Author's response to reviews

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

Authors:

Ingrid Borgmeier (ingrid.borgmeier@ls.haw-hamburg.de)
Joachim Westenhoefer (joachim@westenhoefer.de)

Version: 3 Date: 19 April 2009

Author's response to reviews: see over
Dear Madam and Sirs,

We would like to thank the reviewers for their comments which helped us to clarify our ideas and hopefully improve the manuscript considerably.

In the following we give a detailed reply to the points which were raised by the reviewers.

First, as requested we documented in the Methods section (part study design and experimental condition) the approval of the ethical committee.

Comments to the points raised by Reviewer Eva Grammatikaki (Remarks to the Author):

Major Compulsory Revisions

Introduction:

1) It would be better if the authors clarified in the text what is meant by the big 4 (energy, protein, carbohydrate and fat) or big 8 (big 4+ sugars, saturated fat, fibre and sodium).

Done, we changed the sentence and explained what has to be declared.

2) It should be mentioned that the GDAs usually provide information on a basis of a 2000kcal diet.

Done, changed into:

Examples of detailed labels are the ‘Guideline Daily Amounts’ (GDA) showing the total amount of energy and nutrients in percentage that a typical healthy adult should be eating daily on a basis of a 2000 kcal diet.

3) Please rephrase the last sentence in page 4, paragraph 3 (EUFIC research undertaken…) with better English to better express what you mean.

Done, changed into:

Results from a UK market study on food labels indicated a high awareness of both, GDA and Traffic light label systems. In terms of understanding, the GDA concept is good. The understanding of the Traffic Light concept seems to be characterised by some exaggeration of the meaning of the colours.
4) Page 5, paragraph 2, lines 4-5: Please rephrase the sentence “They understand the common signposting formats in the way that they believe that they understand them” to better capture the meaning of what you want to say…

At this point it is not very clear what you mean exactly.

We changed this sentence to:

In addition most consumers state that they understand the information provided by these labels, which may be characterised as perceived or subjective understanding.

The notion of subjective or perceived understanding is taken up in the next sentence and contrasted with objective understanding.

Methods

5) Page 6, ‘Food Labels Format’

a. The “tick” label format as you present it, already provided a ‘combined’ evaluation of the foods’ nutrient content. This should be mentioned in this part.

We inserted the following sentence to clarify this:

Thus the “tick” label represents a combined evaluation of the food’s nutrient content, whereas the other label formats give separate evaluations for the different nutrients.

b. Another question arising from reading this part of the manuscript is whether the participants were informed or educated in the beginning of the study regarding the different food labels formats. This information should be mentioned in the methodology part.

The Participants were not informed or educated about the different label systems but were only informed about the label system of their own experimental condition. To clarify this we changed the description of the study to:

After answering some questions on demographic variables, body weight and height, each subject received a short, standardised description of the label format of his or her experimental condition and its meaning and after that completed two tasks: 1. Pair-wise comparison of foods and 2. virtual grocery.

6) Please add the details for the statistical package used.
We mentioned that SPSS Version 16 was used as the first sentence in the section on statistical analysis.

7) Please add if any corrections were made for the post-hoc analyses.

We added the following sentence:

Since these post-hoc test had exploratory character we abstained from adjusting significance levels for multiple testing.

8) Was the knowledge of consumers on nutrition assessed prior to the study? If yes, then some information should be added in the methodology part. If not, then please add this to the limitations of the study.

We added the following to our discussion of the limitations:

We did not test or investigate nutrition knowledge of our subjects. Thus we can not exclude that a particular level of nutrition knowledge in our sample has contributed to our findings.

Results

9) It would be good if the authors included the actual needs of the participants or at least the results of NVS II for the specific age group under study. Energy needs tend to differ from age to age. NVS II had a broader age range (14-80yrs) while this study has an age range of 22-50yrs (36±14yrs).

