Author's response to reviews

Title: Perfluoroalkyl acids and time to pregnancy revisited: An update from the Danish National Birth Cohort

Authors:

Cathrine C Bach (ccbach@clin.au.dk)
Zeyan Liew (zeyanliew@gmail.com)
Bodil H Bech (bhb@ph.au.dk)
Ellen A Nohr (eanohr@health.sdu.dk)
Chunyuan Fei (chunyuanfei@gmail.com)
Eva C Bonefeld-Jorgensen (ebj@ph.au.dk)
Tine B Henriksen (tbh@dialnet.dk)
Jørn Olsen (jo@ph.au.dk)

Version: 3 Date: 30 April 2015

Author's response to reviews: see over
Response to reviewers’ comments

Responses are listed below the reviewers’ comments in italics.

Referee 1

Reviewer: Alan Ducatman

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 I have no competing interests

Major compulsory revisions include a justification of or change in the methods, as well as increased clarity of tables.

1. The authors have posed a discrete question about time to pregnancy. The question is not new. It is useful, and it deserves the several studies which have been directed towards answering it. Therefore, this effort is potentially important. And, the authors deserve credit for willingness to look at and be willing to overturn their previous findings.

Thank you. After our first report in 2009 other studies have had contradicting findings. We have chosen to revisit this topic because we have new data and we agree that the topic is of great importance.

2. However, the approach and conclusions appear to rely on a specific use of comparison between stratifications that is already known to be methodologically problematic. The assumption that time to first pregnancy is more valuable than time to subsequent pregnancy is just that, an assumption. It ignores the primary cause of the gender difference in serum PFAs concentrations, which is thought to be menstrual blood loss. The exposure persists through each subsequent pregnancy. (The authors are correct that PFAs transfer to the fetus and placenta also occurs, but the gender difference begins with the onset of menstruation and persists through the cessation of menses.

Articles by Arbuckle et al (Pediatric Perinatal Epidemiology 2013 and Int J Hy Environ Health 2013) address this problem and also outline more appropriate methodological
approaches. Schisterman et al (Epidemiology 2009) and Greenland (Epidemiology 2003) address the importance of not drawing strong conclusions from such heavily adjusted/stratified data (and then there is the influence of small numbers. So that the authors are clear that this criticism is not about a preference concerning the actual findings, I stipulate that the conclusions reached could easily be correct. New associations are often found to be non-causal. Reverse causation is a simple and therefore attractive hypothesis for all or some of the explanation. The concern is not about the conclusion, it is that the methods and the data do not address the hypothesis in the conclusion very well.

Parity has been shown to be a strong predictor of levels of PFAAs in women (see Brantsæter et al. 2013 in Environ Int, Ode et al. 2013 in Environ Sci Pollut Res, and Berg et al. 2014 in Environ Int). Parity is also associated with the TTP through the common causes underlying fecundability or genetics determining the fecundability (see Additional Figure 1). Even though it is our opinion that it is crucial to stratify on parity, and that perhaps results from parous women are not even valid, we have added estimates that are not adjusted for parity to an additional table (Additional Table 5) so that the readers are now able to consider both sides of this controversial issue.

Since we only studied women (not men) we do not agree that menstrual blood loss needs to be considered. Furthermore, the women studied were of rather similar age, and we adjusted for age, which would be a valid proxy of the degree of accumulated menstrual blood loss. We agree that the small cell size in stratified samples is problematic, but we also believe that we have good reasons to do the stratified analysis. We stated in the beginning of the Discussion that we are aware of potential random findings due to the relatively small sample size, which particularly pertain to the stratified analyses.

Less important critiques are below. Many of these are still substantial, but they are not as fundamental as the approach used.

3. In addition to the major critique (above) several details of the methods could also be improved from the perspective of supporting the conclusion that the association of PFAA concentrations with time to pregnancy is due to confounding and reverse causation.

A). The statement “For the pooled analysis of the two samples we additionally adjusted for the sample (two categories)” is ambiguous. The authors are probably describing some form of
normalization. The results that relate to their conclusion depend on this step, thus a description exactly what was done is essential. The explanation that samples done by the different laboratories “would generally end up in the same quartile” is not a replacement for transparent statistical methods. This explanation can be dropped in favor of a complete description of how a normalization was done for the two subpopulations with their different test labs and systematically different serum concentrations.

Thank you for pointing out this ambiguity. We have now replaced “(two categories)” with “(by use of a dummy variable assigning each of the two samples with a different value)” We generated quartiles for each of the two samples. For the small group of participants overlapping between the two samples (analyzed at different laboratories) we compared each individual’s quartile value for the first laboratory to the second. Thus, by “would generally end up in the same quartile” we mean that almost all participants were assigned to the same quartile regardless of which laboratory the sample was analyzed at. The few women who would have been categorized differently by the other sample’s quartiles would have been assigned to a neighboring quartile.

