Author’s response to reviews

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

Authors:
Robert Pryce (r.e.pryce@sheffield.ac.uk)
Colin Angus (c.r.angus@sheffield.ac.uk)
John Holmes (john.holmes@sheffield.ac.uk)
Duncan Gillespie (duncan.gillespie@sheffield.ac.uk)
Penny Buykx (p.f.buykx@sheffield.ac.uk)
Petra Meier (p.meier@sheffield.ac.uk)
Matt Hickman (matthew.hickman@bristol.ac.uk)
Frank de Vocht (frank.devocht@bristol.ac.uk)
Alan Brennan (a.brennan@sheffield.ac.uk)

Version: 1 Date: 09 Apr 2019

Author’s response to reviews:

We thank the editor and reviewers for their time. We respond to the reviewer comments below.

---

Reviewer #1: This is an interesting and well-written paper, largely of UK relevance. It does however tread well-covered ground. Twigg et al (2000, 2002) in Social Science and Medicine developed small area estimates of drinking behaviour. A UK Department of Health review in 2004/5 (Bajekal, Pickering et al) also examined alcohol consumption - and reviewed survey re-weighting. Both produced measures at more local scales than LAs. A key conclusion of both was the difficulty of estimating drinking behaviour due to the limitations of SAE methods and the complexity of the determinants. In the UK, PHE offer several relevant direct estimates relevant to drinking at the county/UA level. The claims to novelty in the present paper are a little overstated.
Response: We agree, we have included a brief discussion of the paper and report mentioned and reference them and have amended the language around the claims to novelty to reflect this.

This said, the reweighting proposed in the paper has novelty and the coverage of a range of consumption measures is interesting (though a straightforward extension of both the proposed method and previous papers). It is great to see that code will be available. Both make the paper useful and worth reporting - but two important issues will require attention:

a) The base data for the small area estimation is drawn from the HSfE. The HSfE is based on a hierarchical sampled and any analysis needs to be multilevel. It is not wholly clear but the paper does not mention using multilevel analysis. It talks of multinomial logistic regression with no mention of ML and no indication in that direction in the results tables. This may be important particularly in SAE where it cannot be assumed (as it seems to be here) that a fixed national association between drinking and age, sex, ethnicity, deprivation and hospitalization applies uniformly everywhere across England. It also has issues for the prediction of measures of uncertainty - confidence intervals - around the estimates - which should also be reported. At the very least, these ML issues need to be investigated, reported and commented upon, ideally reworking the SAEs.

Response: The reviewer is correct, our submitted article did not undertake multi-level modelling. The HSE is used as a nationally representative sample with its sample weights – which are calculated by the HSE to reflect their various data collection procedures. The reviewer is not quite correct that we have directly assumed “a fixed national association between drinking and age, sex, ethnicity, deprivation and hospitalization applies uniformly everywhere across England”. We control for a variety of Local area characteristics too – the region, the hospitalisation rate in the LA and we also tested the mortality rate in the LA. The essence of the multinomial logistic regression is that it accounts for both local area and individual level factors. Of course if the reviewer means that after controlling for local area factors we are assuming that when making a prediction for a locality we assume the same effects of age sex ethnicity and deprivation, then that is the case.

We do not believe that it is absolutely necessary to always undertake multi-level statistical modelling when utilising the HSE to perform statistical analysis. Indeed there are countless papers which do not do so. We have also looked at the references provided on SAE estimation by the reviewer and many of the papers in the review did not do so.
Having said that, we have taken on board the central point that a multi-level modelling approach could in principle be taken i.e. because the coefficients in the multinomial regression for age and sex could be different in different regions say. Therefore, we have done some investigation. The results of this are now incorporated into the paper and appendix. In short, we found problems fitting a multilevel model – with a lack of convergence and when the models tested did converge, a less good fit between model predictions and observed data than we obtain with the non-multilevel model. We have therefore retained our main results presentation as being those from our original regressions, and present the multilevel model findings in the appendix.

b) Related, the HSfE is designed to be representative at the regional level. Using the HSfE to build estimates and then re-aggregating them to the regional level represents a degree of circular reasoning. We need a more convincing approach to 'validation'. Typically in the SAE literature this involves internal validation - assessing the quality of model using standard diagnostics - and external validation as presented in this paper but needing a alternative source and a range of measures of concordance. The paper really should consider both.

Response: Thank you for this comment. We have provided additional information on the quality of model fit using standard diagnostics in appendix X.

We disagree with the reviewer on two points.

First, we believe the comparison of observed versus expected at regional level is useful. It is not entirely circular reasoning. We look at 4 separate statistics – mean consumption, % abstaining, % below 14 units, and % drinking at high risk levels. These local patterns across the spread of the distribution of drinking are the key purpose of the reweighting approach. And it is by no means certain that a model built in the HSE will secure a good fit to observed regional variations in these 4 metrics of drinking that are of interest. Indeed several models we tried do not – including the investigations of multilevel modelling that we have been able to undertake.

