Author’s response to reviews

Title: Urinary phthalate metabolites in relation to serum anti-Müllerian hormone and inhibin B levels among women from a fertility center: a retrospective analysis

Authors:
Yaoyao Du (1877119496@qq.com)
Na Guo (tsguona@126.com)
Yixin Wang (wangyx0203@163.com)
Xiang Hua (huaxiang24@hust.edu.cn)
Taoran Deng (dengtr@foxmail.com)
Xuemei Teng (1625524732@qq.com)
Yangcheng Yao (yydoc@qq.com)
Yufeng Li (yufengli64@tjh.tjmu.edu.cn; yufengli64@163.com)

Version: 1 Date: 22 Jan 2018

Author’s response to reviews:

Responses to reviewers and editors:

Reviewer #1:
This paper presents a study of the correlations between urinary phthalate metabolites and biomarkers of ovarian reserve (AMH and Inhibin-B). The study is fairly large, and homogeneous in terms of race/ethnicity and smoking habits. The biggest limitations of the analysis are that a single spot urine sample was collected and that the collection seems to have occurred AFTER medical treatments.
Response: We really appreciate your constructive comments on our work. We have extensively revised the paper to address the above issue and each of the specific comments below.

Specific comments are detailed below.
1) The Introduction is well-written, detailed, and clear.
Response: Thank you for the positive comment on the Introduction part.

2) Lines 127-128, "…or had oral contraceptive use before hormone measurement were excluded from the study, leaving 415 women eligible for participation." Please describe how long before hormone measurement women could have used oral contraceptives and still be included.
Response: Women who have stopped using oral contraceptives for more than three months before hormone measurement were eligible for inclusion. To make our expression clear, we have rephrased the sentence “…or had oral contraceptive use before hormone measurement were excluded from the study…” to “…or had oral contraceptive use in the last three months before hormone measurement were excluded from the study…” in the revised manuscript (Page 6, Lines 128-129; Methods).

3) Which kit used in the AMH ELISA?
Response: We used the Ansh Labs AMH ELISA kit. In the revised manuscript, we added this information in the Methods section—“The AMH and INHB assays were conducted using the enzyme-linked immunosorbent assay (ELISA) kit (Ansh Labs, Webster, TX, USA) based on the automated DS2 (Dynex Technologies, Chantilly, USA) ELISA processing system” (Page 6, Lines 136 to 138).

4) Line 142, urine was collected on the day of ovum retrieval, presumably after several days of hormonal treatments. Can the authors speculate to what degree medical interventions might have influenced phthalate levels in this urine specimen?
Response: We have tried to estimate to what degree medical interventions might have influenced phthalate levels in the urine specimen in two ways: 1) we referred to references; 2) we examined whether stimulation procedures [i.e. duration of stimulation and total amount of gonadotropin (Gn)] were related to urinary phthalate levels with our available data.

There is only one research that investigated whether use of assisted reproductive technology (ART) contributed to urinary phthalate exposure through medical supplies and pharmaceuticals (Alur et al. 2015). The study found that women conceived with ART had lower concentrations of molar sum of DEHP, as well as individual metabolite of DEHP, compared with women who had a history of infertility but did not use ART. This unexpected result has been explained by changes in lifestyle choices, especially diet, in ART users and altered phthalate metabolism during pregnancy. Another study by Braun et al. (2012) provided indirect insight into the influence of medical interventions on phthalate levels by evaluating the variability of urinary phthalate metabolites before and during pregnancy in a cohort of women receiving infertility treatment. At least two samples (median: 3) were collected from 113 women for phthalates measurement before pregnancy. A median of 12 weeks elapsed between enrollment and the last prepregnancy urine sample collection (range: <1–110 weeks). Thus, the extent of variations in phthalate levels during this period may reflect the contributions of ART procedures on phthalate exposure. Overall, the absolute concentrations of urinary phthalate metabolites did not vary substantially in the course of medical treatment. When using intraclass correlation coefficients (ICCs) to measure the variability of exposure, concentrations of MBP (ICC = 0.40) and MEP (ICC = 0.56) were more stable, compared to DEHP metabolites and MBzP (ICC≤ 0.35). The authors assumed that diet changes may be responsible for the variations in urinary DEHP metabolite concentrations, while the higher reproducibility observed for other metabolites may be due to unchanged behaviors regarding personal care or cosmetic products. Collectively, there might be variations in phthalate exposure during the course of specimen collection. However, it seems that medical interventions were not the source of variability.

To find out whether hormone treatments affected phthalate levels, we first used spearman correlation analysis to assess the associations of duration of stimulation and total amount of Gn in relation to urinary phthalate metabolites. Then, we further explored the relationships using
multivariable linear regression models with adjustment for age, BMI, ethnicity, smoking status and creatinine. Duration of stimulation and total amount of Gn did not correlate with phthalate metabolite levels in bivariate correlation analysis (see Table R1, below). Likewise, no associations were observed between duration of stimulation and total amount of Gn relative to urinary phthalate levels in multivariable regression analysis (see Table R2, below). Overall, the hormone treatments were not associated with urinary phthalate concentrations in the present study. However, because phthalates were primarily used in oral medications and dietary supplements as excipients, women who used these medicinal products during the study period may affect the exposure assessment.

Taken together, although hormone treatments may not contribute to phthalate exposure, the variability in concentrations of phthalate metabolites, especially metabolites of DEHP, suggest that there is a possibility of reverse causation, which we have addressed as a limitation of our study in the Discussion section (Page 16, Lines 373-378).


5) How much time was there between the day of blood collection for AMH and INHB assays and the day of the urine collection for the phthalate assays?
Response: The median interval between the day of blood collection for AMH and INHB assays and the day of the urine collection for the phthalate assays were 72 d (IQR: 46-93.5 d), which have been added in the Methods section of the revised manuscript (Page 7, Lines 149-150).

6) It would be interesting to see if the sum of DEHP metabolites is related to ovarian reserve biomarkers, in addition to, or instead of, the percent MEHP.
Response: We calculated the molar sum of the DEHP metabolites (\(\sum\)DEHP)—MEHP, MEHHP, and MEOHP, by converting the individual metabolite into molar concentration (\(\mu\)mol/L) and examined the associations between quartiles of \(\sum\)DEHP and the ovarian reserve biomarkers. Distribution of \(\sum\)DEHP after adjustment for creatinine levels (\(\mu\)mol/g Cr) were presented in Table 2. The results of multivariable regression analysis were presented in Table 3-4 and Table S1-2 instead of the percent MEHP, as suggested. No associations were observed between urinary \(\sum\)DEHP in relation to serum AMH and INHB in the total population. However, in age-stratified analysis, \(\sum\)DEHP showed a negative association with INHB across quartiles in the younger subgroup (P-trend=0.03). Other changes relevant to \(\sum\)DEHP were made accordingly throughout the manuscript.

7) Line 166, "...were categorized into quartiles, and modeled as an ordinary variable" I think this should say "ordinal variable"? Also Table 3 looks like a categorical variable was used, and not an ordinal variable...can the authors clarify?
Response: Thank you for pointing out the mistake and unclear expression here. Actually, we intended to express that we used the phthalate metabolites quartiles as ordinal categorized variables with assigned integer values (1-4), and performed tests for trend across quartiles in regression models. We have changed the sentences “...were categorized into quartiles, and
modeled as an ordinary variable. Tests for trend were performed to explore the potential dose-response relationships between phthalate metabolites and the hormones” to “…phthalate metabolites (unadjusted) were categorized into quartiles and estimates for each outcome measure were obtained by comparing each higher quartile to the lowest one (reference category). Tests for trend were performed to explore the potential dose-response relationships between phthalate metabolites and the hormones using the exposure quartiles as ordinal categorized variables with integer values (1-4)” (Page 8, Lines 179-183; Statistical analysis).

8) Linear trend p-values are presented in the tables, even where there does not appear to be a linear trend? I would recommend removing all the p-values from the table and only showing linear trend p-values where there appears to be a linear trend. For example, the associations between MEOHP and INHB do not appear to be linear.
Response: We have removed all p-values for linear trend from the tables as suggested, and used the asterisk to denote where there appears to be a linear trend (see Table 3-4, Table S1-S2).

9) AMH can be classified as "low" or "high" to indicate diminished ovarian reserve (yes/no) and subclinical PCOS (yes/no), respectively. It would be interesting to see if phthalates are related to the probability of either of these clinically relevant dichotomous endpoints.
Response: We used an AMH cut-off value of 1.1 ng/mL to classify the study population into “low” (<1.1 ng/mL) or “normal” (≥1.1 ng/mL) ovarian reserve groups and used it as a dichotomous dependent variable in logistic regression models (Ferraretti et al. 2011). We found that phthalate concentrations were not associated with increased risks of diminished ovarian reserve (DOR) after adjusting for potential confounders (see Table R3 below).
Since there is no consensus as to what cut-off values for AMH should be used to define PCOS, we chose a threshold value of 5 ng/mL according to studies by Dewailly et al. (2011) and Iliodromiti et al. (2013), and based on the distribution of serum AMH levels in the study population. Women with serum AMH>5ng/mL were defined as subclinical PCOS while others constituted the non-PCOS group (yes vs. no). Multivariable logistic regression analysis was conducted to explore whether urinary phthalates were associated with the risk of PCOS. An elevated odds of PCOS across quartiles of MEHP was observed in adjusted model (P-trend=0.04), with the fourth quartile conferring a 2.58 increased odds of PCOS compared to the first quartile (95% CI: 1.29, 5.16). No apparent patterns were present for other urinary phthalate metabolites and PCOS (see Table R3 below).
Considering that the manuscript was primarily focused on the effect of phthalates on DOR, we did not present the results of subclinical PCOS in the paper. Taken together, in the revised manuscript, a) we added the following sentence “Since serum AMH levels could be dichotomized at a cut-off value of 1.1 ng/mL to indicate women with diminished or normal ovarian reserve, we additionally explored the effect of urinary phthalate concentrations on this clinically relevant endpoint using multivariable logistic regressions” in the section of Statistical analysis (Page 10, Lines 213-215); b) we presented the results of associations between urinary phthalates and odds of DOR in Additional file 4: Table S3; c) we described the results of Table S3 by adding the sentences “In logistic regression models comparing women with low (defined as an AMH value of <1.1 ng/mL) versus normal ovarian reserve, urinary phthalate concentrations were not associated with an increased odds of DOR after adjusting for potential confounders (Additional file 4: Table S3)” (Page 12, Lines 271-273; Results).


10) I really like the supplemental figures showing the shapes of the associations based on the restricted cubic splines analysis.
Response: Thank you for the positive comment.

11) It appears that antral follicle count was measured in this study, can the authors present the associations between phthalates and antral follicle count in the paper?
Response: We fitted multivariable generalized linear models with a Poisson distribution and log-link function to estimate the associations of urinary phthalate concentrations with antral follicle count (AFC). We found a positive dose-response relationship between urinary MEHHP and MEOHP levels with AFC after adjusting for age, BMI and creatinine (P-trend for MEHHP=0.001; P-trend for MEOHP=0.004). However, when further adjusting for diagnosis of PCO/PCOS (yes or no) in model 2, these positive associations appeared statistically non-significant (see Table R4, below). Thus, one reason for increased AFC with higher quartiles of metabolites in model 1 may be attributed to extraordinary large amount of AFC in PCO/PCOS women.

Taken together, in the revised manuscript, a) we added the following sentence “In addition to hormone biomarkers of ovarian reserve, in the secondary analysis, we fitted multivariable generalized linear models with a Poisson distribution and log-link function to evaluate the associations of urinary phthalate concentrations with AFC” in the section of Statistical analysis (Page 10, Lines 215-218); b) we presented the results of above mentioned analysis in Additional file 5: Table S4; c) we described the results of Table S4 by adding the sentences “In the secondary analysis using AFC as a biomarker of ovarian reserve, a positive dose-response relationship between quartiles of urinary MEHHP and MEOHP levels in relation to AFC was observed with the adjustment for age, BMI and creatinine (P-trend for MEHHP=0.001; P-trend for MEOHP=0.004). However, when further adjusting for diagnosis of PCO/PCOS in model 2, these positive associations appeared non-significant (Additional file 5: Table S4)” (Page 12, Lines 273-277; Results); d) we added sentence “In the secondary analysis, the findings of AFC and phthalates were inconsistent with that of INHB which is contrary to our expectation. The divergent observations between AFC and INHB might be partially ascribed to their different characteristics in reflecting ovarian reserve. Because AFC could not distinguish healthy from atretic follicles (Dewailly et al. 2014), when phthalate exposure increases follicle atresia and consequently affects hormone production, the occurrence of the discrepant results seems plausible” in the Discussion section to explain the results (Page 16, Lines 363-367).

Reviewer #2:
Comments for: Urinary phthalate metabolites in relation to serum anti-Müllerian hormone and inhibin B levels among women from a fertility center: a retrospective analysis

Summary: The main objective of the study was to quantitatively examine the associations of urinary phthalate metabolite levels with serum anti-Müllerian hormone (AMH) and inhibin B (INHB) levels. The authors report on associations of urinary MEHP and MEOHP with serum INHB, concluding that it adds to the growing evidence that certain phthalate species lead to diminished ovarian reserve. Overall, the manuscript was well written and the analysis seems acceptable, though there are some lingering questions that need to be addressed prior to publication. Most importantly, the authors should better clarify 1) the discrepant results observed for AMH vs. INHB and 2) potential impact(s) of the sample collection timeline (outcome measured 4 months prior to exposure).
Response: Thank you for the review of our manuscript and the helpful comments. We have addressed the above issues and each of your comments, point by point, and listed them below.

Major Comments
*Why were the unadjusted urinary phthalate concentrations, as opposed to the creatinine-adjusted concentrations, categorized into quartiles? One imagine that creatinine-adjusted concentrations likely better reflect true exposure levels.
Response: We considered for urine dilution adjustment by using unadjusted urinary phthalate concentrations with creatinine added as a separate covariate in the multivariable regression models instead of creatinine-corrected metabolite levels may introduce bias (Barr et al. 2005). Creatinine levels have been suggested to associate with age, gender, race/ethnicity, BMI, muscle mass, diet, activity, season, and time of day (Barr et al. 2005; Johns et al. 2015). In our context, we found that creatinine concentrations were inversely correlated with age (data not show). Therefore, in order to adjust urine dilution appropriately independent of other variables associated with urinary creatinine concentration, we chose to model the unadjusted phthalate concentrations in multivariable regression analysis and include creatinine as a separate independent variable. For the revised manuscript, we improved our description of the urine dilution adjustment (Page 9, Lines 189-192; Statistical analysis).

*The authors should clarify why the spline models were fitted using unadjusted urinary MEOHP levels as opposed to the creatinine-adjusted levels? From the quartile results, the non-linear relationship is not obvious as estimates do not meaningfully differ between quartiles. It would be very helpful if authors clarified their analysis and thought process behind this.
Response: The reason that we fitted the spline models by including creatinine as a separate independent variable instead of using the creatinine-adjusted levels is the same as the above. We
estimated that there may exist a nonlinear relationship between urinary MEOHP and serum INHB based on two considerations. First, the second to fourth quartiles of urinary MEOHP all had a significant decrease in serum INHB levels compared with quartile one. Moreover, the largest decrease found in quartile two suggested that there might exist a threshold value. Decrease in INHB seemed to reach a plateau when concentrations exceeded the threshold, reflecting a potential nonlinear relationship. Therefore, we attempted to verify our assumption with spline analysis. To make the analysis clear, we rephrased the description in the revised paper (Page 11, Lines 250-252 and Lines 254-256; Results).

*The authors should have an expanded discussion of temporal relationship as the exposure (phthalates) appear to be measured up to 4 months after the outcome. A couple of potential issues arise from this and needs to be discussed:

- Phthalate exposures might be a reflection of prior knowledge (reverse-causation)
- A single urine sample might have some capacity to represent exposure over a few months preceding sample collection, but the urine sample is taken 4 months after the serum outcome measures. Therefore, the measured phthalate concentrations in urine needs to be reflective of a period of time longer than a few months for temporality to be established.

Response: In the revised manuscript, we have addressed the above issue as a limitation of our study in the Discussion section as follows: "Since urine collection for exposure measurement was conducted up to 4 months after AMH and INHB assays, there is a possibility of reverse causation. Phthalates could be found in a wide range of personal care products, foodstuffs, certain types of oral medications and medical devices, thus we could not exclude the possibility that urinary phthalate concentrations might be influenced by changes in daily care routines and dietary habits, as well as medical interventions during the 4-month period. Additionally, misclassification of exposure was likely to occur not only from timing of exposure and outcome but also from our single measurement of spot urine because of short half-life of phthalates and variability in individual behaviors. However, it has been suggested that one spot urine measurement has moderate capacity to reflect average exposure within 4 months (sensitivity ranging from 0.58 to 0.77), when a surrogate category analysis was performed (Dewalque et al. 2015). Given that the median time intervals between blood collection for AMH and INHB assays and urine collection for phthalate measurement were 72 d (IQR: 46-93.5 d) in the study, the single measurement might allow for a moderately reliable ranking of a few months exposure preceding urine collection” (Page 16-17, Lines 373-385).


*In lines 241-247, the authors cited the fact that AMH is produced by small growing follicles as opposed to non-growing primordial follicles as a central reason why it led to a null finding. However, this does not explain why there was a difference between AMH and INHB as INHB is similarly produced from small growing follicles (line 113). The authors should clarify this, as well as any other potential reasons for differences in AMH vs. INHB results.

Response: As indicated by the reviewer, we added a paragraph in the Discussion section to consider this issue: “The discrepancy between serum AMH and INHB in relation to phthalate metabolites was in agreement with an animal study exploring the effect of acute DEHP exposure on fecundability and ovarian aging. Still, the divergent finding was somewhat puzzling.
Nevertheless, several potential explanations were considered. Although both serum AMH and INHB are secreted by small growing follicles, follicles that contribute the most to AMH and INHB levels are differed by diameter. Follicles of 1–2 mm in diameter are probably the main contributors to serum AMH levels with the evidence that the intrafollicular concentrations of AMH decreased gradually with increasing follicle diameters (Dewailly et al. 2014), concomitant with the observation that granulosa cells of secondary, preantral and small antral follicles <4 mm in diameter expressed the highest level of AMH (Weenen et al. 2004). In contrast, concentrations of INHB increased with the growth of follicles until they reached a diameter of 9 mm (Andersen et al. 2010). Hence, larger antral follicles at later developmental stage constitute the primary source of serum INHB. Based on these facts, if a patient whose AFC is mostly represented by small follicles (i.e. 1–2 mm), while phthalates target larger antral follicles, the discrepant findings between serum AMH and INHB relative to urinary phthalate concentrations might occur. Another explanation is probably that women included in the present study had different status of ovarian reserve manifested as altered concentrations of gonadotropin hormones [i.e. follicle-stimulating hormone (FSH) and luteal hormone (LH)] in the early follicular phase. While AMH function as a paracrine factor independent of endogenous hormones, the biosynthesis of INHB is regulated by FSH and LH, and involved in the pituitary gland feedback loop (Wunder et al. 2008). Therefore, the different characteristics of ovarian reserve in the study population may be a potential reason for inconsistency between the AMH and INHB results” (Page 16-17, Lines 341-342 and Lines 346-362).


*Please provide any QA/QC results for all measured biomarkers (phthalates, AMH, INHB).
Response: We have added a description of QA/QC results for phthalates, as well as AMH and INHB in the revised manuscript as follows: a) “For each batch of 60-110 samples, one blank, two quality control samples and six standards were processed along with the urine samples. The blank control contained 1-mL water was used to assess the contaminations during sample processing and analysis. Two quality control samples spiked with 5 and 50 ng/mL of the target phthalates, respectively, were used to determine the validation of intraday method accuracy by calculating the recovery. The average recovery for target compounds ranged from 88.06% to 110.93%, and relative standard deviation (RSD) was less than 10.00%. The calibration curve generated from six analytical standards had a linearity of >0.99, with a range of 0.5-200 ng/mL. If any metabolites in urine samples had much higher concentrations than the linear range of the calibration curves, the samples were re-analyzed after dilution of remaining samples to ensure the accuracy of measurements” (Page 7, Lines 156-164; Methods); b) “For both AMH and INHB
assays, six standards and two quality controls of high and low concentrations were run with
serum samples in each assay to monitor accuracy and precision. The standard calibration curves
were linear (r2>0.99) with a measuring range of 0.06–18 ng/mL for AMH, and a range of 10–
1500 pg/mL for INHB. The RSD of both assays were less than 10%. The inter-assay and intra-
assay coefficient of variations determined by quality control samples were ≤15% and ≤10%,
respectively” (Page 6-7, Lines 138-143; Methods).

Minor Comments
*Figure 1 - Please label green vs. red lines.
Response: We have indicated what green and red lines stand for in the figure legend.

*All abbreviations in tables should be clarified in the footnotes.
Response: All abbreviations in tables have been defined in the footnotes where they first
appeared.

*IQR is erroneously labelled as "IOR" in Table 1.
Response: We have made the correction accordingly.

*While MOP had a low rate of detection, instead of ignoring it completely, it would be
interesting to see it analyzed as a dichotomous exposure (>LOD and <LOD) to take advantage of
the available data. It would have comparable, if not better, power than comparisons of any two
quartiles.
Response: As suggested by the reviewer, we dichotomized urinary concentrations of MOP as
either being below or above the LOD, and examined the associations of serum AMH and INHB
with MOP using multivariable linear regression analysis. Women with MOP concentrations
above the LOD had higher levels of serum AMH compared to those with values below the LOD
after adjusting for age, BMI, creatinine and PCO/PCOS diagnosis. No associations were
observed between serum INHB and urinary MOP. The finding of a positive association between
urinary MOP and higher levels of AMH is puzzling. Nevertheless, it should be noted that
because of the low levels and detection rates of urinary MOP, as well as the multiple
comparisons made in the analysis, the result might be spurious or a chance finding.
Taken together, in the revised manuscript, a) we changed the sentence “Because more than 70%
MOP were below the LOD, it was not considered for further analysis” to “Because more than
70% MOP were below the LOD, we dichotomized concentrations of MOP as either being below
or above the LOD and considered it as a dichotomous variable in subsequent analysis” in the
section of Statistical analysis (Page 8, Lines 174-176); b) we added the sentence “The
associations between MOP and the hormone measures were also evaluated with concentrations
of MOP < LOD as the reference value” in the section of Statistical analysis (Page 8, Lines 183-
184); c) we added the results of regression analysis for MOP in Table 3-4 and Table S1-2; d) we
described the results of MOP by changing the sentence “Similar results were obtained when
additionally adjusted for PCO/PCOS diagnosis (yes or no) in model 2” to “Similar results were
obtained when additionally adjusted for PCO/PCOS diagnosis (yes or no) in model 2, except
for an increase in AMH levels comparing MOP > LOD to those below the LOD” (Page 11, Lines
236-238; Results); e) we changed the sentence “…we failed to identify such a relationship
between urinary phthalate metabolites and serum AMH levels…” to “we found largely null
associations between urinary phthalate metabolites and serum AMH levels” in the Discussion.
section (Page 12, Line 281); f) we added the following sentences “Finally, we observed a positive association between urinary MOP and higher levels of AMH, which was the sole finding that reached statistical significance. Nevertheless, it should be noted that because of the low levels and detection rates of urinary MOP, as well as the multiple comparisons made in the analysis, the result might be spurious or a chance finding” to discuss the results of MOP and serum AMH (Page 13, Lines 303-306; Discussion).