B) The reader cannot be certain that the time frames of the two samples are identical. For the authors’ conclusions, it is essential that serum concentrations be demonstrably unrelated to any independent temporal trends in serum concentrations. It is implied that this was one effort that somehow used two labs. That is fine, but given the difference in laboratory testing, the reader deserves careful reassurance that the sampling time frame is truly at or near identical. If that simplicity is not in place, then the authors face a more complex job in normalizing.

Thank you. The two samples were sampled during the same time period, and we added (first part of the Results section): “...” as well as the year of inclusion were comparable between the two samples”. The only difference is the time of retrieval from the freezer and analyses - but these compounds are shown to be stable in the storage as the reviewer indicated below.

The sensitivity analysis is only about the data that has not been used. There is no a priori reason to hypothesize that those with unknown time-to-pregnancy would fall in a particular quartile (and the authors have extensively proved that). This analysis can be omitted from the paper. If this analysis is included, it is a question whether it should be called a “sensitivity analysis. “ The exclusion is not related to the data used, nor to the relevant conclusion.
We agree that it may seem misleading to call this analysis a sensitivity analysis, and we renamed it a bias analysis, since the purpose with it is to examine the impact of excluding women with missing TTP in our study population (a potential source of selection bias). We respectfully disagree that there is no a priori reason to believe that women with unknown time to pregnancy (most of whom did not plan their pregnancy) would fall in a particular exposure group. On the contrary, many of the women are likely to have conceived “by accident” and may in fact have a high biological fecundity (and therefore a short time to pregnancy). If there exists a causal relationship between PFAA exposure and time to pregnancy it is plausible that these women have lower average exposure levels compared to women who provided a time to pregnancy, which was also the case in our study. This analysis relies on very extreme assumptions that all missing TTPs either belonged in the high or the low group and is therefore only included in order to investigate if results deviate in a particular direction away from the main results.

There are conventional adjustments that could attenuate the existing relationship, and can be considered as part of the primary analysis or as a sensitivity analysis. Do the authors have access to a pre-pregnancy eGFR? If so, this is a potential important adjustment which could support their hypothesis

We do not have measures of eGFR, but we are not sure whether eGFR affects the time to pregnancy.

Body habitus measures (BMI) – are they pre- or during-pregnancy? (Obviously a prepregnancy measure is preferred. The point is, whatever was done should be clear. If there are weaknesses, they should be stated.)

These are pre-pregnancy measures (mentioned on line 108).

Diabetes affects eGFR and diabetes affects fecundity. Is diabetes one of the 3% exclusions? No, the 3 % exclusions were due to missing covariate information only (BMI, socio-economic status, parity, or maternal age).

The authors discuss oversampling male births as reflective of the general population gender ratio. What is the birth gender ratio and what degree of male oversampling was done?
We oversampled mothers who gave birth to boys in a ratio of 4:1 since the offspring of the mothers were part of a case-cohort study of developmental disorders in childhood that are more common in boys. The general population birth gender ratio is close to one.

One method that might be questioned is not a problem. Current literature supports the authors’ belief that PFAA serum concentrations should be robust to ex vivo changes during sampling. If there are differences between the two groups of samples, the authors are correct to look for other sources.

4. And 5. The reader does not doubt the accuracy of the data, given the approach taken. There is no evidence of manipulation. The methods questions raised are more about the adequacy of the approach taken as noted above, and whether the results justify the conclusion reached, as noted below.

A major concern is how well the conclusions relate to the information also presented.
6. The conclusions drawn from the tables are also of concern. The conclusions rely on a single data comparison which the authors regard as crucial, a sort of ratio between how well the findings persist in previously parous versus previously nulliparous women in two groups containing each, and then the two groups combined. The tables presented are not very transparent and have small numbers. Yet, to the degree that we can see the data in the relevant tables, the relationship that the authors reject actually appears to be present in 3 of 4 cases.

Further, the authors have assumed that any difference in time-to-first and time-to-subsequent pregnancy means that the findings are due to reverse causation, because the association in parous women is more likely to be spurious. The epidemiologic problems with that assumption are listed above. The stratification has an apparently inevitable null bias. In addition, there are also physiologic problems that lead to the reader to question - Why is that assumption physiologically true and important? The important gender difference in serum concentration begins in early adolescence in many studies, and then the gender difference increases for a few years and remains in place during the years of female reproduction. Using the logic presented, we could rule out any mechanical cause including menstruation for the gender differences, as each menstrual cycle attenuates the relationship between lifelong exposure and serum concentration. That would be a
decidedly wrong thing to do, and, it follows the same logic. I read and reread the authors’ explanation for the assumption, and think it exceeds our knowledge base.

The author’s conclusion is the simpler guess, and is therefore reasonably likely to be right. The problem for this reviewer is that the reader is left with problems in these areas: how or why the underlying assumption is supported; how the methods are epidemiologically justified; nor how the strong conclusion relates to the data.

- The data in Tables 2 and 3 fail to explain why the findings are reverse causation.
- The data in table 4 is not very relevant to the question.
- Table 5, could be helpful if it had combined the subgroup participants who completed all questionnaires. As constructed, it is not useful to the conclusion, because it provides data internal to one of the groups and does not allow for a full comparison of the most relevant groups. The more defensible summary data from the combination of the two would include the subjects from the two groups who finished all the questionnaires. That summary would have a smaller but more appropriate combined n. It would not address the intractable epidemiologic problems of stratification by previous parity, but at least the combination would be more defensible than the way the groups are combined in the tables currently presented. If the authors then wanted to add a sensitivity test that was related to their data, they could create an additional appendix table that further added back in the survey participants who were appropriately excluded because they completed only one questionnaire. That group has PFAA outcome data, so there is a plausible reason to include them in a sensitivity test. (Again, there is no reason to do a sensitivity test on a group for whom there is no data.)

Finally, my inability to follow the logic of the methods and the meaning that the authors derive from the tables suggests an alternative interpretation of this review that may appeal to the authors. It is possible that this reviewer, with no formal epidemiologic training, has missed something fundamental that unites the several apparent problems in a viable solution. I have mentioned this possibility to the editors, and suggested the kind of highly skilled methodologist who can decide.

The tables may have been unclear as the outcome measures were only mentioned in the table headlines. Therefore we now added the terms FR (fecundability ratio) and OR (odds ratio for infertility) as column headlines. As stated in the Methods section, the fecundability ratio (FR)
denotes the probability of conceiving in a given interval in a group of women with higher exposure compared to the reference group, conditionally on not having conceived in the previous period. Thus, a FR below 1 indicates impaired fecundability as measured by a longer TTP. This was also added as a footnote to Table 2, and a definition of infertility OR was added to Additional Table 2.

The analyses for Additional Table 5 were done in order to investigate whether the different sampling strategies used for the two samples could have accounted for the differing results. The main difference between the source populations for the two samples was that Sample 1 was sampled from women who completed the first interview independently of their completion of the following three interviews, and Sample 2 was sampled from women that completed all four interviews. The analysis was therefore an analysis of the women from Sample 1 who participated in all four interviews (see Methods section). If the different sampling strategies were responsible for the observed differences in the estimates for the two samples, we would have expected that these restricted results would have been more similar to those for Sample 2, but they were not.

Our key message is that we did not corroborate our previous findings of an association between exposure to PFOA or PFOS and time to pregnancy in a new sample from the same cohort, and that we therefore believe the question regarding these exposures and TTP is not resolved and should be further investigated. The aim of our paper was not to determine that reverse causality is responsible for an association between PFAS levels and time to pregnancy, this is just one of many potential explanations for these findings.
Referee 2

Title: Perfluoroalkyl acids and time to pregnancy revisited: An update from the Danish National Birth Cohort

Version: 2

Date: 6 April 2015

Reviewer: Matthew Longnecker

The authors address an important topic. Most of my comments are minor, though two comments about L 111 and one comment about table 3 (see below) indicate additional analyses would strengthen the piece.

Abstract, Results, 1st sentence: would move “in Sample 1” to the beginning of the sentence
Corrected.

Abstract, Results, 2nd sentence: instead of this being a sentence, would begin by inserting an “and” (this will help make it clear that the sentence is referring to Sample 1.
Corrected.

Abstract, Results, 3rd sentence: would replace “did not change after new adjustment” with “were again seen”
Corrected.

Abstract, Results, last sentence: would move “in the pooled analyses” to the beginning of the sentence, and would add a comment on whether the associations (13-22%) were in the parous, nulliparous, or both.
Corrected.

Line numbers from here on are those appearing in the draft reviewed, in the left margin

L 2: Please give an indication of how common and add references
This was added after the sentence as suggested.

L 3: would move “and lifestyle” so that it comes just before “factors” (and would make lifestyle singular)
Corrected.

L 7-10: near the beginning of L 8 (after citations), would add a comma, and insert “they are resistant to degradation in the environment and accumulate in the human organism [7].” (note the similar phrase in L 9-10 can then be deleted).
Corrected.

