Reviewer's report

Title: Individual socio-demographic factors and perceptions of the environment as determinants of inequalities in adolescent physical and psychological health: the Olympic Regeneration in East London (ORiEL) study

Version: 3
Date: 17 October 2014

Reviewer: James Bowes Kirkbride

Reviewer's report:

Thank you for giving me the opportunity to review your paper. I have structured my review according to the BMC headings for peer review and I hope my comments are helpful. All items in questions 1-10 should be considered major compulsory revisions which should be addressed or rebutted as you see justifiable! Items in Question 11 are mostly discretionary revisions for comment except for those marked with an asterisk, which are minor essential revisions.

1. Is the question posed by the authors well defined?

The study seeks to establish the baseline prevalence of physical activity, general health and mental health in a sample of young people (aged 11-12 years) who are defined as adolescents. The study investigates whether the prevalence or risk of these health outcomes varies by various sociodemographic factors, and perceptions of the social and physical environment.

The question is well defined, though the authors should take care in their language used. “Pathways” are mentioned in both the abstract and body of the text, but this study does not specifically look at pathways, since it is cross-sectional and not longitudinal. I think some caution would be warranted here.

2. Are the methods appropriate and well described?

The design of the study and methods used are generally well-described, but I have several comments arising from the analytical approach taken and the reporting of the results. These are as follows

A. The response rate is impressive but insufficient detail is given about the sampling strategy. I understand the full protocol has been published elsewhere but this manuscript should still present sufficient details to understand how the study was conducted. Were all children in a given year group (within the selected schools) invited to participate, or were students sampled? If the latter, how were they sampled?

B. Relatedly, the study clearly has a sampling strategy in which a selection of schools were randomly chosen within each borough (so there is some stratification of the sample too), and then, depending on the answer to point 2A,
further selection of students may have been employed. Given the sampling strategy and non-response why did the authors not derive and apply sampling weights to their estimates of prevalence and risk?

C. A central problem with the final analyses is the adoption of a complete case approach. For some analyses (physical activity, sedentary activity) the CC approach uses just 33% of the sample. For the remainder of analysis this is still only 50%. While you carefully describe the whole sample as generally representative of children of the same age in the catchment area, were there differences (i.e. biases) between the students in the CC analyses and those who were not? This would be important to know.

D. Could the authors apply multiple imputation techniques to enhance the methodological approach of their paper?

E. I have some concern about the methods chosen to analyse the data. Mixed effects models are appropriate, but I am unsure whether the authors used genuine random effects models to inspect the effect of school clustering on outcomes (i.e. the level of each outcome that may be attributable to school level effects vs individual level effects) or whether a simpler cluster approach was taken to adjust standard errors in the models for clustering on the school level? The former would be preferable and a stronger use of the data. If the latter approach was taken, and probably regardless of either approach used, some statements arising from their manuscript are unsound i.e. “During adolescence perceptions of the urban physical environment are as important as the home environment in explaining patterns of health inequality” – I think unless you have explicitly modelled the causal levels in your analysis (individual, home, school, neighbourhood) such statements are unsafe.

3. Are the data sound?

Please see my comments to Question 2, as there is some overlap here between the two sections. Further points arising:

A. While the presentation of univariable i.e. unadjusted prevalence estimates of health outcomes are somewhat interesting for the purposes of descriptive epidemiology, this leads to the presentation of several univariable tables, of which many of the results are not interesting, since the reader would really wish to know whether confounding by other variables influenced any significant associations. In this respect, Table 7 of the manuscript is the most salient and interesting part of the paper. I recommend the univariable associations may be shortened and de-emphasised and possibly included only as supplemental data.

B. The multivariable data presented in Table 7 is of interest, notwithstanding my points arising under Question 2. A further concern is the extent of correlation or confounding between health outcomes as presented in Table 8. I found the presentation of these associations in a separate table difficult to interpret and perhaps misleading. While you shown several significant correlations between health outcomes in Table 8, after adjustment for various other predictors, the
reader is unable to see the effect that controlling for these other health events has on the associations between sociodemographic predictors and the health outcome under study. The authors should present the full multivariable results in Table 7 so that, for example, the association between well-being and gender is already adjusted for depression, physical health and so on. Reading Table 7 alone makes it impossible to conclude whether the reported associations are simply due to confounding by other health outcomes or behaviours.

