Reviewer's report

Title: Urinary Mycoestrogens and Age and Height at Menarche in New Jersey Girls

Version: 0 Date: 13 Nov 2018

Reviewer: Mary Wolff

Reviewer's report:

This very interesting study is quite suitable for publication in EH. It is well written, and the Discussion particularly is excellent. There are a few questions and clarifications. First it is confusing, how the exposure biomarkers are handled, when most are <LOD. Second also confusing are the mixed models and what variables are used, in particular whether they include multiple hts and wts from enrolment to menarche. The way it reads now, in Table 3 and the text, a single weight/height is the outcome. Indeed, the authors undertake an analysis that considers BMI in/out of the pathway, which they justify correctly. But there doesn't appear to be much BMI-mediation here. Regarding mediation and the causal pathway, our puberty studies also examined growth trajectory in relation to several exposures of this kind, with findings that are generally consistent with BMI mediation (phthalates, phenols, elemental lead). Third, the Z urinary concentrations could be put in context with exposures seen in other populations, if any. However, a serious limitation is that only unconjugated (free) Z metabolites were measured. Most studies measure total or some combination of free/bound, and most urinary metabolites of this type are mainly conjugated. The laboratory method did not include the deconjugation step that is done in such studies. This is addressed in the Discussion, but is clearly a major limitation, and accounts for so many of the levels being LOD. It is perhaps likely that the free concentrations are highly correlated with the total, but this is my speculation, though it might be supported by the literature, experimental or human. The Z's are somewhat homologous to some of the other phytos, so maybe analogy can be used to extend this argument. Finally, it is really a stretch to interpret an effect from "(adjusted HR: 0.35; 95% CI: 0.06, 2.00)." The only thing that might be reasonable would be comparison of results to a similar analysis for B2 in the earlier paper, if the estimate is similar to that. Otherwise, perhaps the BMI effects may be responsible for the imprecise estimates in Table 2 for menarche. And the small numbers can flip-flop the models. Rather than overinterpret, rely on the BMI findings which are interesting and consistent with the literature. Either a larger cohort or better exposure measures might improve the models.

Abstr should give some indication of urinary concs or %detect.

HR for menarche - mediation?

Li 88: girls with negative levels - revise "negative"

Li 119 The number of follow-up questionnaires completed ranged from 2-7 with most girls having 4- -provide a number or % for 'most'.

Li 121 - give N for cohort.
Li 122 - Note these were included as right-censored.

Li 133 - Note if external standards only, i.e. no internal standards used.

Li 136 - edit? Grouped = summed concentrations or quantile values?

Li 139: generally if levels are <50% or 60% or 70% < LOD (depending on study), continuous concentrations are not used (Lubin and others). This would affect medians in Table 1, where most investigators do not report univariate data for LODs, except as 'LOD'. LOD should be provided in a note. A table of urinary Zearalonine etc levels should be given.

Li 147 - note how time-to-menarche was treated for right-censored obs.

Li 157 - "matrix were used to assess height and weight z-scores in relation to…" It is not clear if this is a single ht/wt measure or multiple longitudinal measures; perhaps specify the model exactly (y=x). Also, where this is mentioned at li 197, not clear what variables are in the mixed effect models.

Regarding dichotomous Zea, as it is 55% detected, an alternate categorization would be LOD (A), lod-midpoint of detected (B), and > (B) = (C), to provide a little more power and an indication of linearity. Or perhaps as a sensitivity analysis, as (C) might be similar to the small > LOD zeranol group. Alternatively, if "machine" values are available for the LODs, those can be used. There are a number of refs to justify this approach, and it would be further supported by the categorical exposure models.

Li 184: shorter at both enrolment and menarche?

Longer time to menarche: tempo or pace has been defined by others as the time between B2 and menarche; note whether would that have changed these observations.

Li 198: does average weight and height z-scores = average weight- and height- z-scores

Li 199: Zeranol levels - note if this is detect vs non-detect.

Disc li 209 This may mean to say zeranol, rather than ZEA? ("Girls with detectable ZEA concentrations were also less likely to have early menarche, but results were 210 not statistically significant")

li 281, regarding single urine sample for biomarker: in other studies some but not all biomarkers of this kind have been found to be consistent over wks to years (i.e. ICC > .5), so if the Z's come from common, usual exposures a single sample may be expected to be fairly reliable.

Table 1. see notes on handling of LOD values

Table 2. Indicate concentrations of LOD/< LOD
Table 3. Indicate exposure concentrations. If these are ng/ml, then the difficulty of imputing <LOD makes this model problematic. If it is <>LOD, then note if exposure is coded 0,1. Not clear what the note means (Note: n=163 girls, observations used in mixed models= 2,021 (from 2,507).) Are there multiple ht/wt in these models? See also same question above. Describe the model more detail.

**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:

Needs some language corrections before being published

**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