Dear Authors,

Your preprint entitled "Sex-specific relationship between maternal and neonate cortisol in a free-ranging large mammal" has now been reviewed and the reviewers' comments are appended below. As you will see, both reviewers are positive about the study, notably regarding the way it is written, its ecological relevance and its scientific robustness. Additionally, the code was easy to read, ran smoothly, and the results from the code matched those presented in the main text (but see below for some small omissions and discrepancies). Yet the reviewers have several comments that need to be addressed carefully before your preprint can be recommended.

One main comment shared by both reviewers is that the research question and the biological context are not clear. More particularly, regarding potential sex specific effects, apart from mentioning the possible existence of such effects, I could not see any prediction made in the introduction regarding the direction of such effect, and the findings of the work cited when mentioning the possibility for sex differences (Braithwaite et al., 2018; Fishman et al., 2022; Liu et al., 2001) could be briefly developed.

If the authors had no predictions apart from a possible existence of sex specific effects, then it should be clearly stated.

1) We had no clear a priori predictions and thought that it was clear by highlighting that this is an exploratory study. However, exploratory studies can of course still have some predictions and we therefore agree that this needs to be clearly stated. Thus, we have now added such a statement at the end of the introduction (lines 95-96).

I also agree with reviewer 1 on the fact that the statistical procedure to investigate the sex specific effects is not common. They may be a rationale as for why this approach was taken, in which case it needs to be justified and explained to the reader step by step (first a t test, then separate analyses for each sex, then both sexes analyzed together to estimate the sex difference i.e., interaction), with references recommending it if any.

2) Although uncommon, this is the order in which we did the test, and we cannot change that. The reason was that we were reluctant to run a model with an interaction term, because of our low sample sizes. Interaction terms require massive sample sizes, ranging from 4 to 16 times the sample sizes of main effects (see Leon & Heo, 2009, https://doi.org/10.1016%2Fj.csda.2008.06.010; or Gelman, 2018, https://statmodeling.stat.columbia.edu/2018/03/15/need16/). This is why we did not run these first. Given the results from the separate models, however, we decided to run an interaction model afterwards (hence labelling it as *post hoc*), because we thought it would be useful to further understand the patterns in our data. We have now extended the first paragraph of the statistical analysis (lines 162-167) to give the overview in advance, which hopefully makes the steps clearer.

I also have some more specific comments:

Line 55-56: it seems inaccurate to "label" parental effects as maternal effects solely because maternal effects are expected to be stronger than paternal effects. I'd recommend to clarify, or rephrase as e.g., "Parental phenotypes can be drivers of offspring variation, and in many mammalian species, these parental effects are often assumed to be mainly maternal effects since it is the mother that mostly takes care of the offspring." or any other accurate rewording.

3) We agree and have changed wordings in the main text.

Line 135: is there any estimate of individual repeatability of this measure (from this data or previously published)? Could it be mentioned in the dataset which data values come from one sample vs averaged from several samples (ideally, the raw dataset with multiple measures per female and their respective sampling days could be provided)?

4) Sure! We have now uploaded the raw data of the faecal samples as well. From these, one can calculate individual repeatability. We have also provided a script ("Fecal exploration.R"), where we comply with some question from this review. We have added a repeatability test here and as expected, repeatability is quite low for FCMs. Depending on the exclusion or inclusion of the outlier that was mentioned by one of the reviewers (see reply #24), the repeatability is estimated at 0.149 ± 0.106 SE (with outlier) or 0.126 ± 0.1 (without outlier).

Line 138: how was this value defined as "extreme outlier"? Is it from the goodness of fit diagnostic of the male linear model? Perhaps "outlier" is enough (no need for "extreme"), especially since the value does not look particularly unrealistic?

5) We decided to remove this value based on the distribution of the values. Neonate GC levels varied from 6.5 (min) to 29.5 (max, outlier), with the second highest value being 18.3. The first quantile was 12 and the third was 15.40, giving an IQR of 3.4. That means that the outlier was more than 4*IQR above the third quantile, where normally more than 1.5*IQR is considered as a potential outlier. We do not know whether this is due to a sampling error, or whether something else was deviating in this individual. We are, however, skeptical because of the above reasons and do think it is wise to remove this point when investigating general patterns, especially considering our sample size. We agree with the recommender though, that it may be better to remove the term "extreme". We changed the wording and have also added an extra clarification that the value was removed because it was far out of range (line 152-154). There are further explanations regarding this outlier in the manuscript, but we detail them in the specific replies below (#6 and #9).

Line 140-141: although the estimates of the effects are similar, it seems that the interaction is no longer statistically clear when discarding the "outlier"? Please confirm and clarify.

6) This was corrected but we missed this in the previous version, thanks! The estimate of the interaction was actually elevated, but so was the variance. That makes sense because it is an outlier. We have altered the text now (lines 155-158) and have also included the interaction term of the model with the outlier in the result (lines 201-203). Furthermore, we plotted the outlier now in the figure as well, as suggested by the

recommender in a comment below. We hope that it is clearer now and that it is also clear why we removed the outlier in the first place.

Lines 146-147: please add these t-tests in the R code.

7) These should have been in the code from the start, our apologies. They are now added to the code.

Line 169: also specify the value of the "outlier" here.

8) Added.

Figure 1: start the y axis at zero. I would also suggest showing the outlier in the figure (e.g. in light gray and/or with another shape) and mentioning in the figure legend that this data point was discarded from the main analyses.

9) Great suggestion! We did that. The outlier is now shown as a red diamond in the figure. We think that this figure again shows why it makes sense to remove this value, as it is not representative of the main pattern.

I look forward to reading the revised version of this preprint.

Review by anonymous reviewer 1, 03 Jul 2023 11:57

The preprint provides new insights on the relationship between maternal glucocorticoid and foetal glucocorticoid levels. The sex-specific pattern of the relationship found in the preprint is ecologically-relevant and robust, especially in this free-ranging large mammal, the fallow deer.

However, I find it a bit hard to grasp the research questions and the whole picture from the introduction part. I would suggest to reformulate and add several key information (see below).

Title

The authors used 'neonate cortisol' in the title but 'foetal glucocorticoid' in the abstract and the main text. Given the description of how the authors sample the young, I think 'foetal glucocorticoid' should be used throughout the text, including the title.

10) To us it was a bit unclear whether this comment relates to the usage of 'neonate' or 'cortisol' (or both). But we do agree with consistency in the terminology and agree that, at least in the title, it would be better to take over the suggestion by the reviewer. So we have made that adjustment now to the title. We also evaluated throughout the manuscript whether the use of 'neonate' or 'cortisol' was appropriate and have adjusted needed.

Abstract

I miss a sentence clarify the biological relevance/functions of the GC in determining the offspring phenotype in the abstract.

11) Agreed that this would be necessary. We have made some alterations to the start of the abstract to include this (line 41-43).

Introduction

L. 57 References need to be added for the first sentence of the introduction.

12) We added references for the first part of this sentence. We have further improved this sentence (see reply #3) and hope that it is better now.

L.57-61 References in these sentences are only concern the effects of maternal stress/GC on offspring phenotypes. Yet, the authors are talking about maternal effects in general here. Please see and add reviews from Groothuis et al, 2005, 2010, 2015 and others that reviewed more broadly on the maternal effects.

13) Most of the reviews of Groothuis et al were focused on avian species, with a stronger focus on androgens. This is why they were not included in the first place. But we agree with the reviewer that these are still highly relevant and should be included, so we have added Groothuis et al 2005 & 2019 to the introductory paragraph (lines 61-62).

L.62 I doubt whether the authors are looking as acute stress or chronic stress in this study, since the GC accumulated in maternal fecal samples and foetal hairs would reflect a long term and average stress level.

14) Good point. We removed "acute" from the first sentence. It should be more accurate now.

L.66 citation error "(e.g.8)", please correct

15) Corrected.

Materials and methods

L.124 citation error "(e.g. 15,25)", please correct

16) Corrected.

L.134 it is not clear in the introduction or here in the methods whether the Fallow mother only produce one fawn per pregnancy.

17) We agree that this is important information and we have therefore added it to the first section in the methods (lines 105-106).

L.146-147 A confusing sentence, please reformulate. and I wonder if the authors are comparing the GC levels or the FCMs of the mothers? If they are comparing the GC levels of the mother, an explanation of how they convert FCMs to GC levels is necessary. In addition, it is not clear what is the relationship between maternal FCMs and GC levels at this point. How well can FCMs represent the GC levels of the mothers?

18) We understand the confusion. We have now edited this sentence, clearly specifying that we compare maternal FCMs. Furthermore, we have placed the last part of the sentence in parentheses after "t-test", to more clearly emphasize that this relates to the t-test. Furthermore, our method of using FCMs is validated for fallow deer, indicating

that it is suitable for our purpose (https://doi.org/10.1007/s10344-010-0401-1). We have also added more detail on what the validation means in lines 125-128. We hope this clears the confusion.

L.151-153 Where is the results/parameters from the model assumption check?

19) These are displayed and shown in our R-script. We have indicated this in the main text and have also provided the DOI to the document (line 173).

L. 154-155 Specify include these factors as what in the model.

". During the preliminary analysis, we included the number of days between the collection of mothers' faeces and the day of fawn birth in our models."

20) We have added that this was an additional explanatory variable in the models (line 176).

L.157-158 Please give the value of AICs or delta AIC of the models compared, instead of "higher AIC".

21) These are now given in the main text (line 179).

L.160-164 It is counterintuitive to me that the authors ran a post-hoc test by "including both sexes" in the same/previous model. On the contrary, a linear model with neonate GC as response variable and maternal FCMs, fawn sea and heir interaction as explanatory variables should be the overall/main model. Other potential covariates such as the number of days between the collection of mothers' faeces and the day of fawn birth should also be tested in this model. Thereafter, a post-hoc testing the relationship of neonate GC and maternal FCMs separately for two sexes should be carried out.

22) We have already given an elaborate response regarding the order of tests we did (reply #2). Unfortunately, we cannot change the order or rationale of our decisions at this stage. The decision to run an interaction model was made *after* seeing the results of the separate model, hence the term post-hoc. We were reluctant to run a more complex model due to the low sample size, and that is also why we are still reluctant to add more variables to the model. We definitely do understand where the reviewer is coming from. We have decided to upload the raw faecal data, along with the code that explores the relationship between the hour of collection and FCM level. Also, we did the variable the reviewer mentions to the interaction model (also see reply #22) and it led to a higher AIC value, which is now given in the main text. So, we have little reason to think that factors such as the time of day or days between sample and birth have a clear effect on our analysis. Thus, with our limited sample size, we chose not to further complicate the model.

Results

L.172- 178. Like I suggested in the method section. If the statistical analysis changed, this corresponding result section should also be reformulated.

23) We have indeed made some changes to this section. This is already highlighted in reply #6.

Tables and figures

Figure 1B, I am concerning the potential outliners in this panel. There is one data point where the maternal FCM is above 600 ng/g. Please check and re-do the statistics.

24) We understand the concern of the reviewer. This individual is one of those that was sampled multiple times, meaning that this data point is also an average of multiple samples. Since we don't have large sample sizes, we chose to keep the individual in the analysis while seeking another solution. The high value is caused by one of its samples, with a read out of 1318 ng/g, whereas all other values were <800 ng/g (see raw values and associated script from reply #4). This poses the same concern as with the outlier that we removed from the analysis before, namely that the single high value of 1318 ng/g is more than 5*IQR above the third quantile. We therefore decided to remove this one value from the raw faecal dataset, while keeping the other samples from this individual and using those for her average value. This is also mentioned in the main manuscript (lines 147-150). We hope that this satisfies the concerns raised here.

Discussion

L190.-194. This should be consisted and move to introduction, as it is necessary for the reader to know why the authors would expect a sexual difference in the first place.

25) We agree that it was not fully clear why we would expect sex differences in the introduction. We have therefore elaborated on some previous studies regarding sex differences to make it clearer (lines 72-77).

L.194-195 I don't fully understand the sentence until I read the paragraph below. Need to reformulate these information.

26) We have reformulated this sentence and hope that it is much clearer now (lines 219-220).

L.204-207. It seems that it is the male foetusese that control its own GC exposure environment instead of the mother actively adjust how much GC her foetus would expose to. This is very interesting and important. I would suggest the authors add a stand-alone sentence to declare this.

27) We don't think that our findings suggest that male foetuses can control their GC exposure, but we do think that they may interact or influence it. More so, that they have mechanisms that allow them to be less affected by maternal levels, some of which we discuss in the manuscript. We have added a clarification sentence that the offspring can also influence the GC exposure in lines 262-265.

L.211-214 It seems that the authors are suggesting that the GC level of the fetal are extraneous (that is to say. maternally-derive). If this is true, the authors should clarify by which point the fetal can also have endogenous GC? To what extend this endogenous GC would affect the fetal development comparing to the maternal GC. If it is a mixture of maternal GC and endogenous GC that influencing the fetal development. The authors should clarify how maternal GC would affect the embryonic endogenous GC?

28) Gitau et al (https://doi.org/10.1016/S0140-6736(05)60824-0) estimate indeed that about 40% of the foetal GC levels is maternally driven, though this is work in humans. It really is an interaction. See reply #27 for our addition regarding this.

Furthermore, it is still not clear to me what is the relationship between maternal FCMs and her GC levels. How well can FCMs represent the GC levels of the mothers?

29) We agree that this point was not clear before and hope that our changes discussed previously have made this clearer (see reply #18).

Review by anonymous reviewer 2, 24 Jul 2023 14:38

I was pleased to review the article entitled "Sex-specific relationship between maternal and neonate cortisol in a free-ranging large mammal" written by Amin and colleagues.

This manuscript describes an innovative non-invasive ecophysiological method used to link foetal and maternal glucocorticoid levels in mammal. The manuscript is quite well-written, and easy to read and to understand.