In order to estimate the needs energy requirements we would have to make assumptions about the physical activity level PAL which we did not try to assess in order to keep the study simple and feasible for the subjects. Using published recommendations might be problematic as well. The average energy intakes in the German population are according to the NVS II 15 per cent (young women) to 5 per cent lower (as the reference values from the German guidelines. In addition, we would like to point out that 36 ± 14 is the mean plus/minus the standard deviation. This range 22 – 55 covers in a normal distribution only 68 % of the values, not the total range. In our sample the total range ages has been 18 to 71 years which still is slightly less than 14 – 80 in the NVS. However, we think this could stand as it is, because the comparison between our sample and the NVS is only presented to show that the “envisioned intakes” from task 2 are all together in a reasonably plausible range.
10) Page 11, Lines 12-13: “showed that for sodium intakes there was an increase in educational level in the traffic light and colored GDA format” Please rephrase this with better English to better express what you mean.

The sentence has been rewritten to

The higher sodium intake was associated with higher education in the traffic light and coloured GDA condition, but with lower educational level in the simple tick condition (means are not presented in this paper). No systematic variation was found for the two other labels.

Discussion

11) The authors mention in the discussion part that the use of food labeling may have no major impact on individual food choice rather than the production of healthier products. However, they support this argument by describing how the sales of healthier products have gone up (and sales of less healthy products decreased) after some major supermarkets introduced the signposting system. This could truly lead to the production of more healthful products (increased demand leading to increased production). But doesn’t this also mean that people actually made more healthy choices after signposting was introduced?

We changed these arguments slightly to be more precise and picked up the point made by Peter Scarborough in his review:

Despite the finding that food labels have no major impact on individual food choice, one effect of nutrition labelling - in particular with prominent symbols like traffic lights - could lead to the production of more healthful products as manufacturers could wish to present their products with more favourable signposts {Baltas, 2001 #1045}. This may be reinforced by observations from Sainsbury’s and Tesco which seem to indicate that, after introduction of their signposting system (the Wheel of Health, a colour-coded GDA system, for Sainsbury’s and a not colour-coded GDA system for Tesco), sales of some healthier products went up whereas sales of comparable products with less favourable nutrient information went down {Grunert, 2007 #1034}. Such changes would imply changes of individual shopping behaviours. However, to our knowledge, methods and results the Sainsbury’s and Tesco findings have
not been scientifically reported and can therefore be given only very little cre-
dence.

12) Authors should also include in the limitation part that in real shopping situations, promotion activities and offers can also majorly affect consumer choices.

We added the following on page 12

Furthermore, nutrition labelling is only one of several information sources available to consumers. Media, advertising and promotion activities can also affect consumers’ choices {Baltas, 2000 #1045}{Kardes, 2004 #1168}.

In addition, we included a paragraph on the difference between the virtual shopping and real shopping situations as a second important limitation in the discussion.

13) In Table 2, you mention that milk chocolate is healthier than dark choco-
late…However as the percentages of correct answers show, participants probably based their answer not only on the food labels shown but also on the fact that dark chocolate is widely accepted as healthier than milk chocolate (due to antioxidants’ content etc). Although I understand that in terms of the “big 4” dark chocolate may have slightly higher caloric value and fat value (source: USDA food database), participants might have been confused by its higher nutritional value for health. I would suggest that you either remove it from the analyses of your results or at least discuss why participants were unable to identify the ‘correct option’ in terms of body weight.

Removing the chocolate item pair from the results could introduce additional bias in the results. This is particularly true, since milk chocolate as compared to dark chocolate is evaluated differently from people’s expectations in the discussed food label systems. Therefore we prefer to leave this in. However, we followed the suggestion to add an additional remark on this in the discussion – being aware that this increased again the length of the discussion which has been criticised by other review-
ers:

The results point partially to foods or food groups where food labels may be of particular value. In our dataset without food labels only a minority of subjects was able to identify milk chocolate as the healthier alternative in terms of fat, saturated fat and sugar. And even with food labels a large portion of subjects comes to a wrong decision regarding the “healthy variant particularly with re-
gard to a healthy body weight”. Presumably people use in their health evalua-
tions knowledge from the media which portrayed dark chocolate as healthy due to its health protecting ingredients as e.g. flavonoids. Such information seems to overwrite other information even if people are prompted to consider body weight.

