Reviewer’s report

Title: Arsenic Exposure and Risk of Non-Alcoholic Fatty Liver Disease (NAFLD) Among U.S. Adolescents and Adults: An Association Modified by Race/Ethnicity, NHANES 2005-2014

Version: 0 Date: 29 Sep 2017

Reviewer: Andres Cardenas

Reviewer's report:

This is a well written manuscript with the underlying hypothesis that exposure to inorganic arsenic is associated with non-alcoholic fatty liver disease and that race/ethnicity modifies the effect. The authors used a cross-sectional biomonitoring survey design, NHANES, to assess levels of urinary arsenic and alanine aminotransferase as a biomarker of non-alcoholic fatty liver disease. This study has important strengths that include a large representative sample from the U.S. population as well as detailed ascertainment of exposure, biomarker of disease and covariates. Some important limitations are inherited in the cross-sectional design and therefore the interpretation of results must be scaled to match the design. In addition, I have some concerns about the analytical and methodological approach and I think sensitivity analyses are in order.

The discussion is brief and the introduction could be strengthened. I have series of comments and recommendations that should be addressed prior to considering this manuscript for publication:

1. Is there any underlying hypotheses (genetics, environmental or cultural) that explain racial differences in non-alcoholic fatty liver disease? A stronger argument must be made for testing effect modification by race in the introduction, as it is a premise for the study. As of now it appears that the given that only one race was deemed significant it was included post hoc in the title/hypothesis.

2. I'm not an expert in non-alcoholic fatty liver disease but the authors state that it occurs almost exclusively in obese individuals (line 88). In light of this, testing for the association of iAs and ALT in the obese category is in order might help support their findings.

3. Include rationale for excluding Hepatitis B and C in the body of the manuscript. What about hepatitis A? It has also been shown that Hepatitis A serology is associated with iAs exposure in NHANES (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4855991/) this might be an important confounder.

4. A categorical indicator for survey cycle should be included in analyses to control for temporal trends in both iAs exposure and prevalence of non-alcoholic fatty liver disease (or ALT) in the U.S. population. Temporal changes in both exposure and disease biomarker by survey cycle might be a very important confounder of the association.
5. After combining multiple cycles, were the survey weights scaled to match this approach? If not, this needs to be done in order to accurately capture the sampling design as explained here: https://www.cdc.gov/nchs/tutorials/nhanes/surveydesign/weighting/task2.htm. If this was already done please state in the methods.

6. In the methods please provide the % of samples below the LOD and discuss how this might influence the results. The change in LOD by survey cycle supports the inclusion of cycle as a covariate in multivariate adjusted models.

7. The authors should briefly state whether the ALT cutoffs used to classify individual are clinically used or used in research?

8. I couldn't find anywhere the number of individuals both total in the study and by race that had elevated ALT. This number must be provided. Given the small sample size for some races/quartile combinations I'm concerned that this might lead to some spurious associations. Could be beneficial to include in table 4 the number of "cases" or positive individuals for high ALT.

9. Along the same lines, when stratified analyses have small sample sizes the survey weights tend to yield unstable results. I encourage the authors to run non-weighted sensitivity analyses and report in a supplement. In general results should be robust in direction and association but the magnitude might differ.

10. I'm not sure what the rationale is for stratifying by quartiles of exposure, it would very beneficial to see results by continuous iAs exposure measurement. If a dose-response is relevant for ALT why didn't the authors model exposure and ALT as continuous? Or even continuous iAs exposure and elevated ALT in logistic models?

11. It would be very valuable to present joint analyses (all races together). Along the same lines if effect modification is hypothesized I would encourage the authors to conduct a formal interaction test in a model that includes all the races (i.e. continuous iAs exposure and race interaction) and report the F-test p-value. I assume that the borderline test for effect modification by race (page 8) is by categories of exposure and elevated ALT? If so, please clarify.

12. Please refrain from using the word "risk" in the abstract, conclusion and results as the design is cross-sectional and thus one can't infer this relationship.

13. Please consider changing the title as non-alcoholic fatty liver disease was not objectively measured. However, if there's enough clinical evidence that ALT is a good surrogate that this might be OK. Otherwise replacing non-alcoholic fatty liver disease with ALT works and better reflects the scope of this work.
Minor:

1. In the title, non-alcoholic fatty liver disease includes a hyphen, but not in the manuscript. Please make sure there's consistency throughout the manuscript.

2. Page 6, line 131-132. "These measurements reflect approximately the last three days of exposure", please add reference. Or explain how was this estimated.

3. Page 8, line 181: "Due to race and ethnicity differences in iAs and ALT, we ran the fully adjusted model, presenting stratified ORs..." It is not clear if estimates in Table 4 are from one full model and then predicted ORs are presented from that model or multiple similar models stratified by each race. Please clarify if this table are from predicted ORs or truly stratified by race (i.e. multiple regression models).

4. Urine dilution adjustment methods appear under covariates on page 7 and in the methods for iAs exposure on page 6. The wording here is a little confusing please consider consolidating in one place. Was urinary creatinine included as a covariate in models or by dividing exposure/creatinine?

Level of interest
Please indicate how interesting you found the manuscript:

An article of importance in its field

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