I have one crucial comment to address to the authors before recommending their manuscript. I'm not entirely sure to understand whether this manuscript aims 1) to describe a new methodology validated to work on maternal effects in mammals, using the deer or 2) to highlight a sex-difference in the link between maternal and fetal glucocorticoid levels in one given mammalian species through the use of an innovative method. Indeed, on one side, technical and methodological details are missing to be considered as a methodological paper (and some are described in the supplementary material or by citing previous work). On the other side, biological context and discussions would be clearer to be considered as a research paper.

30) Our aim is the latter, to highlight a sex-difference in the relationship between maternal and foetal GC. We hope that, with all the changes made in this revision, that it is clearer that this is our aim.

To help increasing the impact of this manuscript, here are some questions I would recommend the authors to answer in their manuscript:

- For the technical/methodological aspects:

Regarding the "circulating glucocorticoids", did you check that glucocorticoids levels measured in feces are representative to glucocorticoids levels circulating in plasma?

31) Yes, this has been validated for our species in a previous study. We have made this clearer now in our main text (see reply #18).

Because you have wide variations in your sampling protocol, did you verify/consider the potential effect of the time of the day on GC variations? How were stored the feces before being placed in the freezers? How long did the feces were kept in a cooler bag before being placed in a freezer? Did you consider the effect of the duration between collection and freezer storage on GC measurements? How did you determine the gestation stage? Could you confirm that samples were collected before determining the mother-fawn pairing (June vs July respectively)? How did you select the individuals to sample in such scenario?

32) Most of this information is already in the manuscript. We found no potential effect of time of day, for which we also provide the raw data and code ("Faecal exploration.R") now. Faeces were put into zip locked bags and kept in a cooler bag until stored in the freezer, which was always within a few hours (lines 117-120). Gestation state was similar for all does (late gestation) since all the does were sampled in the same period in May (line 113). We checked whether the number of days between sampling and birth played a role, but found that it only made our models worse (higher AIC; lines 175-180). Finally, the faecal samples were indeed collected about 1-1.5 months before the pairing (late

May vs July), i.e. before we knew the pairs. We followed our standard protocol, as also explained in S1, for the pairing and linked the maternal FCMs and neonatal GCs afterwards.

-        For the biological/ecological aspects:

In the introduction, I would clarify why studying GC (and/or the associated stress?) would be relevant in the gestation context. Moreover, I would describe in more details how maternal GC levels play a key role during gestation and can affect offspring phenotype. In addition, whether the aim of your manuscript is to deal with a GC sex-response, I would develop some examples already found in other vertebrate species.

33) These are now added (see reply #25).

Moreover, I would detail how and why can maternal circulating and fetal accumulated glucocorticoid levels be related. Which maternal information is obtained from this relationship between GC (which are accumulated over a couple of weeks) and FCM (which are the immediate/present levels, i.e. when the foetus is already covered in fur, meaning after the synthesis of the hair and the associated GC accumulation).

34) FCM levels are not immediate/present level, but rather an accumulation over several hours (but excreted with a species-specific time delay). We understand that these metrics are on different time scales; however, we aimed to make it comparable by carefully synchronizing our sampling time, i.e. making sure to sample females at the stage of gestation when the foetus is covered in fur. We have further highlighted that again in the methods (lines 114-116).

How maternal FCMs vary all along the gestation?

35) For red deer FCMs increase along the gestation (https://doi.org/10.1002/ece3.1945) as in many other species. We expect the same pattern in our population of fallow deer. Regardless, since all our samples are collected within a period of two weeks (whereas fallow deer gestation is ~9 months), it is unlikely that there is a seasonal effect. This is confirmed in our recently added R-script "Faecal exploration.R".

The discussion part would be reorganize or slightly rewritten to be more precise and clear regarding the different hypotheses that could be emitted regarding both the link between offspring GC and maternal FCMs as well as between female and male progeny.

36) We made some minor adjustments to the discussion (highlighted in previous comments) and also the introduction regarding our predictions. We now emphasize that we did not have clear predictions, and hope that the manuscript is much more clearer now.

I also have a couple of comments line-by-line:

All along the manuscript: I would use the singular when designating 'parental/maternal phenotype'.

Line 42: foetal levels are

37) We kept this as it is, because it reads as "It is not completely understood".

Line 80: there is an extra space at the beginning of the paragraph

38) Fixed.

Line 170: no clear sex difference

39) Fixed.

Line 187: than

40) Fixed.

Line 191: per year

41) We don't quite understand this comment since the text already states "per year".

42) Finally, we have made some grammar/style improvements throughout the manuscript to further improve it. These are also visible in the track changes version that we supply with it.