Reviewer’s report

Title: Hair Product Use, Age at Menarche and Mammographic Breast Density in Multiethnic Urban Women

Version: 0 Date: 14 Aug 2017

Reviewer: Kimberly Bertrand

Reviewer's report:

General comments:

This is a well-written report that is clear and organized and addresses an important hypothesis regarding concerns about early life EDC exposures and predictors of breast cancer risk. The statistical methods applied are appropriate given the study design. The potential limitations of the study (e.g., relatively small sample size and necessity of collapsing some specific exposure categories into a broad group of "other") are adequately discussed and well-balanced by the strengths of the study, including careful exposure and outcome assessment and adjustment for important confounding factors. While I believe the authors' final interpretations are generally supported by the results, I suggest that it would be more appropriate to put the primary emphasis on the race-adjusted results for age at menarche (e.g., in the abstract) rather than the minimally-adjusted results, given that race/ethnicity appears to be an important confounder in these analyses. Below, I also request some additional details regarding the methods (e.g., timing of exposure relative to age at menarche) that I believe will improve the clarity of this manuscript. Other minor specific comments are also offered for the authors' consideration.

Major specific comments:

* Page 3, lines 31-51: Can you please clarify whether all women included in the analysis had no history of breast cancer at the time of interview/mammogram?

* Page 5, lines 5-6: Race/ethnicity is not mentioned as a potential confounder of the association of childhood hair product use and age at menarche in the Methods, although it appears a model including race/ethnicity was fit. Please included these details in the Methods section.

* It is not clear whether "childhood use" of hair products refers to any use before age 13, or any use before menarche. This introduces a concern about temporality of associations: some girls may have been <11 years old at menarche, but started using hair products between ages 11 and 13. Please clarify how the exposure variable was defined.
Page 5, lines 40-45/Table 2: Race/ethnicity seems to meet the definition of a confounder in analyses of childhood hair product use and age at menarche, given that African American are more likely than white girls to use hair oils and other products and African American girls also have earlier menarche on average. In fact, after adjusting for race/ethnicity, the observed positive associations become attenuated. I strongly urge the authors to consider reporting the race-adjusted results as their primary results in the Abstract (and in Table 2 and Conclusions). It is helpful to see the results before adjustment for race/ethnicity, but race/ethnicity appears to be an important confounder here and relevant for interpretation of findings. In fact, the association for hair oils specifically remains borderline statistically significant (95% CI: 1.0, 5.5), even with relatively low power - this is an important finding and still supports the a priori hypothesis.

Minor specific comments:

* Title: "Breast density" could imply breast tissue density or mammographic density. Since breast density was assessed on mammograms (vs. MRI or DXA, for example), it would be more accurate to call this "mammographic breast density," at least in the title and the first time it is defined.

* Abstract, line 29: There is a typographical error: early age a menarche is given as ≥11 years, but I think it should be <11 years based on the Methods and Table 2.

* Page 2, line 60: If possible, it might be helpful to give specific estimates for how commonly these products are used in the general population (e.g., percent of women, preferably by race/ethnicity) vs. simply noting that use is "frequent and long-term."

* Page 3, line 49: The final sample size is given as 248, but this is confusing because on the next page, line 27, a larger sample size is given. At the end of that paragraph, it becomes clearer how the final sample size was calculated. I suggest excluding the first mention of the sample size from the first paragraph of the Methods section.

* Page 3, line 56: I'm a little confused about the assessment of timing of hair product use during childhood. Here, it seems that women were queried about use "before age 13," but based on information on Page 4, lines 15-16, it was also possible to calculate duration of childhood use. Were women asked about use at each age during childhood? Was any information collected about use between the ages of 13 and 20?

* Page 4, lines 28-31: What was the timing of women's mammograms relative to interviews?

* Page 4, lines 42-44: Was information on menopausal status or early life body fatness (e.g., childhood BMI) available? If menopausal status is available, it might be interesting to
evaluate associations in premenopausal women only: one might expect a stronger effect in this group if the hypothesis is related to earlier life exposures. Childhood adiposity is a predictor of both age at menarche and breast density; could it be an uncontrolled confounder?

* Page 4, line 56: Is "multivariable relative risk regression using the binomial link" simply "logistic regression"? For the sensitivity analysis modeling age at menarche as a continuous variable, was this linear regression? Was age at menarche normally distributed?

* Page 5, lines 9-13: Were percent mammographic density and absolute mammographic density both normally distributed? If not, these variables (especially absolute density) may require transformation (e.g., log-transformation or square-root transformation) to meet the assumptions of the linear regression model.

* Page 6, lines 4-6: I think readers would find the actual results of the linear regression model for age at menarche of interest; please consider including the results here in the text (rather than indicating "data not shown").

* Page 6, lines 15-17 and Table 3: "After adjusting for confounders ...." Why isn't race/ethnicity included in the multivariable model? Some recent research has suggested differences in percent mammographic density (adjusted for BMI) by race.

* Page 6, lines 20-22: "... adult hair dye use was associated with higher percent density (beta 3.5; 95% CI: 0.6, 6.4) ..." I respectfully disagree with this interpretation: given the wide confidence interval and the small beta, I don't think this result suggests strong evidence of a positive association. In fact, differences in percent mammographic density of less than 5 percentage points are not generally considered meaningful with respect to breast cancer risk. Similarly, in the Discussion (page 7, line 9), these results are suggested to be "consistent" with the literature showing positive associations between hair product use and breast cancer risk. I disagree; these generally null results for mammographic breast density suggest that the pathway from hair product use to breast cancer risk, if causal, may not operate through mammographic density.

* Page 6, lines 27-29: Readers may be interested in seeing the results for duration of hair dye use in a Supplementary table, if possible.

* Page 7, line 27: "Moreover …" There is a typographical error in this sentence.

* Please be consistent in categorizing variables across tables and throughout the text. For example, age at menarche is dichotomized at <12 vs. 12+ in Table 1, but <11 vs. ≥11 in Table 2. It would be more helpful to see the distribution in Table 1 that corresponds to the way the variables were handled in the analysis.

* Table 1, footnote: Does the p-value represent the test for differences between cohorts?
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