Peer Review File

Article information: https://dx.doi.org/10.21037/tau-23-503

Reviewer A

Comment 1: The authors address the important question of how in utero exposure to phthalates affects testis development and function. Mice were exposed to DEHP by gavage from gestational day 10 (assessed by presence of vaginal plus after mating) until PND0.
A key concern has been identified:
The authors have a similar paper published in Advanced Biology: "Prenatal DEHP exposure induces premature testicular aging by promoting Leydig Cell senescence through the MAPK signaling pathways". DOI: 10.1002/adbi.202300130
The similarity between the materials and methods for this study and the one that is published is quite significant, though the outcomes from in vivo administration of a single dose were reported on in the published work; this one includes three doses.
A significant amount of the content provided here in Figures 2-4 are published in Figure 1 of the Advanced Biology publication, but this other manuscript is not cited.

Reply 1:
Thanks for your concern and advice. In fact, the materials and methods for this study is similar to the study I have previous published, but they do have many differences. Firstly, the focus is different, this study focused on lifelong testicular toxicity and the mechanism of interfering with steroidogenic gene expression, the previous study focused on only testicular aging and the mechanism of premature testicular aging. Secondly, this study also explored the impact of exposure dose on toxicity, but the previous study did not do so. Thirdly, our previous study only did 500mg/kg/d exposure experiment to induce aging model, this study repeated some experiments and acquired similar but not duplicate data or results. To avoid confusion, I will cite the study I previously published and state that"partial results were similar to our previous study"in the revised manuscript. (see Page 11, line 263-264; page 12-13, line 298-299; page 14, line 325-326; page 16, line 355-356)

Comment 2: H&E staining presented in Figure 4 of this submission is a subset of the histology image in Figure 1 in the Advanced Biology paper.

Reply 2: Thanks for your caution, but the H&E staining presented in Figure 4 of this submission is indeed not a subset of the histology image in Figure 1 in the Advanced Biology paper. To avoid misunderstanding, I have replaced the figures.

Comment 3: The authors state that they performed optimisation studies for the MEHP dosage for cell culture of TM3 cells (data not shown). This optimisation was published in the AB paper, but it was not referenced in this submitted manuscript.

Reply 3: Thanks for your caution. This optimisation was indeed published in the AB paper, it was referenced in revised manuscript (see paper16, line348)

Comment 4: With regard to the TM3 cell line data:
The number of replicate experiments is not provided
Outcomes relating to testosterone production was published. Although the description indicates these are the same experiment, the reported data is different, so it is hard to know what data are accurate. The reduction in Testosterone reported in the AB paper is greater than that reported in this manuscript.

Reply 4: Thanks for your comment. I have added the number of replicate experiments in revised manuscript (see figure legend 1). The experiments of testosterone production were duplicated, so the deviation was inevitable, and the data both in this study and AB paper are accurate but with a certain range of deviation.

Comment 5: The LHCGR, HSD3B2 and HSD17B1 protein data (by Western blot) was published in the AB paper. The AB paper differs in that it additionally presents data of the culture without HCG added, but ultimately the data presented are the same.
We must conclude that the only unpublished data presented in this manuscript is the steroidogenic transcript measurements, and this manuscript is essentially a duplication of the authors' own work.

Reply 5: Thanks for your caution, the LHCGR, HSD3B2 and HSD17B1 protein data (by Western blot) was indeed from the repeated experiments which were similar to AB pater. However, we did repeat this experiment and acquired similar but not identical results. To better verify the important role of interfering with steroidogenic gene expression, so, we duplicate this experiment and focused on steroidogenic gene expression. We admit that some results in this study are similar to our own work.

Comment 6: Other concerns relating to how information is presented are provided

below to assist with possible revisions.

Detail in abstract is insufficient, as the exposure interval and mouse strain are not listed.
Abbreviation 'WB' is not appropriate; the words should. E written out.
Referring to 'TM3 Leydig cells' is incorrect; this is a Leydig cell line.

Reply 6: Thanks for your comment and advice. For the word limitation of the abstract, many details including the exposure interval and mouse strain were not present in abstract, we have added details as much as possible. The word WB has been written out. 'TM3 Leydig cells' has been expressed as 'TM3 cells'. (See page 2-3)

Comment 7: The conclusion that 'reproductive aging' (e.g. line 53 in abstract) has occurred is unsubstantiated.

Reply 7: For the word limitation of abstract, the details of 'reproductive aging' are not fully present. Fat, reduced testosterone and semen quality, and atrophic seminiferous tubules at PNM6 could indicate the reproductive aging, we have added these details (see page 3 line 58-59).

Comment 8: Line 87 and following paragraph are not fully correct.
Testosterone synthesis in utero takes place in Sertoli cells, then switches after birth to Leydig cells. The role of STAR is also different in fetal life.

Reply 8: Thanks for your comment. According to previous study, Fetal Testosterone and INSL3 synthesis in utero also takes place in fetal Leydig cells (Zirkin BR, Papadopoulos V. Leydig cells: formation, function, and regulation. Biol Reprod. 2018 Jul 1;99(1):101-111. doi: 10.1093/biolre/ioy059; Voutilainen R. Differentiation of the fetal gonad. Horm Res. 1992;38 Suppl 2:66-71. doi: 10.1159/000182601. PMID: 1292985), not in Sertoli cells, Sertoli cells synthesis AMH. Testosterone, INSL3, and AMH cooperatively promote masculinization. STAR (steroidogenic acute regulatory) promotes the transfer of cholesterol into the mitochondria of LCs, and is very important in gonad differentiation in fetal life. In this part, we mainly emphasize the role of STAR in testosterone synthesis. So, we insist that this part has no obvious errors.

Comment 9: The meaning of the sentence on line 99 is not clear.
While lines 102 – 106 highlight that variation in outcomes relating to testosterone

production following phthalate exposure are reported, the authors make no attempt to justify their own experimental design. Why this strain? Why these doses? Why this treatment regime?

Reply 9: Thanks for your comment. The sentence on line 99 has been revised (see page 5, line 108-110). According to your advice, to justify our own experimental design, we added this sentence “In this study, we not only utilized qPCR and WB analysis simultaneously, but also explored testosterone synthesis and steroidogenic gene expression in the neonatal testis and TM3 cells to verify mechanism” (see page 5, line 117-119).

Comment 10: Can the authors confirm that the animals were treated on the day of birth (Lines 127-128, state treatment extended to PND0)?

Reply 10: It may be a mistake of expression. I want to state that the treatment was extended to delivery day but before delivery. We treat the pregnant mice with no interval before childbearing in the morning. I have revised the expression (see page 6, line 141)

Comment 11: Lines 169-170: What is the relevance of ‘simulating the intrauterine environment’. This appears to be an error.

Reply 11: Thanks for your comment. HCG is massively secreted by placenta and stimulate fetal testosterone synthesis. So, 50mIU/ml HCG (Ruige, Ningbo, China) was added to simulate the intrauterine environment. We think this sentence has no error.

Comment 12: Line 277: what does the following mean? “relatively hypogenetic”

Reply 12: To avoid mistake, I have delete the word ” relatively”.

Comment 13: Line 304: The use of “miraculously” to describe an unexpected result is not warranted.

Reply 13: we have used “Curiously” to replace this word (see page 14, line 320) according to another reviewer’s advice.

Comment 14: What is the rationale for concluding that the outcomes reported in figure 4 represent “reproductive aging”, aside from the authors’ own previous work?

Reply 14: Reduced AGI, fat, Low testosterone and poor semen quality are manifestation of testicular aging. In this study, the meaning of “reproductive aging” is aging of reproductive system, so, the presentation of Reduced AGI, fat, Low testosterone and poor semen quality prompt us to conclude the outcomes represent “reproductive aging”.

Comment 15: Line 311: There is no evidence that changes in mRNA levels were normalised. Are there simply fewer Leydig cells?

Reply 15: Yes, the change in mRNA levels may be due to the fewer Leydig cells, but this study hasn’t explored the mechanism about this.

Comment 16: Line 361: The word “perfectly” is not appropriate here.

Reply 16: I have deleted this word.

Comment 17: Throughout the manuscript, the authors use the word ‘instant’ to describe results evident at PND1 after 9 or 10 days of exposure to high phthalate levels. It is not clear whether 1 day, 4 days or 9 days of exposure resulted in the measured outcomes. Whether this impact is immediate or the result of ongoing exposure is not clear. I’d suggest they use a more conditional phrase such as ‘testicular toxicity evident at birth’.

