Author's response to reviews

Title: Smoking, alcohol and dietary choices: evidences from the Portuguese National Health Survey

Authors:

Patricia Padrao (patriciapadrao@fona.up.pt)
Nuno Lunet (nlunet@med.up.pt)
Ana-Cristina Santos (acsantos@med.up.pt)
Henrique Barros (hbarros@med.up.pt)

Version: 5  Date: 30 May 2007

Author's response to reviews: see over
Porto, May 29th 2007

Dear Dr. Lolu da Silva,

We thank a lot the referees for their valuable comments and greatly thank the opportunity to revise our manuscript. We hope that the point-by-point response to their comments has addressed satisfactory all the issues pointed out, and that the manuscript is now ready for acceptance.

Yours sincerely,

Patrícia Padrão and Nuno Lunet
Reviewer: Aage Tverdal

Major Compulsory Revisions

Comment #1
I note that from the response letter that it is not possible to take dependency within household into account. This is surprising as long as household was the sampling unit. I think that this defect should be clearly stated in the Method paragraph.

Reply #1
We completely agree with your comment, and we also find it surprising. The database that is made available for researchers has one variable that defines the household, but several different households receive the same code, and it is impossible for us to know how many households are registered under the same code (the necessary information to transform this variable was not provided to us), despite our attempts to obtain this information.

This defect is clearly stated in the revised version of the manuscript, as follows (page 7, paragraph 2):
“Unfortunately, the present Portuguese survey database does not include the variables needed to consider the household cluster sampling in the analysis, and no correction was performed for the intracluster correlation.”

Comment #2
On the other hand, perhaps the problem is minor and not major, as participants were recruited from 21808 households. As there were 38225 participants (table 1), it was
only 1.75 persons in each household, on the average. Analysing men and women separately gives only 1 person in each household? In any case, dependencies between observations will decrease the power as compared to independent observations. The estimate of the variance will be too small and the confidence intervals too narrow. Whether the odds ratio estimates will change is dependent on the method of choice, GEE (generalized estimating equations) or random coefficient analysis. (confer for instance the discussion in Twisk JWR. Applied Longitudinal Data analysis for epidemiology, page 140). Thus the discussion part on this issue should be reformulated.

Reply #2

Our previous version of the manuscript was not clear regarding the number of participants per household. The number of households was 21808, and the number of participants 48606, corresponding to an average of 2.2 persons per household. Since our study included only subjects older than 19 years, the sample available for our analyses was 20302 women and 17923 men. This was made clearer in the revised version (page 5, paragraph 1).

Although we cannot compute the average number of persons per household in the sample used in our analyses, by analysing data from men and women separately a large proportion of the sampling units will have only one participant and the average number of participants per household (the average size of the clusters) will be much lower than two.

The impossibility of accounting for the effects of cluster sampling in our analysis contributed to an underestimation of the variances, and consequently a higher power as
compared to an analysis accounting for intracluster correlation (the design effect increases with the size of the clusters and with the intracluster correlation coefficient [1]), but our conclusions can hardly be attributed to the limitations in the analysis. Given the magnitude of the associations observed and the dose-response relations, which strengthen our conclusions, and the characteristics of the datasets analysed regarding clustering of the observations, it is unlikely that the associations shown in this study could be concealed if intracluster correlation was accounted in the analysis. On the one hand, by analysing data separately for males and females only one participant will be sampled in most households and the average size of the clusters will be much lower than two, allowing us to assume a design effect of two as a highly conservative estimate of what would be obtained with analysis procedures allowing to account for intracluster correlation, even assuming that the correlation between subjects within the same cluster can be conservatively assumed to be close to one (near the maximum). On the other hand, when considering the Odds Ratios for soup, fruit and vegetables consumption regarding the two highest categories of cigarette smoking, the standard errors would have to be inflated by more than 3.5 times, on average, to make the associations non statistically significant. A similar factor would have to be applied to the OR for the consumption of higher amounts of alcoholic beverages in men regarding the two highest categories of cigarette smoking, and a factor of 2.5 to the results from women regarding the consumption of beer and whisky. The dose-response relations are even more robust than the above mentioned associations.

An additional issue that needs to be accounted in the interpretation of our results is the fact that the OR estimates may change depending on the method used to perform the analysis accounting for the complex study design, but the effects of different approaches are much more difficult to predict. An empirical comparison [2] of different methods for
analysis of cluster randomized trials (standard logistic regression, standard logistic regression with robust standard errors, generalized estimating equation, random-effects logistic regression, Bayesian random-effects regression) showed differences not larger than 13% between the log odds ratios estimated through standard logistic regression and using any of the other methods. Our estimates could accommodate differences of this magnitude towards the null without compromising the conclusions.

As suggested, the discussion was reformulated to cover the points mentioned above (page 9, paragraph; page 10).

