlled longitudinal design, in which one group of consumers is educated about FOP labels, and have it explained to them that FOP labels can assist them to make healthier food choices, while the other group is not given any information. The main outcome measures can be changes in urinary sodium excretion, blood pressure and blood lipids.

Conclusions
Evaluations of FOP labels vary greatly in methodological rigor, and few methodologically sound studies are presently available. Highest methodological quality and public health relevance is achieved through measuring health effects of FOP labels by using biomarkers in a longitudinal randomized controlled design in a real life setting. We hope our research recommendations will challenge future researchers to further contribute to the interesting research area of front-of-pack labeling.
Conflicts of interest and authors’ affiliations
Annet Roodenburg is seconded at the VU University Amsterdam and employed by Unilever R&D, the Netherlands. The other authors have no conflicts of interest. Johannes Brug is employed by the Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands. The other authors are employed by the Department of Health Sciences and the EMGO Institute for Health and Care Research, VU University Amsterdam, the Netherlands.

References
14. Balcombe K, Fraser I, Di Falco S. Traffic lights and food choice: A choice experiment examining the relationship between nutritional food labels and
Methodological review


