Reviewer’s report

Title: Reweighting National Survey Data for Small Area Behaviour Estimates: Modelling Alcohol Consumption in Local Authorities in England

Version: 0 Date: 19 Jun 2018

Reviewer: Liz Twigg

Reviewer's report:

General comments

This article outlined a method to create synthetic estimates of alcohol consumption across local authorities. Although submitted as a research article, I felt that the article content was very descriptive and the research content was limited. I feel that more could be done in terms of research before it is ready for publication. The authors recognise that steps 1 and 2 of the reweighting method are no different from work that has already been published and it is not quite clear from the written text how the third stage can be 'used directly for statistical analysis or incorporated into more complex modelling work to produce locally representative policy effect estimates'. If this is the original aspect of this work then this should be the focus and provide a detailed research based example.

Admittedly it is sometimes difficult to discuss research around a method or technique but I think the authors have missed an opportunity to evaluate this approach and contribute to the research materials around the techniques involved in small area estimation of health behaviours (including alcohol consumption). There is quite a bit of literature on this and, at the very least, perhaps the estimates at LA level ought to be compared with those produced on the PHE website in their local mapping tool where binge drinking is provided at electoral ward level. Why is this method more useful/effective than other methods? In what types of places and for what types of people might this approach provide better estimate? Perhaps these are not questions that the reweighting survey method aims to address but point estimates do seem to be the end product of this paper and we do have many methods of synthetically producing point estimates of many health risk behaviours. There is some evaluative work but comparing the results with direct estimates from the HSE is a bit circular - the models are derived from HSE data and therefore estimates are bound to be highly associated. How do these estimates differ from other estimation approaches?
I also thought that the modelling approach was rather simplistic and some of the recognised problems with the model could have been rectified with a more complex model. For example I did not quite agree with the argument for separating out the abstainers from those who reported that they did consume alcohol. A better approach would be to simultaneously model a logistic (abstain/consume model) alongside a multinomial model (across consumption categories). This would allow comparison of statistical significance across covariates.

I take the argument that any geography could have been used as the focus for this paper but the focus on LA seem inappropriate. Yes this is the administrative area where public health funding is allocated but it is at sub-LA geographies where problems are targeted and where intelligence is needed to organise and deliver alcohol control initiatives. It is therefore at this more detailed level that this approach should be applied and the results compared/evaluated against other synthetic estimation processes to turn this into a useful research article.

This latter observation leads on to another general point. Place sensitive indicators must be modelled using place sensitive approaches and to assume that the relationship between the right and the left hand side of the regression equation is universal across different types of place or region is a false assumption. The authors do point this out in their discussion but there are now well-rehearsed modelling techniques which take these elements of complex heterogeneity into account and have been used in the estimation of health behaviours. Hierarchical modelling would allow the model probability estimates to be adjusted for regional residuals. Interaction terms in the models would also allow the contingent nature of some of the covariates to be explored and accounted for in the calculation of probabilities across the cells.

In essence this paper reads like the initial results of a work in progress where, eventually, more complex modelling, more detailed evaluative work and a deeper discussion of the third stage of the reweighting method could form the basis of a much more comprehensive research paper.

Detailed comments

Page 2 - background assumes that local equates with local authority in terms of local targeting for alcohol-related public health - needs more evidence to back this up.
Page 2 - there is much modelled estimate data available for public health workers on the PHE Local Health website - ie not direct estimates from surveys.

Page 3 - short summary of production of modelled point estimates is very brief - there are interesting nuanced differences in the methods - and many of these are place sensitive - also some are able to provide distributions of the behaviour in question across gender/age (and other socio-demographic) groups.

Page 4 - the third aim to create a reweighted HSE for each of the 151 UTLAs - more needs to be said about this (see general points above) - most of the paper seems to be about the point estimates derived from this reweighted distribution.

Page 4 - methods - why not place sensitive (ie multilevel/hierarchical regression) methods to model the data.

Page 5 - how was information on the UTLA of residence accessed?

Page 6 - What is the rational for only three ethnicity groups

The general comments above apply to some of the evaluative and discussion work presented in the rest of the paper.

**Level of interest**
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

**Quality of written English**
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published
Declaration of competing interests

Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal