Author's response to reviews

Title: Persistent organic pollutants and non-alcoholic fatty liver disease in morbidly obese patients: a Cohort Study

Authors:

Panu Rantakokko (panu.rantakokko@thl.fi)
Ville Männistö (ville.mannisto@kuh.fi)
Riikka Airaksinen (riikka.airaksinen@thl.fi)
Jani Koponen (jani.koponen@thl.fi)
Matti Viluksela (matti.viluksela@thl.fi)
Hannu Kiviranta (hannu.kiviranta@thl.fi)
Jussi Pihlajamäki (jussi.pihlajamaki@uef.fi)

Version: 2
Date: 13 August 2015

Author's response to reviews: see over
Dear Professor Philippe Grandjean,

Please find below specific answers to each comment raised by the Reviewers 1 and 2 on our manuscript entitled “Persistent organic pollutants and non-alcoholic fatty liver disease in morbidly obese patients: a Cohort Study”.

I have numbered each question raised by the reviewers and made corrections accordingly. In the manuscript and tables changes were made with the track changes function of Microsoft Word. In this document all changes made to the actual manuscript are written in italics to help to distinguish these from other text. Also to easily separate reviewer comments and our response to them, all reviewer text is in grayscale background.

Unfortunately many of the comments by reviewers were such that it was necessary to add text to Introduction, Methods and Discussion. Thus, even though shortening was made where possible especially in the Discussion, overall the length of the manuscript clearly exceeded the recommended word limit.

Finally, we observed a small error in the calculation of the lipid based concentrations of POPs at 12 months. These results were used to calculate the results in Table 2 and Table 5. The impact this error had on the final results was minor and it had no impact on the interpretation of the results. However, these errors are now corrected in Table 2 and Table 5 and in the corresponding places in Results and Discussion sections of the text. This error is also mentioned in the beginning of reviewer 2 comments, because some of those comments have directly to do with the results presented in Table 2 and 5 and thus helps to understand the changes made to these tables.

We hope the answers given may assist in the acceptance of the manuscript for publication in the Environmental Health.

Sincerely yours

Panu Rantakokko
National Institute for Health and Welfare (THL)
Chemicals and Health Unit
Postal address: P.O.Box 95, FI-70701 Kuopio, FINLAND
Street address: Neulaniementie 4, FI-70210 Kuopio, FINLAND
Telephone: +358-(0) 29 524 6395
Email: panu.rantakokko(at)thl.fi

Reviewer: 1

This is an interesting study on the correlation between several POPs and contaminants and liver pathology such as NASH. The study was carried in obese individuals before and after surgery. The biochemical findings in this paper (i.e. correlation between increased ALT values and some POPs 12 months after surgery) are in line with previous findings as outlined by the authors. However the novelty of the paper is that the authors compared the liver histology at baseline with the amount of contaminants in blood. Their findings are extremely
surprising as basically they observe an inverse correlation between the amounts of contaminants and the degree of liver inflammation. This is indeed surprising as there is evidence that 1) these pollutants increase inflammation; 2) inflammation decreases liver capacity to metabolize xenobiotics, 3) there is no reason to believe that inflammatory liver has increased capacity to store these compounds. While these findings are surprising, this is to my knowledge the first time that liver histology is compared to pollutants concentrations in this population. In have some comments.

1) Very often the authors describe variations in the amounts of contaminants when those variations are not statistically significant (even in the abstract). They should be more cautious and describe only what is significant.

Answer: In the revised version of the manuscript description part of the non-significant results in the text were removed. Those retained are well justified, i.e. those match with the overall trend of significant results (are of borderline significance) that deserve a comment or were retained due to comment from reviewer 2.

2) I am confused by the OR columns in tables 3 and 4. Was there a threshold that was set? I may have missed something. Also there is a huge variability. Please explain in more detail.

Answer: No actual threshold was set. In Tables 3 and 4 results from Multinomial Logistic Regression (MLR) analysis are presented. Dependent variable is nominal with 3 or more categories and the first category of the dependent variable is chosen as the reference category, here the liver histology classified as healthy. In MLR separate odds ratios are determined for all independent variables (age, sex, fasting insulin and a particular POP or PFAAs compound included in the model). Of the independent variables results for β-HCH and PCB118 in Table 3 and all PFAAs in Table 4 are tabulated for each non-healthy (or non-normal) category of the dependent variable in comparison to healthy reference category. Due to non-normality of POP and PFAA concentrations, those had to be log transformed in the MLR analysis. This in addition to inverse association found in many cases (especially where results were statistically significant) makes the interpretation of odds ratios more difficult. However, to give an example, the association between lobular inflammation and PCB118 is interpreted that as the concentration PCB118 in the current cohort increases by one log unit (i.e. a 10 fold increase in the concentration), the odds that a particular patient with this increased PCB118 concentration is in the “2-4 foci per 200*field” category in comparison to being in the category “None”, is 0.01. Here it of course needs to be taken into consideration that one log unit covers almost the whole range of concentrations as at baseline 5th and 95th percentiles for PCB118 were 3.1 and 33 ng/g lipid, respectively.

In Table 3 extreme variability in the confidence intervals was observed for liver chirrosis due to too few observations in more advanced cases that caused instability in the model. As all results related to fibrosis stage were non-significant, they were removed from Table 3 (see answer to comment 1). There were also some other cases of broad confidence intervals in both Tables 3 and 4 possibly due to low number of subjects. However, as majority of these results were statistically significant, they were retained.

3) The discussion is long and fails to give any clues to the surprising findings of the paper. Here is one suggestion. The elimination of many POPs and other contaminants is often through the feces. This could be through biliary excretion and decreased intestinal uptake. Is NASH associated with altered biliary excretion or altered biliary-intestinal cycle? Maybe they could use blood chemistry to assess this or at least literature review.
Answer: This was an important comment and helped to interpret the results at least for lipid soluble POPs. For PFAAs not enough background information is available for firm conclusions. However, second and third paragraph of discussion were completely rewritten based on this comment:

“For POPs and PFAAs increased accumulation to liver or increased excretion with disease progression could be proposed as explanations to inverse associations. From animal and human post mortem studies accumulation of many dioxin-like compounds to liver is known to take place by dose-dependent induction of CYP1A2 protein capable of hepatic sequestration of these compounds. However, as PCB-118 and β-HCH did not correlate with CYP1A2 expression and did not accumulate to liver in these studies [16, 43], and as decreased CYP1A2 protein expression levels were observed in liver microsomes of patients with NASH [44], possible hepatic accumulation should take place by other mechanisms. For PFAAs a rat study showed that multiple proteins in the liver are capable for specific binding to PFOA, but those proteins were not specified [45]. Unfortunately, from the very limited amount of liver tissue available it was not possible to analyse the level of POPs and PFAAs and compare them to the levels in serum.

Regarding excretion, the presence of POPs in human bile from autopsy samples with significant relationship to concentrations in blood, adipose fat and liver indicates biliary excretion of POPs [46, 47]. NASH patients generally have increased levels of bile acids in both plasma and liver tissue [48, 49] and an increase of serum bile acids in NASH as compared to less severe stages of NAFLD has been found in obese patients undergoing bariatric surgery [50]. In addition, normally a substantial portion of POPs excreted in bile is reabsorbed by the intestine after deconjugation by intestinal microflora [51], but the possibility of impaired enterohepatic circulation of bile acids in those with NASH has been suggested [52]. Thus, increase in biliary elimination with NAFLD progression could explain the generally inverse associations of serum POPs with NASH and liver inflammation that reached statistical significance for PCB-118 and β-HCH. Contrary to our results, in former lindane manufacturers liver disease (elevated AST, ALT or γ-GT) was positively correlated with longer β-HCH elimination half-life from blood [53]. However, without information of NASH status, the relevance of this compared to our findings is unsure.

For PFAAs, the urine was concluded to be the major elimination route for short PFCAs (C ≤ 8), but for longer PFCAs, PFOS and PFHxS other non-specified routes of excretion likely contributed to overall elimination [54]. Thus, it cannot be deduced whether increased biliary excretion with NASH might explain the inverse associations observed between some PFAAs and lobular inflammation. However, our results at baseline are generally opposite to epidemiological and occupational studies where a slight positive association between PFOA and liver enzymes (ALT and γ-GT) in obese subjects was observed [34, 55].”

4) Minor: there are some typos all along the manuscript.
Answer: Typos were corrected throughout.

In conclusion, the paper provides novel and surprising data. The statistics should be reviewed. Discussion is too long and does not provide real clues.
Answer: Statistics were reviewed in response to the comments of reviewer 2. Discussion was reviewed and shortened where possible, but the requested literature review on the biliary excretion and its relevance to the results was quite long.
Reviewer: 2

General note by the authors: We observed a small error in the calculation of the lipid based concentrations of POPs at 12 months. These results were used to calculate the results in Table 2 and Table 5. The impact this error had on the final results was minor and it had no impact on the interpretation of the results. However, these errors are now corrected in Table 2 and Table 5 and in the corresponding places in Results and Discussion sections of the text.

This is a well written manuscript that evaluates the associations between several environmental contaminants and liver histopathology and function markers in morbidly obese patients that were recruited prior to bariatric surgery and followed up to one year after surgery. Changes in blood concentrations of the environmental contaminants over one year after bariatric surgery are further described. The main strengths of this study include the availability of liver biopsies that are difficult to get in population studies, and of two measurements of numerous contaminants (POPs, PFAAs) and ALT at baseline and at one year of follow-up. The important limitations include the lack of information about changes in weight status prior to recruitment, that are commonly seen in patients prior to bariatric surgery, and also the lack of information of contaminant concentrations in liver tissue that would have permitted to study correlations between liver and blood tissue concentrations and to make more straightforward conclusions about the associations shown. I have some major concerns in relation to the methods and the statistical approach followed that I list more below. I consider that a more clear presentation of the overall causal diagram of the known/hypothesized associations, the repetition and some extension in statistical analysis and a more thorough interpretation of results are required before acceptance for publication.

Major comments:

1) The rational for confounder selection in the multivariate-adjusted models is not clearly explained (page 9). This is a morbidly obese population, however all statistical models were adjusted further for BMI. If BMI is a mediator rather than a confounder in the association between chemical exposures and liver disease, then adjustment for BMI in the models would not be justified. If obesity occurrence was posterior to the alterations in liver function, then adjusting for BMI could lead to collider stratification bias. Also, why adjustment for insulin level was performed in the models? It may be helpful for the authors and also the reader to show a causal diagram of the hypothesized associations between chemical exposures, the outcomes of interest and the additional covariates and base on this the selection of covariates included in the final models. The authors should also consider adding in the manuscript the effect estimates from models adjusted only for age- this could be presented, perhaps, as supplemental material.

Answer: As explained in the introduction, obesity and insulin resistance are considered as major etiological factors in the pathogenesis of NAFLD. Even though there are studies on the association between POPs and obesity (see also the answer to comment on lines 80-81, background section), we treat POPs mainly as a factor that may contribute to obesity and insulin resistance driven NAFLD development. For this reason BMI and insulin level are included as covariates in the statistical analysis. We thus do not treat BMI as a mediator in the association between chemical exposures and liver disease. Also, as obesity is not considered a posterior factor, there should be no risk of collider stratification bias. However, because this is a morbidly obese population, running a model without BMI adjustment is justified. Regarding insulin resistance, because there were large differences in insulin levels between different liver phenotype groups and due to current understanding of its etiology in NAFLD
development, adjustment for insulin levels was also considered justified. However, as it is always reasonably to check the impact of various covariates, results from the models adjusted only for age were included in supplemental material (SI Tables 1, 2 and 3). Also, reference to the models adjusted for age and BMI were removed from the text and replaced with models adjusted only for age.

A short mention for the justification of BMI and serum insulin in the models was added on line 175:

“BMI and serum insulin were added in the models because they are central etiological factors in NAFLD development.”

Finally, because there are publications which have detailed possible mechanistic pathways by which POPs may contribute to insulin resistance and obesity, and thus to NAFLD development, a short reference to them is made on line 88 (1st version of the manuscript) of the introduction because we have theoretically little to add to them:

“Experimental studies suggest that POP POPs may activate nuclear receptors including aryl hydrocarbon receptor (AhR), pregnane X receptor (PXR) and constitutive androstane receptor (CAR). Through these receptors POPs may induce the regulation of genes involved in the inflammatory pathway, mitochondrial function, lipid oxidation, and lipogenesis, thereby contributing to development of insulin resistance and obesity [22, 28].”

2) Associations between chemical exposure and ALT at baseline and at 12 months after surgery are presented stratified by sex (results section and Table 5) and one of the authors main conclusion is that in women the non-significant negative associations shown at baseline turn to significant and positive at 12 months after surgery. However, a similar pattern for many of the chemicals tested is also observed in men even though associations do not reach the level of statistical significance, which may likely be due to the smaller number of the analysed men compared to women. My suggestion is not to completely disregard the associations shown in men and revise accordingly the text where these findings are described in the sections of results and discussion. Further, it is not clear why associations are presented separately in the two sexes. Did the authors test for statistical interactions according to sex? If associations are not shown to significantly differ by sex, then the authors could consider showing results for the population overall in Table 5 and provide results stratified by sex as a supplementary table.

Answer: The comment on the possible interaction with sex was very useful. In the next comment (3) it was asked whether we tried to adjust the models at 12 months for weight loss during the last year instead of BMI. Both of these questions are addressed here.

First. Changing BMI for weight change at 12 months in the linear regression models between ALT and POPs made the B-values much more alike between men and women, at least for PCBs. However, for most compounds those remained significant or nearly significant (p=0.05 – 0.1) for women and non-significant for men (now SI-Table 4). This was possibly due to smaller number of men in the analysis. However, BMI was replaced with weight change in the results to be reported.

Second. We conducted a test for the interaction between POPs and sex in a non-stratified model where the POP*sex interaction term was added as a covariate in addition to age, sex, fasting insulin and BMI (at baseline) or weight change (at 12 months). At baseline only β-HCH had nearly significant interaction with sex. At 12 months POP*sex interactions were stronger especially for those POPs that had significant association with ALT among women, but significance was reached only for β-HCH and Trans-nonachlor. Thus, as only minor proportion of interactions were significant, results for the population overall are presented in Table 5 and results stratified by sex are in SI-Table 4 as suggested above.
These changes caused some change in Methods section describing the test between ALT and contaminants that now reads:

“Associations between log-transformed serum ALT concentrations and log-transformed serum POPs (lipid based) and PFAA concentrations (volume based) at baseline and at 12 months were studied by linear regression. At baseline models adjusted for age and models adjusted for age, sex, BMI and fasting insulin were run. Similar adjustments were performed at 12 months, but BMI was replaced with weight change in kilos as weight loss may be a better predictor of the chemical blood concentrations than BMI at 12 months. Interaction between sex and POPs was tested in separate models at baseline and at 12 months by including sex*POP interaction term to fully adjusted models. Finally, fully adjusted models between ALT and POPs stratified by sex were also run.”

Respective section in the results was completely rewritten:

“Serum ALT was measured both at baseline and at 12 months. In the fully adjusted models at baseline all chlorinated POPs had negative association with ALT that reached statistical significance for β-HCH and PCBs 153, 180 and 170. In the fully adjusted models at 12 months these associations turned positive and were significant for PCB-118, PCB-156 and BDE-153 (Table 5). In the models adjusted only for age associations were much weaker at baseline, but only slightly weaker at 12 months (SI-Table 3). At 12 months most POPs had significant or nearly significant interaction with sex (Table 5). In the analysis stratified by sex significant or nearly significant associations between POPs and ALT were most consistently observed among women at 12 months (SI-Table 4).”

Also in the discussion appropriate changes were also made, but those are not detailed here.

The results for men are now described in proper sections of the Results and Discussion. In particular, on line 283 the following addition was made:

“However, because effect estimates were similar among men for many of the POPs, it is also possible that smaller number of men may in part be the reason for less significant results (SI Table 4).”

3) Since information on chemicals and ALT is available both at baseline and 12 months after surgery, why did the authors not evaluate the association between the change in POPs and the change in ALT levels over the period of one year? The comparison of the effect estimates in the models adjusted and unadjusted by the weight change over this period, it would then help to clarify whether the associations observed between chemicals and ALT are explained by the important weight loss noted in participants after the bariatric surgery. Also, did the authors try to adjust the models at 12 months for weight loss during the last year instead of BMI? Weight loss may be a better predictor of the chemical blood concentrations in this case.

Answer: We ran linear regression models to test the association between the change in POP (log transformed) and change in ALT levels adjusted and unadjusted for weight change (WC, in kilos) and age. This was done for the whole study population and for males and females separately. No significant associations were found between POPs and ALT in any case (p-values mainly ≥0.20). Neither pairwise comparison of effect estimates (computed for each POP using Student’s t-test) revealed significant differences (p-values ≥0.28). Thus, weight loss does not appear to provide a clear explanation for ALT-POP associations.

Regarding the adjustment for weight loss instead of BMI see answer to previous comment.
4) Study limitations should be addressed in discussion (small numbers, false findings due to multiple testing etc). Importantly, the authors should consider how the lack of information about obesity management prior to recruitment may have influenced findings. Physical activity and diet recommendations as well as medication are commonly administered to patients prior to bariatric surgery – may this explain any of the associations shown - eg the negative associations shown between some chemicals and lobular inflammation at baseline? Patients with a better compliance to their physician recommendation prior to surgery would more likely have lost weight which could have led to both an increase in POP levels in blood and an improvement in liver lobular inflammation.

Answer: Study limitations were addressed in the end of the discussion:

“This study has several limitations. First, the number of subjects was limited in each class liver phenotype class. However, clear characterization in different liver phenotypes gave a possibility to compare normal liver to steatosis and NASH. Second, there is a risk of false findings due to multiple testing. Third, analysis of POPs from liver biopsies would have been extremely valuable additional information, but it was not possible due to very limited amount of tissue available. Fourth, follow-up liver biopsies at 12 months would also have been valuable, but receiving they are hard to justify ethically. Fifth, study was cross-sectional in nature and the causality for the main findings cannot be established.”

In the Finnish Current Care Guidelines for obesity surgical treatment for obesity is always preceded by conservative treatment that includes counselling on the thoughts, attitudes, diet and physical exercise. Only if this conservative treatment does not result in a permanent weight loss, surgical treatment is considered. All patients thus had similar obesity management prior to recruitment to Kuopio Obesity Surgery Study (K OBS). Once recruited, all patients included were on a very low calorie diet for 8 weeks before the surgery (800-1000 kcal/day), during which a 5-10 kg weight loss typically occur for all patients. However, this pre-surgery weight loss was not recorded. Because blood samples were taken in the end of the restricted calorie diet period, in the morning of the day of surgery during which liver biopsies were also taken, the impact it had on the liver status and serum levels of POPs in different liver phenotype groups does not affect our results.

A mention of the restricted calorie diet before the surgery was added in the “Study population” section of the methods:

“All participants were on a very low calorie diet for 8 weeks before the surgery (800-1000 kcal/day) during which 5-10 kg weight loss typically occur for all patients.”

A mention of the Finnish Current Care Guidelines and its implications for the possible preoperative changes and study results are given just before study limitations:

“Obesity management prior to recruitment was similar for all patients, i.e. they had non-satisfactory response to conservative treatment given according to Finnish Current Care Guidelines (e.g. counselling on diet and physical exercise). Possible weight losses during pre-recruitment period and pre-surgery low calorie diet period were not recorded. It can be speculated that patients with a better compliance to their physician recommendation prior to surgery would more likely have lost weight which could have led to both an increase in POP levels in blood and an improvement in liver lobular inflammation. Our results at baseline reflect cross-sectional situation on the day of surgery.”
Other important comments by section

Abstract

Background, lines 26-30: In the abstract background should be provided information from prior evidence that justifies the study aims. However, as the text is written the justification of the study aims is not clear. A potential link between POP exposure and the risk of non-alcoholic fatty liver disease has indeed been supported by animal studies but it is not clear whether and why this effect would be “especially seen in obese individuals”. Also, the reason why “levels before and 1 year after bariatric surgery induced weight loss were compared” should be justified here.

Answer: Due to word limit of 250, every chance was taken to minimize the number of words within these corrections to abstract, but hopefully without loss of accuracy.

The whole of the background section was rewritten:
“In animal experiments persistent organic pollutants (POPs) cause hepatosteatosis. In epidemiological studies POPs have positive associations with serum markers of nonalcoholic fatty liver disease (NAFLD) and together with obesity synergistic association with insulin resistance. Because insulin resistance and obesity are critical in NAFLD pathogenesis, we investigated the association of serum pollutant levels with liver histology and alanine aminotransferase (ALT) in morbidly obese.”

Methods, lines 32-33: Please, consider rephrasing this sentence and further specify the period that this study was conducted, eg “Liver biopsies were obtained from 161 morbidly obese participants of the Kuopio Obesity Surgery Study (KOBS) who underwent bariatric surgery between 20??-20??.” The study period of biological sample collection is of great interest as decreasing trends in the levels of exposure for most of the chemicals under study have been noted over time.

Answer: This sentence was rephrased as proposed above with some shortening. Also, some shortening of the next sentence was made. These now read:
“Liver biopsies were from 161 participants of the Kuopio Obesity Surgery Study (KOBS) who underwent bariatric surgery 2005-2011. Liver histology was categorized as normal, steatosis and non-alcoholic steatohepatitis (NASH).”

Methods, line 35: Please, specify in methods that blood samples for the chemical determinations were collected twice and the time of sample collection (during the bariatric surgery and 12 months later?). Also, please consider avoiding the general term of “biochemical variables” and specifying the biomarkers evaluated as study outcomes eg, “biomarkers of insulin resistance (insulin, glucose), metabolism (lipids, adiponectin) and liver function (ALT)”. Also, some of the markers listed here are not later included in analyses, why?

Answer: To answer these questions in a compact manner, the next sentence in the methods was written:
“Liver phenotype at baseline and ALT at baseline and 12 months post-surgery were correlated to serum POP concentrations at respective time points.”

Now “biochemical variables” was thus deleted and replaced with ALT. Also those biomarkers that were not used in statistical test were deleted from the actual methods section of the manuscript.

Results, line 38: Which serum concentrations? The concentrations measured in the samples collected at the time of the surgery or at one year after surgery? Please, specify.
Answer: To specify this, word “baseline” was added in the beginning of the sentence, because this term was used also in the methods section of the abstract (see above). In the actual method section it is clarified further that baseline samples were collected on the week preceding the surgery.

Results, line 39-40: It is not clear why the authors describe results only for women in this sentence. Whether these associations were shown to significantly differ in men should be clarified.
Answer: Please see answer to major comment 2 and other comments below regarding the POP-sex interaction. These sentences were rewritten:

“ALT had negative associations with many POPs at baseline that turned positive at 12 months after major clinical improvements. There was an interaction between some POPs and sex at 12 months, and in stratified data positive associations were observed mainly in females but not in males.”

Also the second sentence in the conclusions was rewritten and the whole conclusion now reads:

“We found a negative association between serum concentrations of PCB-118, β-HCH and several PFAAs with lobular inflammation at baseline. Positive POPs-ATL associations at 12 months among women suggest that increased POP concentrations may decrease the degree of liver recovery.”

Background

Lines 66-69: The use of the POPs and most of the PFAAs evaluated in this study has been banned in developed countries and decreasing time trends of exposure are noted. Perhaps, the authors would like to clarify this.
Answer: This mention was added and now these lines read:

“Production and use of many POPs has been banned by international treaties. Decreasing time trends of exposure have been observed in the Nordic Countries e.g. for polychlorinated biphenyls (PCBs) and organochlorine pesticides (OCPs) since 1960’s and 1970’s [13], for polybrominated diphenylethers (PBDEs) since 1990’s [14] and for perfluorinated alkyl acids (PFAAs) since 2000’s [15].”

Lines 69-76: As written it is not clear that the findings described in relation to liver toxicity occur from experimental studies and not from human studies. Please, revise and rephrase accordingly to clarify this.
Answer: Text was rewritten to make this clearer. One reference was also replaced with a more fitting one and now reads as a whole:

“Due to their lipophilicity POP are stored in humans primarily in adipose tissue and also in the liver [16], but animal and some human data indicate that dioxin-like compounds are also
selectively sequestrated to liver [16, 17]. Contrary to other POPs, PFAAs are not lipophilic and they have been shown to bind to proteins in blood and especially in the liver of animals [18] and to various protein rich tissues also in humans [19]. In rats single high oral dose of certain PCBs [20] and OCPs [21] increases deposition of triglycerides to liver. These effects may also depend on the diet. For example, in rats POP contaminated salmon oil (but not decontaminated) caused hepatosteatosis [22], while in mice POP contaminated whale meat did not [23].”

Lines 80-81: “Serum levels of certain lipophilic POPs have been consistently associated… type 2 diabetes [24], obesity [25] and insulin resistance [26] in cross-sectional epidemiologic studies.”: This is a very general statement that could be expanded and specified. Which POPs have been associated with obesity and diabetes? The direction of the associations (positive or negative?) is also of interest and should be specified. Also, there are several cross-sectional but also prospective studies suggesting that exposure to POPs (mainly DDE and PCBs) is associated with increased risks for obesity and diabetes and perhaps the authors would like to replace the individual studies cited here by some of the many literature reviews currently available. Prospective studies provide stronger evidence compared to cross-sectional studies as can reassure that exposure occurred prior to the disease, thus the evidence from prospective and not only from cross-sectional studies should be highlighted in the introduction. Further, I would argue that although fairly consistent positive associations between DDE and obesity outcomes have been reported across previous studies, both positive and negative associations have been previously reported for other POPs (eg PCBs) and obesity, and associations with diabetes have been much less explored in humans so far. Given the current state of evidence, in my opinion “consistently associated” is a rather strong statement in this case.

Answer: Considering the complexity of associations removing the word “consistently” is justified. However, regarding POPs and T2D there is an abundance of studies even to the point that a US National Toxicology Program Workshop Review held in 2011 concluded that “the overall evidence is sufficient for a positive association of some organochlorine POPs with type 2 diabetes” [EHP, 2013, 121, 774-783]. Regarding POPs and T2D, this study and another further expansion of this review were now cited plus a review on prospective studies. Also a review on POPs and obesity was cited plus 2 available prospective studies on this topic. As a whole, this paragraph now reads:

“In humans serum levels of certain lipophilic POPs have been associated with NAFLD related conditions (e.g. T2D and obesity). Regarding T2D, reviews of cross-sectional studies suggest/support a positive association for certain organochlorine POPs, such as trans-nonachlor, dichlorodiphenyldichloroethylene (p,p'-DDE), and PCBs [25, 26]. A recent meta-analysis of 7 existing prospective studies indicated towards the temporal precedence for hexachlorobenzene (HCB) and total PCBs, but the data was insufficient to establish causality [27]. Regarding obesity, a current review concluded that OC pesticides (especially p,p'-DDE) tend to be positively associated or not associated with obesity, but PCBs have also shown inverse associations in many studies [28]. Two prospective studies have been conducted. One study among young adults observed that p,p'-DDE and PCBs with ≥7 chlorines had inverted U-shaped association with increased body mass index in 18 years follow-up [29]. In another study among elderly adults low-dose exposure to less chlorinated PCBs, p,p'-DDE, and dioxins were associated with the development of abdominal obesity in 5 years follow-up [30]. In addition, POPs and obesity together have a synergistic association with T2D or insulin resistance [31, 32].”

Line 84: Why would the elevation in ALT levels linked to PCB exposure be “unexplained”?

Answer: This was an unnecessary word from abstract of the article of Cave et al. Deleted.
Lines 89-90: I understand the rationale for evaluating associations in the most susceptible population subgroups (in this case morbidly obese subjects) however other populations (eg the general population) are also “suitable” for evaluating these associations. I think this sentence should to be rephrased.

**Answer:** End of sentence was rewritten to better describe the idea of selecting morbidly obese patients. The whole sentence now reads:

“Based on the proposed interaction of obesity and insulin resistance with POP exposure, morbidly obese represent a population where the association between POPs and NAFLD is most likely seen if that exists.”

**Line 95:** The detection of significant correlations between POPs and liver function markers is not enough to support “an association with liver disease”. Were effect estimates calculated in the cited study? Effect estimates adjusted for other factors would be more convincing for the existence of associations between POP exposure and liver function than simple bivariate correlations.

**Answer:** Kim et al (Environ Health Perspect 2011, 119(3):377-383) presented details of this data and statistical methods in the Supplemental material tables. Tables contain partial correlation co-effecients and p-values. In all models age and sex were considered as confounding factors. To represent more accurately methods and results of Kim et al, two sentences on lines 94-97 were modified and now read:

“Pre-surgery serum POP levels had significant positive partial correlations (adjusted for age and sex) with liver dysfunction markers (ALT, AST and γ-GT). Using similar test increased post-surgery POP concentrations were associated with a diminished improvement of liver related parameters [33].”

**Methods**

**Lines 105-107:** This sentence could be moved at the end of the section entitled “study population”.

**Answer:** Moved as proposed.

**Line 110:** Please, specify the chronological period of recruitment. The selection criteria of the study and the participation rate of the initially eligible cases should be also stated in methods.

**Answer:** Following text was added after the first sentence of the paragraph:

“Recruitment and surgery of study subjects took place during the years 2005 – 2011. Selection criteria for the study were according to Finnish Current Care Guidelines that are based on the National Institutes of Health Consensus Development Conference Statement from 1991 [40]. Detailed criteria were 1) BMI > 40 kg/m², 2) BMI 35-40 kg/m² and comorbidity or its risk factor, such as T2D, hypertension, sleep apnea, osteoarthritis of weight bearing joints or polycystic ovarian syndrome, 3) failure of dietary and drug treatments, and 4) no other contraindication for operation. Over 90% of the eligible recruited individuals participated.”

**Lines 116-117:** Please, specify the percentage of patients with high alcohol consumption that were excluded from analyses. What was the rationale for dropping these cases from analyses should be clarified. If alcohol consumption is considered to be an important confounder for the associations of interest, why did not the authors adjust for this covariate in the models instead of dropping these observations?
**Answer:** This sentence was slightly incorrectly formulated. Alcohol consumption of 2 or less doses per day was one of the primary inclusion criteria for the study and for that reason we did not consider adjusting for this variable necessary. Wording was corrected to the form that represents the situation more precisely:

“Alcohol consumption of \( \leq 2 \) doses per day was applied as study inclusion criteria.”

**Line 120:** Please, clarify whether blood samples were collected at fasting or non-fasting conditions - Lipids may be of reduced validity if non-fasting blood has been used for the determinations. Also, more information about samples collected “at baseline” should be provided. Was sample collection performed the exact same day of the surgery? Prior or posterior to the preoperative preparation of patients? If the latter is the case, the authors should consider whether medication or liquids administration at the preoperative period may have influenced the biomarker determinations.

**Answer:** Some of this information was already included in the section “Study population”. However, to make the requested issues more clear and compact, this information was deleted from that section and to the paragraph “Clinical determinations and blood samples” the following text was added:

“During the week preceding the surgery every participant had one-day visit including an interview on the history of previous diseases, current drug treatment and an evaluation of cardiovascular risk factors. On the morning of surgery before preoperative preparation of patients fasting blood samples were drawn after 12 h of fasting. Fasting blood samples were also collected at 12 months after surgery.”

**Lines 130-135:** In the abstract is previously stated that liver biopsies were obtained from 161 subjects, but here it is stated that biopsies were obtained from 119 subjects. The numbers should be revised and corrected accordingly to match throughout the text. Also, later at the end of this paragraph is stated that 105 subjects were finally classified based on histopathology - what about the rest of subjects with a liver biopsy available?

**Answer:** Number 119 was a copying mistake from a previous publication. Even though liver biopsy was obtained for all 161 patients, clear distinct liver phenotype could be determined only for 105 of these 161 patients. To make this clearer, sentence was corrected to:

“Based on this histological assessment a clear distinct liver phenotype (normal, simple steatosis or NASH) could be determined for 105 of these 161 patients (Table 1).”

**Line 145:** “vast majority if samples” meaning >50% of samples?

**Answer:** Cut off was set at 70% and “vast majority” was modified to “more than 70%”.

**Line 166:** The authors perhaps would like to explain why they have used Kruskal Wallis and not ANOVA tests for continuous variables – Were all continuous variables not normally distributed?

**Answer:** This test was used here as most of the variables in Table 1 were not normally distributed according to Komogorov-Smirnov, Shapiro-Wilk tests and also histograms indicated non-normality. Mainly age, BMI at baseline and cholesterol at 12 months showed clear normal distribution. For this reason we have also used in our previous liver biopsy study nonparametric Kruskal–Wallis test for comparisons between groups (see e.g. Pihlajamaki et al. Journal of Hepatology 2012 vol. 56, pages 663–670). However, we have not mentioned this minor issue in the text.

In addition, we removed results for adiponectin from Table 1 because this parameter was not used further anywhere in this publication.
Lines 170-171: Please, explain how the % of change in mean levels was calculated (by
resting the mean of baseline concentrations out of the mean of concentrations measured at 12
months and dividing this by the mean concentrations at baseline? For this, concentrations
were used in original scale or after log-transformation? etc)

Answer: Calculations were briefly described in the footnote of Table 2 as follows: “Mean,
median, 5th and 95th percentiles of serum PFAAs and POP concentration changes from
baseline to 12 months were calculated from changes in individual study subjects.” This text in
the footnote was modified for clarification and now reads:

“Mean, median, 5th and 95th percentile of serum PFAAs and POP concentration changes from
baseline to 12 months were calculated from percent changes in individual study subjects who
had serum sample available at both time points (same n as in the 12 months after surgery
column”).

For the main text this explanation was expanded for further clarification and now reads:

“Percent changes in the concentrations of each PFAA and POP compound from baseline to
12 months were calculated separately for each individual study subject who had serum
sample available at both time points. Percent changes were calculated using concentrations
in the original scale as 100*[conc. at 12 mo]-[conc. at baseline]/[conc. at baseline]). For
each compound mean, median, 5th and 95th percentiles were calculated from percent
changes in individual study subjects.”

Also, in the revised version more attention is paid in the Results and Discussion sections as to
whether specific ∆PFAAs or ∆POPs refer to mean, median or percentiles.

Results

Line 204: The statement in relation to a decrease in BDE209 concentrations does not seem to
hold based on the results shown in Table 2. The mean of concentrations is a bit higher at 1
year compared to baseline and the % of change is estimated to be 148%! Thus, concentrations
are shown to have increased over time. The mean change seems unexpected though and
should be revised. The authors should revise all numbers shown in this table also for other
chemicals and make sure that numbers are correct.

Answer: First, please observe the general note by the authors in the start of the response
document regarding the error in the calculation of the lipid based concentrations of POPs at
12 months. This had an impact on the POP results in Table 2, e.g. median change of BDE-209
raised from 0 to 11%. However, the difference between mean and median for BDE-209
remains large for the same reason as for PFHxA, i.e. a few extreme changes cases raise the
mean without affecting the median as large proportion of results were <LOQ. For clarity, a
comment regarding the impact these extreme values had on the relationship between the mean
and the median change for PFHxA was added as footnote c to Table 2:

“Few extreme values rise the mean of percent change, but median remains 0 as substantial
number of results were <LOQ both before and after surgery.”

Erroneous statement as regards to concentration of BDE209 was corrected, and the whole
sentence now reads:

“However, the behaviour of PBDEs was more variable: for BDE47 and BDE209 median
changes in serum concentrations were small (12 and 11 %), but for BDE153 all changes
(mean, median, percentiles) were similar to those for chlorinated POPs (Table 2).”
All numbers in Table 2 were also checked and rounding of some numbers were corrected for uniform style.

**Lines 221-223:** A similar pattern of negative to positive associations is also shown for men, although associations are non-significant. Comparisons in terms of statistical significance in the associations shown in men versus women are not straightforward due to the important difference in the number of observations in the two subgroups. To evaluate whether significant differences in the associations may exist between the two sexes at baseline or 1 year, effect modification by sex in the associations of interest should be tested in the models. Other metabolic biomarkers have been listed previously in methods (e.g. lipids, adiponectin) but are not presented in the results section. It is not clear whether other biochemical outcomes except from ALT were analysed in this study.

**Answer:** Please see answer to major comment 2 where the question of effect modification by sex has been answered.

Regarding other metabolic biomarkers, those were mainly retained as a copy-paste from previous publications, but were not used in this study. For example, association of POPs with adiponectin were tested, but no significant results were found. Also, of the liver biomarkers only ALT was available at baseline and at 12 months. Thus, biomarkers that were not mentioned in the Results and Discussion were removed also from the Method section.

**Discussion**

**Lines 237-239:** Please, consider using the terms “is associated”/“is less clearly associated” instead of “a strong”/”less strong”, as the meaning of strong is ambiguous.

**Answer:** Corrected as proposed.

**Lines 240-242:** Since effect modification by sex has not been tested, the comparison of effect estimates between women and men is not clear. Please, see also my comments in relation to this more above.

**Answer:** Please see answer to major comment 2 above. Sex*POP interaction terms are included in the new version of Table 5 where the non-stratified associations are now presented. For this reason and to be more exact in the use of terms, sentences on lines 239-242 were modified:

“Second, ALT had negative association with POPs at baseline that turned positive at 12 months after major improvements in the liver function tests. For most POPs these associations were significant or nearly significant at both time points (Table 5). In addition, there was an interaction between some POPs and sex at 12 months (Table 5), and stratified data revealed positive associations between ALT and POPs in females but not in males (SI-Table 4).”

**Lines 242-244:** It is not clear what would the authors like to say here. How would “weight loss may reveal some of the associations between POPs and liver disease after resolution of NASH”? – Please, rephrase and expand the rational to clarify this.

**Answer:** If POPs have a minor contributory effect to the ALT increase, it is possible that this effect is only seen after obesity and insulin resistance are levels in the study population, i.e. there is less “noise” that enables to see this signal. This issue was already described more clearly later on in the discussion, only 3 last words were added now (rows 288-289 in the first version of the ms):
“As a whole, at 12 months increase in POP levels plus improvement in the general clinical situation both increase the likelihood of POP-related ALT signal to be observed if such exists.”

As this is mentioned later on, the sentence on rows 242-244 was removed.

**Lines 279-283:** Statistical interactions by sex should be tested to justify a higher susceptibility in women in this study, if indeed there is any. What are the limitations of this study?

**Answer:** First, please see answer to major comments 2 (interaction) and 4 (limitations), and comment above regarding lines 240-242 (interaction). Start of this paragraph was modified to conform the text with previous changes as follows:

“Change of associations between POPs and ALT from negative at baseline to positive at 12 months and their sex specificity was very interesting (Table 5). Positive POP-ALT associations among women at 12 months (SI Table 4) are in line with other metabolic outcomes. For example, in a 25-year follow-up study, ...”

Some of the following sentences were also shortened to remove redundant text.

**Conclusions**

Conclusions should be revised accordingly to the changes made in other sections.

**Answer:** Modified as proposed. For negative association between PCB-118, β-HCH and several PFAAs with lobular inflammation we added the following sentence:

“Reason for this is unclear, but increase in biliary elimination with NAFLD progression could be proposed as one explanation especially for POPs.”

**Lines 330-331:** The statement “suggesting potential interactions with gender and weight loss” is unjustified here as interactions have not been tested in this study.

**Answer:** Please see answers to major comment 2 and 3 and answers to other comments regarding POP*sex interactions. This sentence was modified:

“ALT had significant or nearly significant negative associations with most POPs at baseline that turned positive at 12 months after major improvements in clinical status. There was an interaction between some POPs and sex at 12 months, and in stratified data positive associations were observed mainly in females but not in males.”

**Lines 331-332:** Positive associations with ALT were shown only at 12 months after surgery and are possibly explained by the important weight loss noted after surgery...

**Answer:** Please see answers to major comment 3 where this issue has been treated in detail.

**Tables**

**Table 1:** Please, provide for categorical variables also %, not only absolute numbers, to facilitate comparisons.

**Answer:** Numbers were added.

**Table 2:** The % of mean change for PFHxA has to be erroneous. Mean and median levels at baseline and 1 year after are identical, thus a 150% of change cannot be justified – please, revise. Also, the % of change in the mean levels for other chemicals seem to be unexpected based on the means shown in Table. For example, the mean concentrations for PFDA are identical (=0,27) but a mean change of 4.4% is then shown. May this be due to the different subgroups analyses at baseline and 1 year later (n=161 versus 118)? The authors should
consider showing in this table descriptive statistics only for the subgroup of patients with available concentrations both at baseline and at 1 year – This would permit more straightforward comparisons.

**Answer:** First, please see the answers to comments in the Methods (Lines 170-171 of ms) and Results (Line 204 of ms) sections. Values in Table 2 were checked several times and no mistakes were found. Highly skewed distribution with a lot of LOQ/2's in the results and the way the change-%'s were calculated can cause such results.

Below is the table from SPSS that shows results for PFHxA and also the highly skewed histogram for PFHxA change-%. As can be seen in the below table, mean at 0 months changes only little if n = 161 or n=118. Because these values change relatively little and the way the calculations were performed are now explained in detail in footnotes, original values with larger n at baseline are retained.

Table 2. Descriptive statistics for PFHxA with n= 161 (baseline) and n=118 (baseline and 12 months).

<table>
<thead>
<tr>
<th>Statistics</th>
<th>PFHxA, ng/ml, 0 mo</th>
<th>PFHxA, ng/ml, 0 mo</th>
<th>PFHxA, ng/ml, 12 mo</th>
<th>PFHxA, Change-%</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Valid 161</td>
<td>118</td>
<td>118</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td>Missing 0</td>
<td>0</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>Mean</td>
<td>.15559</td>
<td>.16393</td>
<td>.15684</td>
<td>149.8286</td>
</tr>
<tr>
<td>Median</td>
<td>.02500</td>
<td>.02500</td>
<td>.02500</td>
<td>0.0000</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.955</td>
<td>1.877</td>
<td>1.849</td>
<td>3.880</td>
</tr>
<tr>
<td>Std. Error of Skewness</td>
<td>.191</td>
<td>.223</td>
<td>.223</td>
<td>.223</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.034</td>
<td>3.182</td>
<td>3.449</td>
<td>18.115</td>
</tr>
<tr>
<td>Std. Error of Kurtosis</td>
<td>.380</td>
<td>.442</td>
<td>.442</td>
<td>.442</td>
</tr>
<tr>
<td>Percentiles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>.02500</td>
<td>.02500</td>
<td>.02500</td>
<td>-92.3674</td>
</tr>
<tr>
<td>25</td>
<td>.02500</td>
<td>.02500</td>
<td>.02500</td>
<td>-44.9425</td>
</tr>
<tr>
<td>50</td>
<td>.02500</td>
<td>.02500</td>
<td>.02500</td>
<td>0.0000</td>
</tr>
<tr>
<td>75</td>
<td>.22660</td>
<td>.19460</td>
<td>.21049</td>
<td>108.9139</td>
</tr>
<tr>
<td>95</td>
<td>59086</td>
<td>60922</td>
<td>55847</td>
<td>1107.7400</td>
</tr>
</tbody>
</table>

Other minor comments and typographical errors:

**Line 66:** Please, specify “the consumption of fatty fish [10]”
Answer: Specified as requested.

Line 81: “type 2 diabetes” should be “T2D” as this abbreviation has been already used more above in the text.
Answer: Corrected.

Lines 187-188: “p-values <0.05 were considered significant” should be erased as it is repeatedly stated also in the next sentence.
Answer: Corrected.

Quality of written English: Needs some language corrections before being published.
Answer: Language corrections were performed where found appropriate to make the text easier to read and more understandable.

---

Additional corrections/additions/modifications by the authors

Results, Line 212: Addition to text was made to highlight trends:
“In addition, serum concentrations of both β-HCH and PCB-118 generally decreased with impairment of all liver histology measures (Table 3).”

Results, Line 212: Sentence was slightly modified to highlight the negative, but non-significant associations for other POPs as well:
“Associations for other POPs were also mainly negative but non-significant and are not tabulated.”

Discussion, Lines 309-317: This paragraph was deleted because: 1) results in Table 2 were changed due to correction mentioned in the General note by the authors, 2) Reviewer 1 asked to shorten the discussion, 3) due to space constraints detailed analysis of the POP levels as a result of weight change is slightly outside the scope of this paper