Reviewer’s report

Title: Heavy Metal Exposure and Nasal Staphylococcus aureus Colonization: Analysis of the National Health and Nutrition Examination Survey (NHANES)

Version: 0 Date: 07 Sep 2017

Reviewer: Helen Crabbe

Reviewer's report:

General comments:

I welcomed the opportunity to review this manuscript. The subject matter is of increasing importance in the public health battle against anti-microbial resistance. The merging of current topical infectious disease issues and that of non-communicable exposures is most welcome. Further exploration of outputs and re-analysis of data from NHANES is always interesting and continues to add to our understanding of environmental contaminant exposures and effects on human health.

The paper is concisely written and explores the novel relationships between human blood lead and cadmium levels and nasal staphylococcus aureus colonization.

The authors assume a level of familiarity of NHANES and the nature of the purpose of the study and nature of the variables collected. The paper would be strengthened if more explanation of the background of the NHSANES survey, setting, cohort descriptives and population could be briefly added.

Good definitions of the variables and explanation of the models used to calculate risks. More detail of the variable selection and omission, both a-priori and post-priori is needed to explain and justify the model building process. Why were some variables left out on purpose, e.g. smoking in the Lead models?

Throughout the paper there are strong assertions made for a strong and consistent association found. This conclusion was made on the evidence presented that seems more like mixed findings and perhaps weak signal strength associations that are found after multiple testing and pre-defined models. Did the authors consider the effect of making multiple comparisons, that you would find 1/20 results as significant (type 1 error) and test the effects of adjusting the p value appropriately (e.g. Bonferroni correction or other)? Model choice and inclusion of variables also seemed to be very prescriptive. Did you conduct a sensitivity analysis to assess the effects of other variables and confounders?

The authors use frequent references to 'Pb/Cd exposure' when referring to the blood lead/cadmium levels measured in study participants. It is (probably) unknown what the level of
environmental exposure resulted in the blood level, or what 'exposure' was when the samples were collected. The blood concentration would reflect the body burden and intake/uptake of the metal, but not necessarily the concurrent environmental exposure. Some reference or explanation of this fact needs outlining, and perhaps reference to 'blood Cd/Pb levels' would be a more accurate description/label.

It would also be beneficial to add a description of the half-life of Pb and Cd in blood and some thoughts on when and how the metal exposure could have occurred; i.e. how and what the blood levels actually represent. Also further explanation of how colonisation of MRSA/MSSA occurs. Can carriage vary (come and go)? Once you acquire it does your status change? Do we know if it is recently required etc? Can you be colonised by both MRSA and MSSA at the same time, at different body sites, and/or swap status? This may help to explain the associations between the two (or is this a limitation of the study if unknown?).

The authors list a few limitations of their study, and it is an important first step to exploring the relationships, but there are other limitations that were not addressed, e.g. could there be alternative explanations for the associations found? Other (residual) confounding, interactions, biases or chance? What might these be? Was the analysis limited to the format of NHANES outputs? What other information would be useful to collect to explore these associations (e.g. current smoking status, SES, occupation, etc.)? What other information do we need to collect to further explore this issue?

On the whole an interesting study and good use of NHANES data. With some revision, I would recommend publication of the study in Environmental Health.

Specific comments on the sections include:

Abstract:
Sample size is missing.

The wording suggests cause and effect, e.g. 'both metals were protective against MSSA'. No causal relationships are proven, and the associations were often weak or not-significant. The wording around the strength of the associations in the whole manuscript could be toned down.

Line 26: Can the authors explain what 'co-select' means in this application?

Results- add sample size. Some sweeping statements are given regarding metal concentrations found and MRSA. On further examination these differences between the GMs are not statistically different, so either the language needs to reflect this or p values given.
'both metals were protective against MSSA': This assumes causation. A dose response relationship was shown, but with adjustment for confounders, there is little difference in the ORs of the quartiles (CIs overlap).

Conclusions: the authors refer to 'current' population exposure when in fact the samples were taken 13-16 years ago and environmental exposures and indeed blood Pb/Cd levels may have changed during this time. How do the authors know that the metal exposures and colonisations was concurrent?

Background:

Line 105: Explain what 'co-select' means in this context.

Line 112: could the authors explain (somewhere) what literature searches they undertook to make this statement (search strategy or evidence)? This could be outlined as a strength of the study in the discussion.

Line 116: A 'health-care' acquired infection? Not just hospitals?

Line 123- please add a reference to this sentence or assertion.

Line 133-135: This is a causation argument and authors suggest a sequence of events that exposures causes colonization. However, the authors later on in the paper suggest that the order is unknown (fig1), so this is not consistent.

Line 137: Refers direct to metal exposure and associations with colonisation when in fact the relationship is between blood lead/cadmium levels measured and concurrent colonisation ( see comment before). We assume both samples were taken at the same time?

Line 138: define adults.

Methods:

Data sources: How was the data extracted, by whom and how? Can you give more details about NHANES as some readers maybe unfamiliar with this survey? How were the participants selected and recruited? More detail about when and how the biological samples were taken would help (.e.g. concurrent - on the same day/clinic visit?). The reference refers to an analytical guide (33), not necessarily giving details of the background to the survey and how the sample population / cohort was selected and recruited. Are they representative? What were the waves of the survey - does this matter?

Line 150: Table 1 suggests age was 0+.
Line 151-2: Was ethics required? If publically available can you give a reference for this (url?)

Line 165-167: Is this a standard assumption made for colonisation? Can this be referenced?

Line 176: Why was smoking not included in the lead model? Lead is known to be present in tobacco smoke and cigarettes. What was the method of the choice of variable to include in the models and why were some variables pre-defined and left out of models? Did the authors consider backwards or forwards stepwise model building? Did this have an effect on the risk estimates?

Line 177: Was it possible to separate current smokers and ex-smokers, as this could explain both the exposure and outcome variables? If this data was not available, it should be discussed as a limitation of the study?

Diet data: Line 179 onwards. It is not clear how the different diet data were calculated for the different metals. Can more explanation of what was included for each model be given? Was green leafy vegetables separated or used as a variable? What differences were considered? Did this make a material difference?- line 188- can you give a reference for the food codes?

Statistical analysis: Line 199- explain what is meant by a DAG? How did this affect variable choice? It seems that only a-priori confounders were considered in the models. What was the effect of other variables? Were the effects of other confounding and effect modification considered? Was a sensitivity analyses undertaken? How was the choice of different dietary options optimised for each model?

Results.

Descriptive statistics need to include some background to the population cohort, numbers sampled, how recruited and how representative the sample is to the population. Are the differences in colonisation rates between the groups real? Do the CIs overlap? Can you provide some statistical test of real differences between groups?

What levels of Pb and Cd were found? Mean, min, max, GM etc? What were the levels of the quartiles? How do these levels compare to other populations? Are the results transferable to other populations?

Table 1: Define the years in the title of the table. Define the % given in the headings (i.e. row %ages?). It would be good to see the breakdown of the sample into the categories, i.e. by adding a totals column and add % of the columns (e.g. % of the sample in each age group, % of males to females, % ever to never smoked, etc). There are discrepancies with the 0-17 age group with the text of a minimum age of 1 year.

Table 3: What are the units of these measurements (grams?)? Use label: fruit and vegetables.
Line 231: 'Exposure to Cd is associated with decreased odds of MSSA carriage…' but model 3 (Table 5) shows that by adding fruit and vegetable consumption, then the OR are not significantly different from 1. Can you explain this?

Table 4: what was the effect of adding smoking to the Pb models?

Discussion.

Some of the assertions made assume that the differences found are real. Is there any unexplained residual confounding, is the result due to biases or chance? These points should be discussed and if not addressed listed as limitations of the study.

Line 240: I do not wholly agree with the statement that increasing Cd exposure was associated with MRSA colonisation. In the fully adjusted model (3) in Table 4 the OR do not increase and are only significantly reduced for the highest quartile.

Line 245 onwards- what searches were carried out and on what databases to make this assertion? How complete were the literature searches?

Line 264: 'over time'.

Line 265- can the pattern be explained? (relates to point above)

Line 273-275. What Cd levels does this relate to? What is high? How is green leafy vegetables measured in this study? Would it not be the case that both Cd and Pb be a factor via this route? Could you explore the datasets in this study to test this?

Line 284-285. Perhaps weak and not consistent evidence is presented. Many OR were not statistically significant and there may have been multiple testing issues and residual confounding not explained. There seems to be some pattern that suggests a possible association.

Line 293- suggest being specific to 'nasal' colonisation, as colonisation of other parts of the body have not been explored.

What other limitations apply to the study? Could anything else explain the associations found? Do you think there is residual confounding or interactions? Was current smoking status explored, or the effects of household second hand smoke, occupation, etc?

Line 307- can you please explain what is meant by 'reverse causality' in this context?

Line 312: Explain MDROs. Also add to list of abbreviations - line 325.
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:

Acceptable

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