Minor Essential Revisions

1) Page 3, paragraph 1, line 7: please change to “…and more than 100% increase…”
   done

2) Page 4, paragraph 5, line 4: change “lead” to “led”
   According to other we deleted these sentences in order to shorten the introduction

3) Page 13, paragraph 1, line 4: …household size have an effect…
   done

4) Figure legends: d) each portion…
   done

5) In Table 2, no percentages are marked in bold.
   done

6) In Table 2, please change “preserve” to “jam” to be consistent with the text.
   done

7) Table 1: in females please correct spelling of ‘middle’ educational level
   done

8) Please check the style of the citation in text. In some points it has not been done in the proper way.
   a. Page 3, paragraph 1, line 4: …(KiGGS) [3,4] has…
   c. Page 6, paragraph 4, line 1: …fat, saturated fat, sugar and sodium…
   d. Page 8, paragraph 2, line 4: …nutritionists…
   e. Page 12, paragraph 3, line 4: …range” [32,41].
We revised everything.

9) **Please also check the format of the references in the reference section.**

We used endnote with the file biomedcentral.ens to create the reference list. However, the reference type *report* and *government document* seems not to work correctly. Nevertheless, we hope that now everything should be ok.

Comments to the points raised by Reviewer Peter Scarborough (Remarks to the Author):

**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.**

Done, revised into:

Overweight and obesity are an increasing problem. Worldwide more than 1,6 billion people (age 15+) are overweight, and approximately 400 million adults are obese.

2. **In Europe nutrition labelling is compulsory if a nutrition claim is made (big 4, big 8).**

We changed the sentence. Please see as well Ms Grammatikaki’s 1st comment on page 1.

3. **More or less simple labels are the health logos used in Australia, New Zealand or USA.**

Done, the sentence is revised:

More or less simple labels are the health logos “pick the tick” in Australia and New Zealand, a tick symbol for approved foods low in total fat, saturated fat, added sugar and sodium, and the “smart spot” in the USA for products meeting similar nutrition criteria.

4. **They understand the common signposting formats in the way that they believe that they understand them.**
Changed (see Ms Grammatikakis comment 4):

In addition most consumers state that they understand the information provided by these labels, which may be characterised as perceived or subjective understanding.

5. *Foods from different food categories were photographed, revised, and printed in postcard size.*

Changed to:

Foods from different food categories were photographed, and the photos were 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.*

Changed to:

Intakes of energy, nutrients in gram and nutrients as percentage of energy intake did not differ significantly between the examined label formats.

7. *However, the direction of this association is not quite clear. Perhaps ‘causal nature’ would be better than ‘direction’.*

We followed the suggestion.

*The references need some attention (both how they appear in the text and in the reference list).*

We revised the reference and reference list in accordance to the journal style

**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?*

We did not really think that this approach had been too restrictive, but we had the impression that it is somehow straight forward. Therefore we did not consider other approaches. Nevertheless, it is true that this restrictiveness may have affected the results. The added a comment on this in the discussion as suggested.
A forth limitation lies in the criteria which we used for assigning the “tick” label. The “tick” label was only given, when all of the criteria to award the “green” colour for fat, saturated fat, sugar and sodium were fulfilled. This may represent a rather restrictive approach. Therefore results may be different if other approaches or criteria are used.

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 foodlabel group, to show whether any potential bias may have been introduced by the randomisation process.

Done, results of the different socio-demographic characteristics are presented for the different food label groups.

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.

We would prefer to leave the table as it is.

Group differences were examined to identify which signpost label enables participants best to identify healthier products; “No Label” was tested as an experimental condition itself and not as control condition. Therefore the table shows the correct answers in each experimental condition (for clarification the highest percentages are printed bold – which was unfortunately lost in the original manuscript we’ve sent).

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.

We have split this in to two tables, table 4 and 5. The comparison with NVSII is now included in the separate table 5.

There does not appear to be a table 3!

By mistake the table was not presented in a different file but in the manuscript.
The description of the results focuses too much on the potential influences of demographic variables and BMI (when no substantial differences are found).

We included the information on the lack of major influences of socio-demographic variables, because we have the experience from oral presentations of the study and its results that people ask exactly, whether such differences exist. Therefore we anticipated that the readers of the article would ask similar questions and hence answered these questions in advance.

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.

We tried to clarify what we mean with subjective and objective understanding, picking up the terminology used by {Grunert, 2007 #1034} in rephrasing this to:

Studies show, that consumers think they do understand the nutrition information on food packages correctly which has been called subjective understanding {Grunert, 2007 #1034}. However, less work has been done to assess actual and objective understanding of such information. In our study we assessed objective understanding of nutrition information given in different signpost food label formats in the sense that we tested whether subjects we actually able to identify the healthier food variants.

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.

We added this point to the discussion.

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).

We amended the paragraph with the limitation:

Since the interviewers recruited the study participants themselves, the potential for interviewer bias might be increased, as the interviewers might have been more prone to recruit specific socio-demographic types that could have influenced the 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’.

Obviously we expressed our idea very poorly. The “with respect to body weight” is related to the notion of “healthy food”. It was not meant that the analysis of different weight groups is the major focus of the paper. Hopefully this is clarified by rephrasing it to:

Aim of the present paper was to investigate whether different formats of signpost food labels help consumers to differentiate between healthy foods, healthy particularly with respect to body weight, and less healthy foods,

We added a corresponding limitation point in our discussion:

A fifth limitation is related to our conclusions regarding different body weight groups. We calculated BMI from self-reported weight and height which is known to be inaccurate and have restricted validity. Thus some of the subjects might have been misclassified as normal weight or overweight, and we can not exclude that results from weight related analyses would be slightly different if we had actually measured weight and height. However, all other inferences should not be influenced by this limitation.
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).

We explained the selection rationale in the methods section:

The selection of the food pairs was made with the attempt to produce a representative sample of foods from the different food categories that are generally available and commonly eaten. Dieticians with experiences in applied nutritional counselling were interviewed to find the appropriate foods. Foods were selected if a healthier and a less healthy alternative was available on the market and both variants are commonly eaten.

We added the following to the discussion of the limitations

In addition to the limited number of foods, some bias might result from the fact that within the different food groups a different number of foods or food pairs were selected for the experimental procedure. The results of the pair-wise comparison task indicated that different labelling systems favoured different products. Thus the results may be biased by the over-representation of some product groups, e.g. we used 8 dairy product pairs in which the traffic light system most often produced the highest percentage of correct responses, but only three grain/cereal comparison in which the simple “tick” logo yielded two times in highest percentage of correct responses. Therefore a different composition of the number of foods within each food group may yield different results.

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

We changed the discussion slightly to allow for this argument which is valid. Now the corresponding part is:

However, given this limitation, analyses taking into account the different levels of education did not find any hints that educational level differentially affected the understanding of food labels, the perceived healthiness of foods resulting from the understanding of the food labels or the intended/envisaged consumption of food labelled in different formats. Nevertheless, increasing the sample size or including a substantially higher number of subjects with low educational status might increase the power to detect differences related to education which we missed with our convenience sample.

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.

We added the following to discussion of the limitations to acknowledge these arguments:

In addition, the experimental procedure and the limited number of foods may contribute to the lack of differences in the virtual grocery task. The pair-wise comparison task showed overall that there were only modest differences between the different label formats. Given these small differences it may be expected that subjects select a similar diet if they tried to present a healthy food
choice to the interviewer as might be implied by the experimental context. Thus, we can not exclude that a wider selection of foods might yield a better chance for the different labels format to produce differences in food choice.

Comments to the points raised by Reviewer Klaus G. Grunert:

One small limitation that is not mentioned concerns the fact that different interviewers recruited respondents for different cells of the design - since this is a convenience sample (interviewers’ friends etc.?) this is a small deviation from complete randomization.

We added the following in the discussion of the limitations to acknowledge this aspect:

Since the interviewers recruited the study participants themselves, the potential for interviewer bias might be increased, as the interviewers might have been more prone to recruit specific socio-demographic types that could have influenced the results.

The introduction is too long and could be edited down. Especially the political aspects could be taken out. It will be enough to position the study in the context of previous research on the topic.

The introduction is edited down. Especially, the political situation and labelling discussion in Germany was taken out.

The discussion section is too long as well and references some additional material that is not directly linked to the study.

The discussion has been even more extended due to the requests of the other reviewers to include some additional points and limitations. But we also tried to include now only references which are directly linked to the results and conclusions of our study.

The paper needs editing and a language check.

The paper is edited in accordance to the details of all referees. Furthermore a language check has been done by a native English speaker. We hope that this enhanced the manuscript sufficiently. But we have to admit that being no native English speakers presents some sort of handicap in preparing such a paper.