Reviewer's report

Title: Impact of different food label formats on healthiness evaluation and food choice of consumers: A randomized-controlled study

Version: 2 Date: 2 February 2009

Reviewer: Peter Scarborough

Reviewer's report:

Impact of different food label formats on healthiness evaluation and food choice of consumers: a randomized-controlled study

The manuscript describes a randomized study that investigated whether different front-of-pack labelling systems influenced a) attitudes towards the healthiness of foods, and b) intention to purchase foods. The subject is particularly relevant at the moment and the authors appropriately reference other similar studies that have been conducted in different locations. The uniqueness of the location, and the value added by the investigation of people's intention to purchase foods make this a valuable addition to the literature. However, there are a number of issues that should be addressed before publication.

Discretionary revisions:

There are a number of sentences that could be revised to help with understanding. These include the following:

1. Worldwide more than 1.6 billion people aged 15 years and older are overweight, and approximately 400 million adults are obese, rising continues.
2. In Europe nutrition labelling is compulsory if a nutrition claim is made (big 4, big 8).
3. More or less simple labels are the health logos used in Australia, New Zealand or USA.
4. They understand the common signposting formats in the way that they believe that they understand them.
5. Foods from different food categories were photographed, revised, and printed in postcard size.
6. No significant differences between the examined label formats were found neither for any nutrient in gram, any nutrient expressed as energy percent nor energy density of chosen food.
7. However, the direction of this association is not quite clear. Perhaps 'causal nature' would be better than 'direction'.

The references need some attention (both how they appear in the text and in the reference list).
Minor essential revisions:

The criteria used to decide whether a food achieved a ‘tick’ were quite strong (requiring green traffic lights for sugar, fat, saturated fat and sodium). The restrictiveness of this system may have affected the results for this label group, and this should be discussed. Did the authors consider using a validated nutrient profile model for this categorisation?

The presentation of the results is sometimes misleading. In table 1, I think it would be more useful to show the socio-demographic characteristics by food label group, to show whether any potential bias may have been introduced by the randomisation process. Table 2 only shows whether there was significant variation between the five different label groups. I would prefer to see results of each of the four labelling systems versus ‘no label’, which would test the individual hypotheses of whether any of the labelling systems in itself is effective (compared to the null hypothesis of no labelling). The results comparing all of the systems can then be handled by the mean number of correct answers across all foods. Table 4 is not presented very well. It is too busy, and would benefit from a column that compares the results for the different labelling groups. The comparison with the NVSII data is valid, but should perhaps be presented in a separate table. There does not appear to be a table 3!

The description of the results focuses too much on the potential influences of demographic variables and BMI (when no substantial differences are found).

In the first paragraph of the discussion the authors state that ‘in our study we assessed objective understanding of nutrition information given in different signpost food label formats’. Given the limitations of the ‘virtual grocery’ section of the study, this seems a little strong. In the context of the paper, it seems that the subjective understanding was actually being assessed.

The results drawn from sales data from Tesco and Sainsbury’s that is referenced in the discussion (reference 32) are taken from press releases from the two supermarkets and there has never been any publication of the methods that were used to derive them, and hence they should be given very little credence.

Major compulsory revisions:

There are a number of limitations to the study design which influence the interpretation of the results. These limitations have either not been mentioned by the authors, or have not been discussed with enough prominence.

Some limitations that have not been mentioned by the authors include:

1. The randomisation process involved interviewers being randomised to only two of the five experimental groups. This introduces the potential of ‘interviewer bias’, where the responses of the participants are biased by the method in which the interviewer conducts the survey. Was there an attempt to standardise the data collection methods? If so, this should be reported. Was the potential of
interviewer bias tested for? In addition, since the interviewers recruited the study participants themselves, the potential for interviewer bias is increased (as the interviewers may have been more prone to recruit specific socio-demographic types that could influence results).

2. The calculation of BMI was done by self-reported height and weight measurements, which are known to be inaccurate. This is particularly a problem since the authors state in the discussion that the aim of the study was ‘to investigate whether different formats of signpost food labels help consumers to differentiate between healthy and less healthy foods, particularly with respect to body weight’.

3. There were a different number of foods drawn from each food category, and it is not clear how these foods were selected. Was the selection of the 28 pairs of food made with any attempt to produce a representative sample (e.g. of foods that are eaten or of foods that are generally available)? Since the results show that different labelling systems favour different products (‘no label’ seems to be the best condition for turkey breast) the results may have been biased by over-representation of certain food categories (e.g. the eight selected dairy products).

In addition, I would like to see more discussion of the following limitations:

1. The sample is over-loaded with people with a high educational background. The authors discuss this, but conclude that it is probably not too much of a problem because education level did not seem to have any impact on the results. This may only be the case because the educational level of the sample was high – a similar study of nutrition labelling with a sample of low educational level could have found a different impact of education.

2. The authors appropriately address the fact that the ‘virtual grocery’ was inadequate to accurately measure how people’s purchasing behaviour is affected by signpost labelling, and their conclusion that ‘there is little reason to believe that the labels would result in differences in real behaviour whereas they did not result in differences in our virtual, experimental situation’ seems reasonable. However, it seems that the lack of impact of food labels may have (partially) been a result of the restricted number of foods that were available to the participants. Since there was little substantial difference between the different labelling systems in the first half of the study, and the authors assumed that the participants had a reasonable idea of what was under investigation in the second task (and hence would try and select a healthy diet to ‘please’ the interviewer), it seems predictable that the resultant diets would have little differences between the different labelling groups. Given a wider selection of foods the differences between the labelling groups may have had a chance to present themselves.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Needs some language corrections before being
published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.