Author's response to reviews

Title: Family structure and breakfast consumption of 11-15 year old boys and girls in Scotland, 1994-2010: a repeated cross-sectional study

Authors:

Kate A Levin (kl1@st-andrews.ac.uk)
Jo Kirby (jilmk@st-andrews.ac.uk)
Candace Currie (cec53@st-andrews.ac.uk)

Version: 3 Date: 7 March 2012

Author's response to reviews: see over
I appreciate the authors' efforts in responding to reviewer comments. Two concerns remain. I consider both of them major. However, once the authors address those, I leave it to the editors' discretion about whether I need to see the paper again.

1) I appreciate why the authors had to choose 7 versus less as their cut-off. However, under circumstances, I think the MAIN results presented should use the less than 4/4 or more as the cut-off. It would also really help if the authors detail the distribution to the answers to those questions, so the readers can see how many of the 'less than 4' were actually 0s, and how many '4 or more' did NOT eat breakfast daily.

We believe that given the choice between the three cut-offs, daily breakfast consumption provides the most robust indicators, especially in light of the fact the question wording changed. As use of all three possible cut-offs result in identical conclusions, we have not changed all the analyses for the reviewers' preference. We have, however, already carried out the final analyses for all three possible cut-offs, and have mentioned this in the manuscript, offering the reader the opportunity to contact us for results using the two alternative cut-offs. We feel this is sufficient.

2) While the authors' concerns about multicollinearity with SES may be justified, I am still interested in knowing how sensitive the results are to its inclusion.

The short answer is not very. As we have already noted, multicollinearity was only one reason for not adding SES (or FAS). FAS items have changed over time with the addition of holidays in 1998 and computers in 2002. It was therefore not possible to split the distribution into low, medium and high tertiles for the 1994 dataset.

In a 'next stage' of analysis (currently in submission), not only SES but all contextual variables collected by the 2002-2010 HBSC surveys related to family members, relationships and the home environment are studied. Modelling the data from this study for daily breakfast consumption shows that family structure inequalities remain, with effect sizes only marginally affected, after adjustment for SES- see table below (note that this study included only 4 family structures). This is in line with most other studies of breakfast consumption (see Pearson et al, 2009). You can also see suggested increasing family structure inequalities in the table, although you need to model all data together with an interaction term to confirm that these changes are significant.

<table>
<thead>
<tr>
<th></th>
<th>2002</th>
<th>2006</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 1</td>
</tr>
<tr>
<td></td>
<td>Parameter</td>
<td>Parameter</td>
<td>Parameter</td>
</tr>
</tbody>
</table>
3) My biggest concern is the continued use of interaction terms in a logistic model. I realize BMC Public Health is not a leading methods journal, but it is still better to avoid methodological approaches that may be flawed. The non-linear nature of a logistic function makes it difficult to interpret interaction terms. Authors are referred to a recent article in Health Services Research (http://onlinelibrary.wiley.com/doi/10.1111/j.1475-6773.2011.01314.x/abstract) that outlines the challenges.

It is important not to confuse ‘flawed methods’ with ‘methods which are difficult to interpret’. Use of interaction terms in logistic regression is not a flawed method (I refer you back to the reference you cite), but is difficult to interpret and we absolutely agree with the conclusions of the paper you cite, that it is important to understand why interaction terms are included in nonlinear models to be clear about their substantive interpretation.

Sophisticated or complex methods that are difficult to interpret should not be replaced with inappropriate methods such as linear regression for the sake of easy interpretation. Of primary consideration when choosing a modelling method is the distribution of the data and the set of assumptions which the data must then adhere to under the modelling procedure used. Ignoring underlying assumptions of the model, such as data distribution, results in spurious results. Binary data therefore generally cannot be modelled using linear regression, as was previously suggested by the reviewer. I refer the reviewer once again to JS Long, 1997, pp 38-40.

We agree that interaction terms in probit/logit models are difficult to interpret however as we have interpreted these models for the reader in the results, and as we have provided untransformed regression coefficients for all variables (including the constant term) in the tables for more quantitative readers to combine and transform (exponentiate) if they so wish, we see no problem here.

HOWEVER we agree that we have made a small error in our interpretation of the interaction term- family structure differences in breakfast consumption increase are not linear, as previously implied- see final paragraph of the Results section for amendments highlighted in blue. Many thanks to the reviewer for spotting that something was not quite right!