Author's response to reviews

Title: The contribution of risk factors to socioeconomic inequalities in multimorbidity across the lifecourse: A longitudinal analysis of the Twenty07 cohort

Authors:

Srinivasa Katikireddi (vittal.katikireddi@glasgow.ac.uk)
Kathryn Skivington (kathryn.skivington@glasgow.ac.uk)
Alastair Leyland (alastair.leyland@glasgow.ac.uk)
Kate Hunt (kate.hunt@glasgow.ac.uk)
Stewart Mercer (stewart.mercer@glasgow.ac.uk)

Version: 2 Date: 29 Jun 2017

Author’s response to reviews:

Dear Dr. Samuel

Thank you for considering our revised manuscript. We also thank the reviewers for their time in looking at our revisions. We provide a point-by-point response to the additional review below, detailing line and page numbers where changes have been made, and have tracked changes in the Microsoft Word manuscript submission. We have updated results tables following our reworking of the multiple imputation strategy based on the reviewer’s comments. We would like to note that there are only very small changes in actual figures (mostly less than 0.04 change in odds ratios) and no meaningful changes in results. The meaning of the findings and the discussion and conclusions did not therefore change.

Reviewer #4: Thank you to the authors for providing additional clarification in this revised version of the manuscript.

My main concern regards the multiple imputation analysis. The authors state that they have multiply imputed the data 15 times for item non-response and for occasional wave missing (not drop-out) and then they conclude that results are similar to complete case analysis, which in fact might not be true.

There is a table in the additional file (Table S10) that gives the results from the main analysis using complete case data. There are no differences of note between the complete case analysis results and the results using imputed data (Table 1 in the main paper).

I would like to see addressed the following important issues:
1) What was the proportion of missing data being imputed? Based on this proportion the authors should choose the number of multiply imputed data. No justification is given for the choice of 15. A table with the missing values for each of the variable imputed should be reported.

There were potentially 16,525 cases for analysis. There were 9,277 (56.14) complete cases. Data were not imputed for 3,664 cases that had dropped out or were dead at t-1; data were imputed and used for 3,584 cases. Total sample in the imputed datasets was 12,861. We have added an additional supplementary table (Table S-6) to show the missing values for each of the variables in the analysis plus the auxiliary variables. The outcome variable, multimorbidity, had the highest proportion of missing overall (20.0%). At one exposure/outcome time point (wave 2 to wave 3), 24% of the outcome multimorbidity was missing and imputed. We have therefore now used 24 rounds of multiply imputed data for our main analysis. We have updated the text (line 181) to reflect the change.

2) Did the authors impute missing data on covariates only or also on the outcome? This is an important point, because if the missingness of the outcome was imputed, then the authors should have added variables that were correlated with that missingness and/or with the outcome.

We imputed missing data on the outcome as well as covariates and did not originally include any variables that were not also included in the analysis model. We thank the reviewer for pointing out that auxiliary variables should have been added if imputing missing outcome, and have added two variables to this effect: social class and self-rated health. These variables are correlated with the outcome. We have added this detail into the methods section (line 180/181). Following changes to the number of imputations used (see point 1) and the addition of auxiliary variables we have updated Tables 1 and 2 in the main text accordingly. Very slight changes to the ORs and CIs reported in the text have also been made (to the results section of the abstract and main text lines 221). Changes to number of imputations and addition of auxiliary variables did not lead to any significant or notable changes in results. Note the larger change in physical activity/alcohol was because of a typographical error in these lines in the original (actual differences are low).

3) Multiple imputation has the advantage over complete case analysis of potentially increasing the MAR assumption if auxiliary variables are included, the authors have not mentioned including any auxiliary variables.

As noted above, we have now included auxiliary variables and have added text to the methods section (lines 180/181) to make this clear.

4) The number of multiply imputed data should be reported in the footnote of each relevant table

The analysis that used the multiply imputed data were based on 24 imputed datasets. This detail has now been added as a footnote to each relevant table (Tables 1 & 2).

5) The predicted probabilities used in the graphs are they based on one single imputed data? Or on complete case analysis or have the authors pooled the predicted probabilities obtained from each imputed data using Rubin's rule?
The predicted probabilities used in the graphs are based on complete case rather than multiply imputed data. Given that the results were similar using the imputed and the complete case data we used the complete case data to create the graphs. We have added text to the methods section (line 188) to make this clear.

We look forward to your further consideration of our manuscript.

Dr. S Vittal Katikireddy and Dr. Kathryn Skivington (on behalf of the authors)