L 8: would begin new sentence before “these compounds” by inserting “Furthermore”
Corrected.

L 8: would insert “some” before “imported” and add a reference
Corrected and reference added.

L 8: would insert “being” before “replaced”
Corrected.

L 31: would replace “and” with “;”
Corrected.

L 33: would delete “accepted to” and add a “d” at the end of “participated” L 45-48: this sentence can go without harm
Corrected.

L 62: would insert “at ambient temperature” after “by mail”
Corrected.

L 62: hemolysis could have occurred during shipping, correct? If so, it would be a good idea to verify that when this occurred it caused no effect on the measured plasma concentrations of PFAA.
This is an interesting consideration. Hemolysis might have occurred during initial transport from the sampling sites (general practitioners) to the Statens Serum Institut, and we did not compare levels between samples that were hemolyzed and samples that were not. However, the laboratory staff centrifuged the samples prior to analyses, and we expect proteins to be removed and do not
expect any difference based on this. Samples for both Sample 1 and 2 were separated at the Statens Serum Institut shortly after blood sampling. The only differences between the two samples in the handling of the plasma used for PFAA analysis was the duration of storage after separation, the shipping duration and the differences in the two laboratories that performed PFAA analyses which we believe is unlikely to account for any differences between the two samples.

L 71-72: Would replace this sentence with “For Sample 2, the analytic results for five samples were not usable and were thus excluded.”
Corrected.

L 80: would insert an “e” at the end of “on”
Corrected.

L 86: would replace “fixed” with “response”
Corrected.

L 100: would indicate here if the transformation was natural or base 10.
Corrected.

L 111: as part of the pooled analysis, it would be a good idea to evaluate whether the association of PFAA and TTP was homogeneous across sample 1 and sample 2.
There were no statistically significant interactions between sample number and the exposures.

L 111: whether to adjust for parity is a controversial issue. If PFAS are related to TTP, then PFAS could also have a causal influence on parity. See Weinberg CR. Toward a clearer definition of confounding. Am J Epidemiol 1993;137(1):1-8. Best approach would be to comment on the effect of not adjusting for parity as part of the sensitivity analysis.
We added estimates not adjusted for parity (adjusted for BMI, age and socio-economic status) to Additional Table 5. We also inserted a small section concerning this in the Methods and Results sections.
L 132 or thereabouts: stratification by parity was one of the analyses done – shouldn’t the stratified analyses be mentioned in the Methods?

This is mentioned in lines 91-92: “We performed all analyses with and without stratification by parity (nulliparous versus parous women)”. 

L 163-164: a comment for the authors, not necessarily meant to result in a change in the manuscript: multiple imputation is becoming widely accepted as the method of choice for dealing with missing data, and would have been a better choice here.

In general we agree that multiple imputation is a very useful tool in many circumstances, however, in our dataset most of the missing values were missing outcomes, and in our opinion time to pregnancy (mainly due to unplanned pregnancy) is most likely not missing at random.

L 166: would change “that” to “who”

Corrected.

L 196: should reference 11 be cited when describing the results of Whitworth et al.?

Corrected.

L 214-215: would replace “was obtained which is expected to correlate closely with levels at that time” with “about 3-6 months after conception”, and levels at that time probably correlated closely with those at initiation of the attempt to conceive.”

Corrected, however we replaced ”3-6 months after conception” with “in the first or second trimester”.

L 254: to avoid presenting new data in the Discussion, would move this posthoc analysis to the Methods and Results.

Corrected accordingly.

Table 1: would give the n’s in sample 1 and sample 2 somewhere, perhaps as part of the column headings.

Corrected; listed in column headings as suggested.
Table 1: would define higher and middle/lower SES somewhere, perhaps in a footnote

*This information has been added in a footnote.*

Table 1: time to pregnancy for sample 2: would move entries for 1-2 months, etc. to left, so they line up under the results for < 1 month

*Thank you for pointing out this mistake, which we have corrected.*

Tables 2 and 3: would add log results to Pooled analyses entry. Even though the absolute values were not the same in the two samples, the increase in OR per increase in lnPFAA should be comparable across samples 1 and 2.

*We added this in the Methods section, the tables, and the Results section.*

Table 3: for sample 1, PFOS, parous results, the categorical FOR are above one yet the log result is < 1. This suggests that an influential point is having a large effect on the continuous results. Would comment on this in the text and do some type of influence analysis to address this.

*We observed no influential points in the data. The differences in the analyses using continuous and categorical exposure variables may be due to low precision because of the relatively small sample size after stratification. There is a high degree of overlap between the confidence intervals.*

**Level of interest:** An article of importance in its field

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.