Reply 17: I agree with you in this part. But if use phrase such as ‘testicular toxicity evident at birth’, the expression may be too complicated and bureaucratic. So, we use immediate testicular toxicity for simplicity in revised manucsript.

Comment 18: Line 376: The following should be rephrased for clarity “Due to the poor therapeutic effect… and adult infertility”.

Reply 18: I have revised this sentence (see page 18, line 394-396).

Comment 19: Line 407” The word “fantastic” is not appropriate.

Reply 19: I have used “interesting” to replace it.

Comment 20: Histograms should show individual data points so that the spread of data are evident to the reader.

Reply 20: I have revised the figures and showed individual data points.

Comment 21: The sentence in the Acknowledgement “Who contributed”… is not clear.

Reply 21: I have revised this sentence (see page 21, line 464).

Reviewer B

Comment 1: This manuscript reports the results of two related experiments designed to test the effect of gestational phthalate exposure on lifelong male reproductive health in mice. First, mice were exposed from gestational day (GD 10) to postnatal day (PND) 0 to 0, 100, 500, or 1000 mg/kg/d DEHP and assessed on either PND 1, PND 56, or postnatal month (PNM 6) for reproductive health parameters and expression of genes involved in testosterone biosynthesis. Second, the authors treated the TM3 mouse Leydig cell line with MEHP and reported steroidogenic gene expression. The study is well-designed in several ways, including the critical consideration of using the active metabolite, MEHP, for the in vitro exposure. The choice of postnatal endpoints is justified by the biology, and the sample size is very large, which suggests a well-powered study.

Reply 1: thanks for your comment and encouragement.

Comment 2: There are several inconsistencies between the present report and prior phthalate studies that result in conceptual issues that should be addressed. First, it is generally accepted that in the majority of prior studies using fetal mouse models, phthalates have no effect on testosterone or lead to an increase in testosterone (see reviews by Johnson et al. 2012 doi:10.1093/toxsci/kfs206; and Albert and Jegou 2014 doi:10.1093/humupd/dmt050). In this experiment, there are effects on testosterone stemming from fetal exposure, as early as PND 1. Did the authors investigate whether these effects begin in utero in their model? If so, how can this difference with prior studies be explained? If not, how do they arise so soon after the relatively insensitive gestational period? The mice used in the present study are C57BL/6J mice, so a strain difference seems unlikely, but I would suggest adding a fetal timepoint to this study to determine whether testosterone levels are reduced during gestation, which would be a difference from prior studies.

Reply 2: According to our study, no hypospadias was found in male mice model either

in neonatal or adult stages after prenatal DEHP exposure, and only prenatal exposure to 500 and 1000mg/kg/d DEHP show reduced testosterone in neonatal stage. So, we speculate that the mouse is more insensitive to prenatal DEHP exposure. In addition, the mice's testis is too small, which may restrict the measure of testosterone. Based on these situations, we collected bilateral testes (about 4mg) to measure intratesticular testosterone, other than the serum testosterone. We think the sample difference may be the cause of the result difference between our study and the previous study. Our ingenious method would improve the accuracy of results. At PND1, the male mice were exposed to DEHP recently, and at this timepoint, the accumulative effect was the most significance. So, I think this timepoint (PND1) was enough. My experiment has finished and has no condition to add experiment in fetal timepoint, I beg your understanding.

Comment 3: A less significant conceptual question relating to life stage pertains to the TM3 cells. TM3 cells, if I am not mistaken, are juvenile (postnatal) Leydig cells. These are different in life stage (gestational vs. postnatal) from the mice used for the in vivo exposures. They also potentially are derived from different precursors, as fetal and adult Leydig cells arise at different times. This should be discussed, as it is a potential limitation for comparing the in vivo and in vitro studies presented here.

Reply 3: thanks for your comment, I agree with you. Fetal and adult Leydig cells have many differences, but few cell line represents fetal Leydig cells. So, this study also selected TM3 cells. I have discussed this shortcoming in the revised manuscript. (see page 20, line 119-452: Lastly, prenatal DEHP exposure mainly impact fetal Leydig cells, the TM3 cells can’t simulate fetal Leydig cells perfectly, the primary Leydig cells extracting from fetal testis may be more representative for vivo experiment).

Comment 4: Second, in prior mouse in utero phthalate studies, administration of high phthalate doses to mice prior to GD 14 led to significant fetal toxicity, including litters with >70% resorptions (Gaido et al. 2007. doi:10.1093/toxsci/kfm049). Here, the authors report no fetal toxicity. Again, is there an explanation for this, such as a strain difference? Given the significance of these differences, if a mouse strain difference is the likely cause, it would be worth conducting a direct comparison of strains to test for strain differences.

Reply 4: this study indeed found no difference in fetal resorption after DEHP exposure. We also feel puzzled for this phenomenon. The mouse strain difference may be a cause. In fact, during experiment, we found that the gavage induced absorption, we found fetal resorptions both in vehicle and experiment groups, but the difference

of fetal resorptions showed no statistically significant difference. So, we speculate that the gavage may be the main cause of fetal resorption, other than the DEHP exposure.

Comment 5: Finally there is a recurring issue beginning on line 36 and throughout the manuscript, with the use of the term DSD. Disorders (or in some texts, differences) of sex development (DSD), as I understand them, encompass a range of mostly genetic conditions including congenital adrenal hyperplasia, trisomy of sex chromosomes, AR deficiency, etc. Phthalate exposure is associated with some endpoints related to Testicular Dysgenesis Syndrome (TDS), but not to any DSD of which I am aware. I suggest changing the references to DSD to make reference to TDS or simply persistent reproductive toxicity, unless the authors can cite a link between phthalates and DSD. Citations 3 and 4 appear to refer to reproductive malformations and kidney injury. The former are a component of TDS (cryptorchid testis, hypospadias), but they do not constitute DSD.

Reply 5: Disorders/differences of sex development (DSDs) include a broad range of congenital conditions in which the development of chromosomal, gonadal, or anatomical sex is atypical (Hughes IA, Houk C, Ahmed SF, Lee PA (2006) Consensus statement on management of intersex disorders. Arch Dis Child 91:554– 563. https://doi.org/10.1136/adc.2006.098319). The male genital malformations such as hypospadias, undescended testis, and testicular dysgenesis syndrome are typical phenotypes of male DSD. In this study, the term DSD was used to describe the developmental toxicity of prenatal DEHP exposure. Previous studies also found the relationship between prenatal DEHP exposure and male congenital genital malformation. (Sathyanarayana S, et al. J Clin Endocrinol Metab. 2017 Jun 1;102(6):1870-1878. doi: 10.1210/jc.2016-3837. PMID: 28324030; PMCID: PMC5470772; SCHIESARO M G, et al. Endocrine, metabolic & immune disorders drug targets, 2022, 22(7): 686-703.). According to DSD consensus, these malformations are DSDs. Therefore, we use DSD to describe the genital malformations.

Minor comments follow:

Comment 6:
Abstract
lines 52-58: This description does not include many details of the results. How is instant testicular injury defined? Histologically? What histological measurements? There should be some summary of the significant outcomes.

Reply 6: Thanks for your comment. For the word limitation of abstract, many details were not presented. I have added some details to state instant testicular injury the in the revised manuscript (see page 3, line 56-57).

Introduction
lines 77-79: There are also several rat studies that investigate the latent effects of in utero phthalate exposure in adult rats. These include decreased spermatogenic output and testicular histopathology in adults, and this is worth mentioning here.

Reply 6: Thanks for your comments. I have mentioned them in revised manuscript (see page 4, line 86-87).

Comment 7:
Methods
line 130: How were the DEHP doses chosen? Some justification would be useful.

Reply 7: the DEHP dose was based on experience from previous studies.

Comment 8: line 131: This is a very minor comment, but the term “euthanized” is preferable to “sacrificed.”

Reply 8: the term “euthanized” has been revised as “sacrificed.”

Comment 9: lines 141-143: There is evidence that testis weight does not vary with treatments that affect body weight during postnatal development in the rat (Chapin et al. 1992. Fundam Appl Toxicol 20:15.), because even caloric restriction sufficient to reduce body weight by <30% will not lead to a reduction in testis weight. Because of this, it is preferable to report testis weight, rather than testis weight/body weight index.

Reply 9: Thanks for your advice. We reported testis weight in revised manuscript according to your advice.

Comment 10: line 147: I do not believe the acronym WB has been defined, or is considered a standard acronym like PCR.

Reply 10: the acronym WB has been defined in introduction section (page 5, line 116), thanks for your comment.

Comment 11: line 178: What quality control criteria were applied for RNA samples (purity based on spectrophotometric ratios)? Were the samples analyzed for integrity using a bioanalyzer or similar?

Reply 11: The total RNA was adjusted using the bioanalyzer (Nanodrop ONE System, Thermo Scientific, USA), this has been stated.

Comment 12: line 183: Please provide the cycling parameters for the real-time PCR reactions and state whether negative controls (no reverse transcriptase, no template) were included in the reactions, how specificity was confirmed, and what threshold cycle cutoff was used, if any.

Reply 12: according to the manufacturer's instructions, the cycling parameters for the real-time PCR reactions was controlled at " stage 1 95°C 30 s; stage 2 95°C 5 s, 60°C 30 s", negative controls also were set. Due to word limitation, these details were not presented in manuscript. We beg your understanding.

Comment 13: lines 234-238: Please confirm that the statistical unit was the litter (i.e. littermates were not considered separate replicates), and indicate how littermates were handled.

Reply 13: Statistical unit was the littermates from at least 2 litters. We have illustrated how littermates were handled in figure 1.

Comment 14
Results
lines 264+: The effects described here, including reduced testosterone and testicular histopathology, do not constitute DSD. They do constitute testicular toxicity.

Reply 14: Poor development of the genital system (reduced testis, epididymis, and seminal vesicle weight; reduced AGI and penile length) is the phenotype of DSD. In revised manuscript, the DSD were defined (See page 12, line 279-283:This study found that the weight of the unilateral testis in mice without DEHP exposure was > 85mg. If the testis was poorly developed, the whole genital organs (including the epididymis, seminal vesicle, and prostate) were poorly developed. So, if the weight of unilateral testis < 85mg, DSD was defined). DSD is also presentation of testicular toxicity, so, the effect does constitute DSD, as well as testicular toxicity.

Comment 15: lines 284-286: I suggest removing the editorial comment that "it was not hard to find…", which is better suited for the Discussion. Additionally, I do not agree that this is a clear conclusion. The papers cited above (e.g. Johnson et al. 2012, Gaido et al. 2007) clearly show that phthalates disrupt fetal development of the mouse testis without altering testosterone. Persistent effects on morphology may be caused by changes in Sertoli and/or germ cell development, rather than subsequent differences in testosterone levels.

Reply 15: thanks for you comment and advice, I have revised the sentence (see page 13, line 299-301). We think that the change of testis morphology also refers to the Leydig cells. The change of Leydig cells especially the stem Leydig cells will bring about long-term impact in testosterone synthesis, and the change of Leydig cells may impact Sertoli and/or germ cell development. We admit that we can't prove that Persistent effects on morphology was caused subsequent differences in testosterone levels.

Comment 16: fig 3I: It is not clear what the arrows in these figures are meant to indicate. The arrow in the 500 mg/kg/d group appears to point to a gap in the seminiferous epithelium that may be an artefact (a similar gap is present in a large tubule in the lower right corner of the control. The authors state in the results that the tubule diameter is decreased, but this is not apparent from the figures, and it should be systematically measured (i.e. in all or a majority of randomly/systematically selected tubules from each testis section) to support the claim. The authors could also assess the testes for evidence of testicular injury, including germ cell sloughing, Sertoli cell vacuolization, retained spermatid heads, disorganized or atrophic tubules. None of these effects are apparent from this image (although some would not be visible at the present magnification).

Reply 16: thanks for your comment. I have added more illustrations in the figure legend.

Comment 17: line 304: the word "miraculously" should be removed. Also, there is prior evidence of a similar pattern of phthalate toxicity persisting into young adults, with spermatogenesis recovering later, in the rat (Dostal et al. 1988. Toxicol Appl Pharmacol. 95:104).

Reply 17: thanks for your advice, we have used "Curiously" to replace this word, according to another reviewer's advice.

Comment 18: lines 309-310: This is one possible interpretation, but it is also possible that this is not an accelerated aging effect, but a latent Sertoli cell dysfunction.

Reply 18: thanks for comment. The latent Sertoli cell may also be a cause of premature productive aging.

.

Comment 19: figure 4L: The same comments as above generally apply. There is not an obvious difference between these images. The arrows do not indicate atrophic tubules, and only possibly show some degree of germ cell loss, but this may be an artefact. If real, the same feature is pointed out in the control and treated samples, so the interpretation is unclear.

Reply 19: thanks for your comment. I have added more illustrations in figure legend. Here, we want to state that the diameter of seminiferous tubules was smaller and the lumen of seminiferous tubules was dilated, so we use “atrophic” to describe this phenomenon. We usually use “atrophic” to describe the damaged tissue even at the prepuberal stage, to state the effect of premature testicular aging, therefore, we use “atrophic”.

Comment 20: figure 5H: What explains the multiple bands on the Western blot for HSD3B2?

Reply 20: it may be due to the experiment technique detail, I speculate that the relatively higher voltage during electrophoresis may cause such a phenomenon, because HSD3B2 and HSD3B1 share similar molecular weight. But we don’t think that will impact the results.

Comment 21:
Discussion
lines 376-377: This is a somewhat minor issue, but the toxicity of DEHP is well-established, and the effect levels observed here (largely 500 and 1000 mg/kg/d) exceed human exposure levels. There is a body of literature on phthalate dose-response in utero and the effects of mixtures of phthalates that may be useful to interpret the risk to humans.

Reply 21: I agree with you, human receive various EEDs at any time, so the real exposure dose of EEDs is great. I have revised this sentence properly (page 18 394-396).

Reviewer C

Comment 1: The manuscript is well written and contains all the information necessary to understand the study. However, I suggest some improvements in the text:
- Line 37 to 42 of the background (However...). Structure it better so that it becomes more fluid.

Reply 1: thanks for your comment, I have revised this sentence (see page 2 37-43).

Comment 2: - Improve the conclusion by showing the novelty of the study. It was already expected that DEHP would induce testicular and other changes associated with sexual development, but what are the new contributions of this study (for example, WB)?

Reply 2: thanks for your comment, I have revised the introduction section (page 5, line 117-119), so as to state the new contributions.

Comment 3: - Line 137, complement the figure legend with more details of the experimental design.

Reply 3: thanks for your comment, I have added more details (page 6, line 151).

Comment 4: - Figure 3, I (HE staining), add an insert (with zoom) in the figure to better visualization of the hypogenetic characteristic.

Reply 4: thanks for your advice. Because the hypogenetic characteristic mainly presented with smaller diameter of seminiferous tubules, so I present the picture with such a magnification times, add an insert (with zoom) may have little function.

Comment 5- Line 304, replace Miraculously with Curiously.

Reply 5: thanks for your advice,we have done so.

Reviewer D

Comment 1: This manuscript uses protein expression of steroidogenic enzymes to explore the mechanism underlying phthalate-induced effects on the male reproductive tract after in utero exposure in mice.

• Overall, the findings of changes in steroidogenic enzyme expression observed (in either in situ testes or in vitro Leydig cell cultures) do not necessarily indicate that these changes are the underlying mechanism of phthalate-induced changes of the male reproductive tract.

Reply 1: thanks for your comment. We admit that the evidence is not enough. Testosterone plays a key role in male reproductive development and function; steroidogenic enzyme expression is important for testosterone synthesis; we found the steroidogenic enzyme expression showed change after phthalate exposure, therefore, we speculate that interfering with steroidogenic enzyme expression is one of the underlying mechanisms. Considering the testis did not show fundamental changes, our mechanism could at least explain the testosterone decrease induced by phthalate.

Comment 2: • Throughout the paper, terms used are not accurate. For example, the authors state that phthalates cause disorders of sexual development (some of which are genetic and include female sex organs), which are a different class of disorders than those classically seen with in utero phthalate exposure (see work by Richard Sharpe and many others). The testicular dysgenesis disorders typically seen in rodents are not accurately described or cited. One of the major limitations is that mice show different effects than rats and are probably not the best model to use – see https://doi.org/10.1093/toxsci/kfs206.

Reply 2: thanks for your comment. We have considered the mice may not be the best model, and the results also showed that the mice were more insensitive to DEHP, which is similar to the previous study. This is a limitation of this study, we have added this limitation in discussion.
Disorders/differences of sex development (DSDs) include a broad range of congenital conditions in which the development of chromosomal, gonadal, or anatomical sex is atypical (Hughes IA, Houk C, Ahmed SF, Lee PA (2006) Consensus statement on management of intersex disorders. Arch Dis Child 91:554– 563. https://doi.org/10.1136/adc.2006.098319). The male genital malformations such as hypospadias, cryptorchid testis, and testicular dysgenesis syndrome are typical phenotype of male DSD. In this study, the term DSD was used to describe the developmental toxicity of prenatal DEHP exposure. Previous studies also found the relationship between prenatal DEHP exposure and male congenital genital malformation. (Sathyanarayana S, et al. J Clin Endocrinol Metab. 2017 Jun 1;102(6):1870-1878. doi: 10.1210/jc.2016-3837. PMID: 28324030; PMCID: PMC5470772; SCHIESARO M G, et al. Endocrine, metabolic & immune disorders drug targets, 2022, 22(7): 686-703.). According to DSD consensus, these malformations are DSDs.

Poor development of the male genital system (reduced testis, epididymis, and seminal vesicle index; reduced AGI and penile length) is the phenotype of male DSD. In revised manuscript, the DSD of male mice were defined (This study found that the weight of the unilateral testis in mice without DEHP exposure was > 85mg. If the testis was poorly developed, the whole genital organs (including the epididymis, seminal vesicle, and prostate) were poorly developed. So, if the weight of unilateral testis < 85mg, DSD was defined. See page 12, 279-283). DSD is also the presentation of testicular toxicity, so, the effect do constitute DSD, as well as testicular toxicity.

Comment 3: • Lesions seen with phthalate-induced testicular dysgenesis (hypospadias, cryptorchidism, gonocyte effects, seminiferous tubule malformations) were either not evaluated or not shown.

Reply 3: we have evaluated the incidence of hypospadias, cryptorchidism, but the mice with DEHP exposure hadn't shown hypospadias, cryptorchidism, but presented with smaller testis, reduced testis, epididymis, and seminal vesicle weight; reduced AGI and penile length, we think these were also the manifestation of testicular dysgenesis.

Comment 4: • The authors claim to present evidence of "lifelong" toxicity but the oldest evaluation date appears to be in 6 month old animals. Similarly, the claim of "instant" toxicity does not correspond with examination at PND1, given than in utero exposure began at GD10.

Reply 4: thanks for your comment. This is one limitation of this study, we have stated in the discussion section.

Comment 5: • It is not clear what is meant by "reproductive aging," and there was no evidence of atrophy/degeneration shown.

Reply 5: thanks for your comment. "reproductive aging" meant aging of the reproductive system. In this study, we think the obesity, reduced AGI and serum testosterone, and increased epididymis and seminal vesicle weight, poor semen quality, atrophic seminiferous tubules are the evidences of "reproductive aging", though these evidences are not enough strong.

Comment 5 • Absolute testis weights should always be presented, since testis weight, like brain weight, is conserved. The sole use of the "testis index" is misleading.

Reply 5: thanks for your comment. We have presented testis weight according to your advice, see figure 2-4.

Comment 6: • Figure 2: Being able to detect a difference in penile length in PND1 mice seems challenging. Data for absolute testis weight and AGD should also be shown, especially given no difference in birth weights.

Reply 6: thanks for your comment. We have presented testis weight and AGD in figure 2.

Comment 7: • Regarding the PND56 animals, the authors claim there is insufficient development of the testis, epididymis and seminal vesicles. Histology should be shown for epididymis and seminal vesicles to support this. The histology images of the testis in Figure 3I do not show hypoplasia or any lesions as compared with the control, contrary to the interpretation by the authors. The gross images of the testes in Figure 3A show no appreciable differences; again, absolute testis weights should be shown. This reviewer sees no evidence of "hypogenetic" tubules in the images shown. Similarly, the authors claim that there was seminiferous tubule atrophy at 1000 mg/kg in the 6-month-old males, but this is not being shown in Figure 4.

Reply 7: thanks for your comment and advices. We have not done Histology of epididymis and seminal vesicles, because we ignored the importance of this results. The Figure 3I did not show great changes between the control and experiment groups, but it did have differences, such as smaller diameter and length of seminiferous tubules. Figure 4 also did not show great changes between the control and experiment groups, but it did have differences, such as smaller diameter of seminiferous tubules and dilation of the lumen of seminiferous tubules. We have presented data of testis weights to better show the difference in testis development.

Reviewer E

Comment 1: Summary: This manuscript reports experimental data from an in vivo mouse study with maternal oral exposure to DEHP during pregnancy. The authors report on male reproductive endpoints in offspring at three timepoints.
Comments:
1. The citations of previously published studies on this exact topic is severely lacking. There have been numerous developmental studies examining the male reproductive effects of in utero exposure to DEHP specifically, mostly in a rat model. It appears that many of these studies are not cited or included in the Discussion of the present

results. These include:
Blystone et al. (2010) doi: 10.1093/toxsci/kfq147
Gray et al. (2009) doi: 10.1093/toxsci/kfp109
Li et al. (2013) doi: 10.1016/j.etap.2012.10.006
Moore et al. (2001) doi: 10.1289/ehp.01109229
Parks et al. (2000) doi: 10.1093/toxsci.58.2.339

Further, Dorman et al. (2018) doi: 10.1080/10937404.2018.1505354 published a systematic review and meta-analysis specifically on the male AGD effects of prenatal exposure to DEHP and reported 6 human epidemiology studies and 19 animal studies. Clearly, this has been a well researched topic. The authors need to re-review the literature and revise the manuscript to reflect the much more advanced state of knowledge that is actually present, compared to the lack of knowledge the authors report in this draft manuscript.

Reply 1: thanks for your comments and advice. We have already cited many references. Certainly, the references you mentioned are very valuable for better discussion. But this study mainly cited recent references (< 5 years), we admit that these references may not reflect the researched topic perfectly. Our research topic is to explore the mechanism of DSD and adult male disorders by mice model, so we mainly cited references about these. We revised the manuscript according to your advice (see page 4-5) and added many new references you mentioned (Dorman et al. (2018) doi: 10.1080/10937404.2018.1505354 in the introduction,).

Comment 2: 2. The Methodology regarding how sample sizes were handled is inadequate. The Methods do not report how many pregnant mouse dams were dosed at each dosing level. Further, the statistics were not done appropriately by accounting for litter-based effects. The authors report the total number of offspring measured at each dose level, however this is inappropriate. The appropriate sample size is the number of litters, with data analyzed with individual offspring nested within litter. For example, the Methods report that 10 males at 0 and 1000 mg/kg were euthanized on PNM6 – are these 10 offspring from the same litter or different litters? If different litters, how many different litters? I very much doubt that there were 30 to 40 litters per dose group as shown in Figure 1.

Reply 2: thanks for your comment, in this study, the sample size means the offspring from 10 (0 and 1000mg/kg/d groups) or 8 pregnant mice (100 and 500mg/kg/d groups), this has been illustrated in the revised manuscript. However, for prenatal exposure experiments, it is impossible to treat the mice from the shared mother

mouse, which may impact the results. Mice in each experiment were from more 2 litters, which may reduce the error. We have added more details in figure 1 and method section (see page 6, line 144-145)

Comment 1: 3. The data for AGD and male sex accessory tissues are all reported as and “index” (i.e., relative organ weight). This is not appropriate for male sex accessory tissues. The index values can be calculated and reported, but the absolute weights must also be reported. Further, for AGD, which is a 1 dimensional measurement, the appropriate calculation of AGDI accounts for the 3 dimensional nature of body weight by using the equation AGDI=AGD/(cube root of bodyweight).

Reply 3: thanks for your advice, we have add data of AGD in revised manuscripts.

4. The Methodological standard for measuring AGD, particularly in neonatal rodents is the use of a stereoscope and ocular micrometer. I find it particularly difficult to believe that an accurate AGD could be measured in a 1 day old mouse pup using a handheld caliper.

Reply 4: thanks for your comment and advice. The penile and anal opening are very obvious in neonatal stage for they have no hair. It’s not hard to measure AGD in neonatal mice using electronic vernier calipers, whose measurement accuracy reach 0.001mm. Certainly, we admit that there exists deviation. We usually measured the AGD three times and used the mean value.