**Minor Essential Revision**

Comment #4

I still think the discussion moves into topics not supported by these data. One example is the smoking relation to homocysteine. It is nothing on homocystein in this study. It is probably true that there is a dose-response relationship between smoking dose and homocysteine level. But I am not sure what would be the best opinion on homocysteine and cardiovascular disease. The authors refer to one article from 2002 claiming that the link between hyperhomocysteinemia and cardiovascular disease has been established, but two randomised studies (HOPE-2 and NORVT) found no cardiovascular benefit from reducing homocysteine with folic acid and B vitamins. New Engl J Med, 2006 (march, I think).

Reply #4
We agree with these comments and the discussion was reformulated as suggested. The issues regarding the smoking relation to homocystein were excluded.
Reviewer: Andrew Roddam

Major Compulsory Revisions

Comment #1

The main problem is that the authors have failed/been unable to appropriately account for the survey design within the analysis. This is particularly problematic because it is highly likely that multiple respondents from households will have correlated measurements which have not been accounted for in the analysis. However as the authors perform all analyses separately for men and women this effect is likely to be somewhat small although their discussion of this issue is rather limited. As has been suggested at the very least this should be reformatted to talk about possible biases and potential changes to estimates of standard errors and their corresponding impact on statistical power. Furthermore the sampling strategy was a probability weighted sample and the authors should discuss the effect that not accounting for this survey design might have on the analysis.

Reply #1

We fully agree and these aspects were thoroughly discussed in this revised version.

Regarding the last point raised in this comment, the authors of the National Health described the sample as autoweighted, and provide no weights for the analysis (all participants should be given the same weight).

Comment #2

Another important limitation that the authors have failed to address is the issue of reporting validity. It is well known in dietary surveys the accuracy-validity is often
dependent on personal characteristics and therefore it is possible that smokers may, on average, be more likely to under or over report their consumption of a food. This was of some concern to me as the results in Table 4 seemed to suggest that the more you smoked the less likely you were to have eaten most foods the day before. Surely it would be reasonable to suggest that if you were less likely to eat certain foods you must (to maintain energy balance) be more likely to eat other foods. Is it simply that these foods were not captured by the very limited questionnaire? Or could reporting bias be a real problem here?

Reply #2

It was not a goal of the Portuguese National Health Survey to study quantitative aspects of food intake, and the food questionnaire was not specially designed to assess food intake. This survey employed generic classifications of food groups, rather than specific varieties or species, or quantitative measures, only a limited number of food items was considered, and we cannot estimate the specific composition of food recorded neither the quantity.

In relation to the method validity, no published studies are available, although the results showed good consistency and reliability (data not published). However we don’t have information on potential reporting biases.

The reporting of dietary habits is known to be influenced by personal characteristics and the association between smoking and the patterns of food and beverages consumption could be attributed to differential reporting in smokers and non-smokers, unless the increased intake of other items could be demonstrated to show the maintenance of the energy balance. Unfortunately, the questionnaire used in this survey includes a very
restricted number of food items and limited information regarding the intake of each specific item, and we are not able to discuss the energy balance. However, it is described in the literature that although smokers tend to eat less of some food groups (the more healthy, like vegetables and fruit) [3, 4], they tend to report increased intakes of other food groups (like chips, fatty meats, sugar) [5, 6], that were not included in this survey. Besides that, we have to bear in mind that alcoholic beverages (energy rich, therefore contributing for the energy balance) are more much frequent among smokers.

These aspects were further discussed in this revised version of the manuscript (Page 9, Paragraph 2)

Comment #3
The authors should combine Table 2 with Table 4 so that it is easier for readers to see the N’s and percentages in each category.

Reply #3
We agree. Tables 2 and 4 were combined (now presented as table 2).

Comment #4
In table 3 average beverage consumption is reported by sex and smoking status – but in the text there is report of associations between beverage consumption adjusting for age and education – why not report adjusted mean consumptions rather than raw consumption? Also why not include p-values for trend in the Table and make it more like Table 4?
Reply #4

We agree. The table 3 was substituted by table 3 and 4, presenting the association between smoking and consumption of beverages in a format similar to the adopted in table 4 (presented as table 2 in this revised version). Since the information on the consumption of beverages was more detailed than the available for foods, instead of presenting ORs obtained from logistic regression models using a dichotomous dependent variable (non-consumers and consumers), in tables 3 and 4 we present age- and education-adjusted ORs computed through multinomial logistic regression with a three-category dependent variable (non-consumers, consumers of amounts of beverages below the median observed in consumers and consumers of beverages in an amount above the median consumption in consumers). p-values for trend are also included in this table. The methods section was changed to describe this data analysis (page 7, paragraph 1).

Comment #5

I found some of the discussions – specifically relating to the health effects of fruit and vegetables etc to be inappropriate in this manuscript. This is a very complex subject in so far as smoking is a risk for chronic disease. F&V consumption might offer some protection (evidence more limited than what the authors seem to suggest) but smokers eat less F&V. I think the authors should significantly shorten this part of the discussion and stick to conclusions which they can support form their analysis – i.e. patterns of consumption of foods.

Reply #5
We agree with these comments and the discussion was reformulated as suggested, and some of the information was used in the introduction.

References