Author’s response to reviews

Title: Socioeconomic Patterning of Vaping by Smoking Status Among UK Adults and Youth

Authors:

Michael Green (Michael.green@glasgow.ac.uk)
Linsay Gray (linsay.gray@glasgow.ac.uk)
Helen Sweeting (helen.sweeting@glasgow.ac.uk)
Michaela Benzeval (mbenzeval@essex.ac.uk)

Version: 2 Date: 03 Jan 2020

Author’s response to reviews:

Reviewer 3:

“Having said that, I have one major concern that requires, at the very least, some clarification. It also possibly may require some additional analysis to ensure variances (and confidence intervals) are correctly estimated. In particular, on page 7, lines 9-12, the authors state that "Multiple imputation was applied to include all respondents with valid weights, and 93.2% of the adults and 72.1% of the youth with valid weights had complete data on all analysis variables. Multiple imputation was conducted with an unconstrained model of all analysis variables (i.e. allowing each variable to predict all others) and all results were averaged across 25 imputed datasets." This statements are unclear. Typically, multiple imputation imputes values for observations that have missing data a specified number of times. Multiple values are imputed so as to recognize the uncertainty about the true value. This additional uncertainty has to be incorporated into the analysis phase, so that variance estimates (standard errors, p-values and confidence intervals) account for this additional uncertainty. In other words, the authors state that results were averaged across 25 imputed data sets. However, in multiple imputation, the analysis (e.g., regression model) is estimated "x" number of times, once for each imputed data set (in this case, it would be 25 times, since 25 imputations were conducted). Parameter estimates (coefficients and standard errors) across all 25 sets of analysis are then pooled for statistical inference. This process allows for the additional uncertainty attributable to imputation to be incorporated into variance estimates, standard errors, and confidence intervals. If the authors did indeed perform multiple imputation, they need to clarify this in the methods section so that readers are assured that reported p-values and confidence intervals reflect this additional uncertainty. If not, then p-values and confidence intervals might be under-estimated (too small) leading to spurious conclusions about statistical significance. If the authors did analyze the data this way, it needs to be reported as such. If not, the authors should be encouraged to run the analysis using correct multiple imputation procedures (e.g., PROC MIANALYZE in SAS or other packages such as Stata).”
We thank the reviewer for the opportunity to clarify this point. We have edited page 7, line 12-13 to read “and results were averaged across 25 imputed datasets using Rubin’s rules”. The (now cited) manual for the software utilised clarifies that “parameter estimates are averaged over the set of analyses, and standard errors are computed using the average of the standard errors over the set of analyses and the between analysis parameter estimate variation”.

On a minor note, the authors should report the statistical software used to conduct the analysis.

-we have added on page 7, line 6-7 that analyses were conducted in Mplus 8.