Second – we agree and disagree that external validation is required for publication. In principle yes we would agree. In practice – there is no solid external source for such an external validation. We were hopeful at one stage that there would be. Indeed, we obtained the results of a pilot set of surveys done by PHE on drinking at LA level in 25 different local authorities. When we compared our results against these it was clear that there were substantial differences. PHE have not released these survey results publicly and we understand though it has not been said explicitly that this is because the survey results themselves produce implausible patterns – inconsistent with patterns seen at national and regional level year after year in HSE – and that therefore the survey is somewhat flawed. This is the reason why external validation against another source is unfortunately not possible.
Reviewer #2: This paper reports a method for reweighting national survey data, which its authors then apply to the Health Survey for England. The method combines survey and publicly available data from sources external to the survey, to produce simulated locally representative survey data and provide statistics of alcohol consumption for each Local Authority in England. Based thereon, the authors identified a 2-fold difference in estimated mean alcohol consumption between the lightest and heaviest drinking Local Authorities, a 4.5-fold difference in rates of abstinence from alcohol, and a 3.5-fold difference in harmful drinking, defined as more than 35 units (approximately 12 fluid oz.) of ethanol, or about 20 U.S. standard drinks, per week for women, and 50 units (about 17.2 fluid oz.) of ethanol, or about 29 U.S. standard drinks, per week for men. The method was found to compare well to direct estimates from the data at the regional level. The authors conclude that, in addition to important policy implications in its own right, their method and the reweighted data it produces can also be used for modeling local policy effects, and for other public health small area estimation where locally representative data are not available.

Clearly there is a need to develop and implement methods that are both valid and user friendly for small-area estimation from larger regional or national survey data, since mounting small-area surveys with large enough samples for reasonable statistical power and precision quickly becomes resource prohibitive. The method presented by these investigators appears straightforward and adaptable beyond the immediate context in which it was applied in the present report. I only have a few comments:

1) It seems a bit odd to use consequences of alcohol consumption (e.g., hospital admission and mortality rates) to estimate consumption, especially since there is wide interindividual variability in the level of consumption required to trigger alcohol-related medical complications. Some further explanation and justification of the use of these measures as part of the estimation methodology would be useful.

Response: We have added some text to explain the rationale a little more. Essentially, areas with higher proportions of drinking would be expected to experience greater levels of harm – based on epidemiological studies worldwide that have found this. By logic this means that areas with higher levels of harm are statistically more likely to have higher numbers of heavy drinkers. It is this logic that we are using. We are not saying that the higher harm causes heavy drinking, just that the two are going to be statistically associated and because both are observed we can estimate the statistical association between them.
2) More generally, the independent variables the authors use to model consumption are specific to the British context. This is of course appropriate given their specific task. However, because the readership is international, I would recommend that they take a step back, whether in Methods, in the Discussion, or both, to describe in more generic terms the kinds of measures that can be brought to bear in adapting their method to the estimation of consumption in other national contexts.

Response: Thank you this is useful. We have added three sentences to the discussion paragraph around generalisability of our approach and what factors might be useful in contexts outside the UK.

3) These investigators are correct (p. 6, lines 158-159) that multinomial logistic regression allows the most flexibility in modeling and does not impose an assumption of proportionality of odds. However, there are other ordinal models that might be appropriate here and that do not require an assumption of proportionality of odds. If there truly is ordinality in these data, then the use of the multinomial model comes at a cost of loss of statistical power. If they have not already done so, the authors should examine this issue with expert biostatisticians.

Response: We considered a wide range of statistical models but our preferred model is the multinomial logistic model. Whilst there is ordinality in the data in that a higher consumption band means higher consumption, there may not be a monotonous relationship between dependent and independent variables. For example, alcohol consumption may be polarised for a given independent variable. For example, men may be more likely to either drink very heavily or not at all. A “male” coefficient in an ordinal model will assume a monotonous relationship between “male” and consumption. The multinomial logistic regression allows full flexibility.

Reviewer #3: 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.
Response: We have undertaken research to estimate LA level consumption of drinking. The other two reviewers were clear that this is valuable research. No action required.

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?

Response: Whilst the PHE binge drinking estimates are a useful resource, direct comparison is not possible as the outcome modelled is maximum daily alcohol consumption rather than mean consumption. The modelling is also done at local-area level with no individual-level explanatory variables so is not very similar to the method used in our paper.

On the second point – we have responded to this same comment above made by reviewer 1.

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.

Response: We have now tried a simultaneous model. And present the results in the appendix C.
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.

Response: We partially disagree. This work was used to support further research on alcohol pricing policies at local level. It is being used to consider the potential impact of local authorities implementing a minimum unit price for alcohol – funded by NIHR. We do agree that other levels of locality analysis could also provide benefit. But it was clear that LAs and regional representatives at two major stakeholder events each with 80 people in the North west and North East of England were keen to understand these LA level results and use them.

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.

Response: We have now done multilevel modelling.