*Lines 169-170 - given that in linear regression analyses, the addition of covariates will likely not lead to a loss of power, the authors should consider adding ethnicity and current smoking as covariates in the model and compare the resulting estimates to the models presented in the manuscript.

Response: We re-ran the linear regression analyses by additionally adding ethnicity (Han vs. other ethnic groups) and smoking status (never smoked vs. ever smoked) as covariates in the models. Because only two women were current smokers, they were incorporated into the ever smoked group, defined as a current or former smoker. As expected, further adjustment for the covariates did not change the estimates materially (see Table R5-6, below). Therefore, we decided to retain the original models presented in the manuscript.

Assistant Editor comments
1. line 98: the authors note that exposure to phthalates is ubiquitous. Can the authors explain how detection of phthalates among women seeking infertility treatment implies a specific association with reproductive outcomes, as opposed to being expected in any population given the ubiquity of exposure?

Response: The rapid growth in infertility population over the past decades has been paralleled with a huge production of phthalates (Hamilton and Ventura 2006; Hannon and Flaws 2015). Moreover, a growing body of evidence has linked phthalates with impaired human fertility. Therefore, though exposure to phthalates is ubiquitous, women seeking infertility treatment may represent a population at higher reproductive risk due to the possibility that they may have the highest exposures or more vulnerable to phthalate-induced toxicity. To clarify this issue, we changed lines 98-99 to “Notably, in our previous study, phthalates have been detected in the follicular fluid of women seeking infertility treatment, a population at higher reproductive risk due to the possibility that they may have the highest exposures or more vulnerable to phthalate-induced toxicity” in the Background section (Page 5, Lines 98-101).


2-6. There are a number of issues to be addressed regarding aspects of the statistical analysis that suggest additional text in the methods and/or edits to the text otherwise.

Response: Throughout the revised paper, we followed the suggestion of the editor to address the issues regarding the statistical analysis and listed the point-by-point responses below.

2. Natural log transformation of AMH and INHB as dependent variables is described on line 164, so that linear regression models include the transformed outcome and quartiles of
phthalates. The authors should describe the impact of this transformation on interpretation of regression coefficient estimates, and what approaches were taken to use model output to reflect some meaningful number (e.g., geometric means, % difference compared to reference, etc). This is issue is highlighted by the alternating use of “beta (95%CI)” and “percent change (95%CI)”.

Response: In the revised manuscript, we added the following sentences “Given that in the regression analysis serum AMH and INHB were ln-transformed, we calculated the percent change with the regression coefficients (β) to allow for easier interpretation of the results. The percent change and corresponding 95% confidence intervals (CI) were calculated as follows: \[\exp (β)-1\] * 100, which indicates a percent difference in the outcome comparing each of the higher category of exposure to the lowest one (reference category)” (Page 9, Lines 195-199; Statistical analysis) to described the methods we used for transformation and the interpretation of percent change.

3. As noted by reviewer #1, lines 164 through 166 describes use of quartiles for phthalates but it is not clear how these were included in models, due in part to use of the term “ordinary variable” in the absence of further information otherwise, along with the results in tables 3 and 4 which are more consistent with inclusion of the quartile variable being modeled with indicators. Further, if it is the case that phthalate quartiles were modeled in the non-parametric fashion described above, no assumption of linearity would be implied and the role of the restricted cubic spline models in the analysis should be modified slightly to reflect this.

Response: In addition to comparing each higher quartiles of phthalates to the reference category (for MOP, >LOD vs. <LOD), we used the phthalate metabolites quartiles as ordinal categorized variables with assigned integer values (1-4), and performed tests for trend across quartiles in regression models (except for MOP). In the revised manuscript, we have changed the sentences “…were categorized into quartiles, and modeled as an ordinary variable. Tests for trend were performed to explore the potential dose-response relationships between phthalate metabolites and the hormones” to “…phthalate metabolites (unadjusted) were categorized into quartiles and estimates for each outcome measure were obtained by comparing each higher quartile to the lowest one (reference category). Tests for trend were performed to explore the potential dose-response relationships between phthalate metabolites and the hormones using the exposure quartiles as ordinal categorized variables with integer values (1-4)” (Page 8, Lines 179-183; Statistical analysis).

4. Line 166 notes tests for trend, but methods for these tests are not described and should be added if these tests are retained in the paper at all (see comments from Reviewer #1).

Response: We added a description of the methods in the revised manuscript (Page 8, Lines 181-183; Statistical analysis).

5. Use of restricted cubic spline models was noted as a strength by the referees, but some more details regarding their use and interpretation are important and would strengthen the manuscript.

Response: Following the suggestion of the editor, we added additional text to provide more details regarding the use and interpretation of restricted cubic spline models. Taken together, in
the revised manuscript, a) we added the following sentences: “The optimal number of knots was selected based on model fit and biologic plausibility. We chose the 3-knot RCS function because it had lower Akaike Information Criteria (AIC), which suggests that the model could better explain the observation, while easier for interpretation in the biological context. The location of knots was selected as usually recommended, because it has been suggested to have a little impact on the shape of the dose-response association compared to the number of knots (Durrleman and Simon 1989). The Wald chi-square test was used to assess the overall and nonlinear associations between phthalates exposure and the hormones (Desquilbet and Mariotti 2010)” to detailed the use of spline analysis (Page 9, Lines 202-208; Statistical analysis); b) we added the following sentences: “The test for the overall association between MEOHP and INHB in the younger subgroup was significant (P for overall association <0.05), which means whatever the shape of the association is, MEOHP was significantly correlated with INHB. The null hypothesis of the test for nonlinearity that assumed a linear relationship between MEOHP and INHB was rejected (P for nonlinear association=0.01), suggesting there exists a nonlinear association. Collectively, we observed that within a range of relatively low levels of exposure, serum INHB was negatively associated with urinary MEOHP in a dose dependent pattern. However, when a threshold level was met, the decline in INHB was attenuated” to described the results and explain how nonlinearity was evaluated in these models (Page 12, Lines 264-270; Results); c) we added a comment on this result as follows: “Although the specific mechanisms by which phthalates affect ovarian function remains unclear, there is evidence that phthalates could exert their effects through binding steroid receptors (Craig et al. 2011). Thus, one possible explanation for the attenuated decrease in INHB at higher MEOHP level might be that the receptors at the target tissue or cells are gradually saturated, leading to the observed nonlinear biological effect. In addition, our findings do show some cause for concern over the low-dose effect of phthalate exposure and need further verification in both animal and human studies” (Page 14-15, Lines 329-335; Discussion).


6. On line 180, the authors note an interest in evaluating effect modification and, to this end, describing running stratified models. Please provide information regarding how these models were used to test effect modification and provide support for statements regarding whether associations were modified by age (such as line 204)?
Response: To test effect modification, we included a product term between metabolite quartiles and strata of age (<35 years versus ≥35 years) in the regression models. We found that the effect of urinary phthalates on hormone measures were not modified by age (P for interaction >0.05). Taken together, in the revised manuscript, a) we added the following sentence: “We examined whether age modifies the effect of urinary phthalate metabolites on serum AMH and INHB levels by adding a product term between metabolite quartiles modeled as ordinal variables and strata of age in the multivariable regression models” in the Statistical analysis section (Page 9, Lines 210-212); b) we changed lines 218-220 to “In the age-stratified analyses, women in the
younger subgroup had apparent decrease in INHB levels for MEHP, MEOHP and ∑DEHP quartiles, whereas women ≥35 years had no significant change in INHB levels (Additional file 3: Table S2). Although the product terms were not statistically significant (P for interaction >0.05), this finding may imply a potential effect modification by age” (Page 11-12, Lines 257-261; Results); c) we changed lines 273-274 to “In line with the study reported by Messerlian and Souter et al., younger women were at higher risk of phthalate-induced ovarian toxicity, as indicated by the stratified analysis” (Page 15, Lines 335-336; Discussion).