4. Do the figures appear to be genuine, i.e. without evidence of manipulation?

No figures are published in this manuscript but I have no doubts about the correct reporting of the analyses conducted by the authors.

5. Does the manuscript adhere to the relevant standards for reporting and data deposition?

Yes.

6. Are the discussion and conclusions well balanced and adequately supported by the data?

I have some concern that your discussion and conclusions go beyond the strict interpretation of cross-sectional data. Some tempering of your discussion would be welcome, and in particular for the conclusion, which makes unsafe statements regarding the relative importance of home vs neighbourhood and implications for policy. Other matters on this topic:

A. Line 321 – When you say evidence for socioeconomic inequalities is weak, please clarify whether you mean in general or in your study specifically

B. Line 331-332: You interpret your findings as evidence that neighbourhood perception is more salient than socioeconomic factors in determining adolescent health outcomes. But could it be reverse causation at play? In fact children with worse physical or mental health outcomes are more readily likely to perceive their neighbourhood as poor? This might apply more strongly to mental health outcomes and some comment on reverse causation is warranted.

C. Line 334 – There were surprisingly few ethnic differences. Could chance have explained those observed?

D. Line 351 – 354: Did you have any data on the onset of puberty? What was the rationale for describing 11-12 year olds as adolescents and not children?

E. Line 356-358: I don’t think the reference to the longitudinal study being a valid means of understanding causal processes given your cross-sectional data is valid until other issues are addressed.

F. Line 368: Please check whether reference 49 specifically included black Caribbean and black African groups in its “Afro-Caribbean” group – normally I thought this term was restricted to the black Caribbean group only, which might have an impact on your sentence about ORiEL advancing knowledge in this regard.
G. Line 382: Even longitudinal studies cannot “confirm” causality.

7. Are limitations of the work clearly stated?

Some limitations are addressed, but this section will need revisiting given the comments arising here.

8. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?

Yes

9. Do the title and abstract accurately convey what has been found?

Yes

10. Is the writing acceptable?

Yes

11. Other Issues arising

I have a few minor issues also arising for your attention
A. The abstract could be shortened by omitting the unadjusted results
B. In the abstract you state the survey was anonymous but I have reason to doubt this given the longitudinal nature of the follow up. Presumably you will be linking baseline responses to later follow-up outcomes, meaning the strict definition of anonymity is not correct here? If I am mistaken, I apologise!
C. Line 126: In the preceding section the ORiEL study is described as a longitudinal cohort study, but here it is described as a prospective controlled quasi-experiment. This might need some clarification to avoid confusing the reader.
D. *Line 128: Please state when (calendar year) the children were assessed in order to put it in the context of regeneration of East London
E. *Line 143: Individual and household level socioeconomic demographic characteristics are listed here as outcomes but used as exposures throughout. Please clarify this.
F. *Line 151-153: Attitudes to the Olympic Park – I did not see these reported in the results. If this is not part of the present study this section can be omitted.
G. Line 163-164: Do you mean “risk of depressive symptoms” or actually “depressive symptoms”? Could a stronger (i.e. more clinical) cutoff have been applied?
H. *Line 169-170: Is 75 hours of total activity or more really high? Assuming that corresponds to total waking time, that implies 13.5 hours of sleep per day, which seems a lot, even for adolescents!
I. Line 183 & throughout: “White UK” and “White Mixed” are not official 2011 Census categories. Please define these as White British and the specific mixed
ethnicities to which “white mixed” refers.

J. Line 239: Have there been recent influxes of white migrants from Europe (i.e. since 2011). I thought most of this migration was before 2011, meaning this would have been expected and captured in the last UK census

K. *Table 2 & Lines 248-249: How were univariable differences tested? Please present the test statistics used.

Good luck with your revisions – I hope you can address some of my major concerns to strengthen the quality of your findings. Best wishes.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests