**Peer Review and Communication History**

**MS Title:** Hormonal Contraception and Sexuality: Causal Effects, Unobserved Selection, or Reverse Causality?

**Author Names:** Laura J. Botzet, Tanja M. Gerlach, Julie C. Driebe, Lars Penke, and Ruben C. Arslan

**Submitted:** March 6, 2021

**Editor First Decision: Revise & Resubmit**

July 16, 2021

Dear Laura J Botzet,

Thank you for submitting your work to Collabra: Psychology. I sent the paper to two experts who provided outstanding feedback as you will soon discover. These individuals are extraordinarily well qualified to review this paper and I thank them for their service to this journal. I independently read the paper and then consulted the comments from the two reviewers.

As you will read below, the reviewers were positive about the work. I concurred with their positive sentiments. The paper was straightforward, clear, and interesting. My reading is that the Reviewers raised a few questions and offered some constructive suggestions but I believe all of their comments and concerns are addressable. Thus, I would like to move forward with a revise decision. I do not plan to send a revision back out for review and thus the time lag to the final decision should be greatly reduced if/when you submit the revised manuscript. My plan would be to review the new version and the letter of response and then make a final up or down decision. I still reserve the right to consult with one or both of the reviewers if anything drastically changes.

The reviewers did an outstanding job and you should address each of their concerns either in the revised text or in the response letter. Addressing their concerns will make this a stronger paper. I will highlight some issues that came up as I read your paper. I acknowledge that you might disagree with some (or all) of these points so feel free to pushback against any suggestions you believe will harm your work. Just describe your counterpoints in the letter.

1. I understand the value of analyzing existing data and I believe the paper is straightforward about this issue. I think it would be valuable to detail existing papers from this dataset in a bit more detail and to be clear about all of the measures in the dataset (see Reviewer #1, esp. points #4 and #7).
2. Reviewer #2 had several technical questions that deserve your attention.
3. Most of my reactions were minor comments on style and phrasing.

a. Can you define selection effects, attrition effects, and reverse causality on page 5 and perhaps ground the brief explanations with reference to this substantive area. This does not need to be extensive but just outlined for readers right away to anticipate the material on page 10.

b. I agree with Reviewer #1 about the need to provide a bit more support and justification for the hypotheses about hormonal contraceptives and sexual behavior given the outlet.

c. I would refer to personality traits rather than to the broader concept of personality on page 10.

d. I think some reviewers might need a bit more orientation to DAGs. Again, this does not need to extensive but just something a bit more detailed. I suspect readers here might be better versed with old school path diagrams so a few sentences about the similarities and differences would be helpful.

e. I had a similar reaction to Point #2 from Reviewer #1 about the causal language. I think I understand the various arguments about causal language so I am not insisting on anything but I think using associative terms would be cleanest and mitigate any reactance in readers.

f. I think a few passages defining weakly informed priors for non-Bayesians on page 20 would be helpful.

g. Would including percentages for income and relationship duration in Table 2 make sense?

h. Please check the first sentence of the Conclusion. I think the restatement of the results is incorrect (or perhaps my caffeine levels waned).

1. I like to see formal statements of constraints on generalizability in papers at this outlet. Would you consider formally including such a section and citing the Simons et al. paper? This would not need to be extensive and could pull from some of the material already in the Limitations section. I also wonder if it would be useful to note that other “controls” not measured in this study could result in different conclusions.

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. Perspectives on Psychological Science, 12(6), 1123-1128.

1. The paper is long. That might be unavoidable but I would just ask you to consider anyways to make this work more streamlined for readers.

Those were the most salient issues that occurred to me in reading the paper and the reviews. Please feel free to contact me for any clarifications.

To access your submission account, follow the below instructions:

1. login to the journal webpage with username and password
2. click on the submission title
3. click ‘Review’ menu option
4. download Reviewed file and make revisions based on review feedback
5. upload the edited file
6. Click the ‘notify editor’ icon and email the confirmation of re-submission and any relevant comments to the journal.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org. Could you have the revisions submitted within 90 days? If you cannot make this deadline, please let us know. Good luck revising this work. Thank you for trusting us with your paper.

Sincerely,

Brent Donnellan

# Reviewer 1

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

**General Comments:**
This manuscript investigates three core questions: selection effects regarding choice of contraceptive method (i.e., no/non-hormonal vs hormonal), potential effects of contraceptive method (and contraceptive congruency) on relationship quality and sexual functioning, and sensitivity of these effects when unobserved confounders are considered.
The authors report that a complex model of selection effects did not substantially improve on the simple model when predicting hormonal contraceptive use or change in contraceptive method. They also find no evidence for an effect of hormonal contraceptive method, congruent contraceptive use, and their interaction, regarding partner attractiveness, relationship satisfaction, sexual satisfaction and libido. They do find evidence for positive effect of hormonal contraceptives on vaginal intercourse, and a negative effect of hormonal contraceptives on masturbation frequency. Potential observed and unobserved effects were considered.

I think the paper is well written overall, and the results (although they are not what the authors predicted) demonstrate commitment to good science. I also want to note that I really appreciated the authors providing access to the analysis and codebook. The contraceptive and ovulatory literature has a need for large scale studies that utilise conservative statistical techniques and I enjoyed being able to see the exact measures and confirm what was described in text. I think this manuscript could be published with some revisions.

**Major Comments**
Background and Literature Review

1. The authors discuss relevant literature regarding hormonal contraceptives, the congruency hypothesis and cycle shifts (e.g., partner preferences, attractiveness ratings, etc). However, more detail could be provided regarding why the authors hypothesize that hormonal contraceptives may impact sexual functioning. I recommend this issue is addressed because the audience of this journal may not be familiar with hormonal contraceptives and theories relevant to the research aims.
2. The authors write that the terms “effects of hormonal contraceptives” and “effects of congruent contraceptive use” are used to remain constant with previous research (footnote 2, page 5). We recommend the use of language that is appropriate for correlational research, to help improve the accuracy and quality of the literature.

Method
3. The authors provide sufficient evidence regarding participant demographics overall. However, more information regarding hormonal contraceptive (HC) type could be included in text. I recommend that the authors include (from the supplementary documentation) information about the type of HC (e.g., pill, patch, implant) and the proportion of each type in the current dataset. If available, information regarding estrogen dose (particularly regarding pill users) is also valuable.
4. The data within this manuscript was derived from an existing dataset. Can the authors confirm whether the measures analysed in this article been published elsewhere?

Analyses
5. Some measures (i.e., frequency of vaginal intercourse and masturbation) were collected daily but were aggregated to produce a single outcome value. Can you explain why the decision was made to aggregate the data, rather than using the daily measures?
6. Fertility is known to be associated with extra- and in-pair desire in naturally cycling women (Arslan et al, 2018; cited in the manuscript). I wonder if fertility should be considered in the sensitivity analyses.

Results
7. In the section regarding ‘Further Unobserved Confounders’ (page 49), the authors write that some variables (i.e., sociosexual orientation (SOI), satisfaction with body image and life satisfaction) were not included due to the potential for colliders or mediators. The original codebook includes SOI, however it was unclear whether satisfaction with body image and life satisfaction were recorded in the original diary study. Can the authors clarify whether these variables were available in the dataset, or if their exclusion was justified retrospectively?
8. Footnote 3, page 9, justifies the inclusion of all HC users as one group, while highlighting that previous research has often included only oral contraceptive users. Follow up analyses, contrasting data from oral contraceptive users and naturally cycling women, would be a valuable addition to the analyses.

Minor Comments and Suggestions

* On page 6, the authors write “Hormonal contraceptives contain synthetic progestin, which suppresses the natural production of estrogen and progesterone.”. It would be more accurate to write that hormonal contraceptives contain synthetic progesterone (i.e., progestin), as all progestins are synthetic progestogens.
* In addition to Wiegratz et al (2003), page 6, a useful reference for testosterone and oral contraceptive use: Zimmerman, Y., Eijkemans, M. J. C., Coelingh Bennink, H. J. T., Blankenstein, M. A., & Fauser, B. C. J. M. (2014). The effect of combined oral contraception on testosterone levels in healthy women: A systematic review and meta-analysis. Human Reproduction Update, 20(1), 76–105.
* For clarity, on page 20, when referring to the outcome measures, clarity would be enhanced if the authors explicitly state which measures were from the baseline (as they have for the diary measures).

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

# Reviewer 2

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

## Summary

This paper aims to assess the causal effect of hormonal contraceptives on several behavioral and psychological outcomes, such as perceived partner attractiveness, relationship satisfaction, frequency of vaginal intercourse and masturbation. The paper uses a survey data of 1,179 women, regression analyses to adjust for observed confounders, and sensitivity analyses to assess robustness to unobserved confounders. For most outcomes, it concludes that it does not find much evidence for an effect. The two exceptions are the effect of contraceptives on sexual and masturbation frequency, which are robust to the inclusion of observed confounders, and appear to be somewhat robust to potential unobserved confounders. However, reverse causation and attrition biases cannot be ruled out.

## Comments

This paper is well written and very transparent in its methodological choices. Causal inference with observational data is difficult, and most social sciences, in particular psychology, have been doing it casually, either by naively avoiding causal language (but interpreting the results causally) or simply by making causal claims ignoring the role of causal assumptions altogether.

This paper is refreshing in its transparency. It clearly states its causal goals, it explains the rationale for controlling for certain covariates and avoiding others (eg., potential colliders and mediators), and it is not afraid to point out limitations of what can be inferred from the data. The supporting information provided is also very through and complete.

There are some technical issues that still need to be addressed as I detail below, but overall I am positive about the paper.

1. Be more specific about the estimand

Although the paper is clear regarding its causal goal, it still does not explicitly mention which causal effect it is trying to estimate. For instance, is it the average treatment effect (ATE) E[Y1−Y0]? This is not a particular limitation of this paper, but a general limitation of most papers in the social sciences. Only recently this has been gradually changing due to advances in causal inference research and education. I suggest the authors be even more specific about their target causal quantity. As a reading suggestion, the following may be useful:

Lundberg I, Johnson R, Stewart BM. What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory. American Sociological Review. 2021;86(3):532-565. doi:10.1177/00031224211004187

1. Choice of effect size, priors and ROPE

I personally like the choice of the authors of not focusing on statistical significance, and instead focusing on estimating effect sizes and regions of practical equivalence. I may have missed, however, which effect size is being reported: for instance, is it a regular or standardized regression coefficient? This is particularly relevant for a Poisson or logistic model, as I’m not sure the default “effect size” would directly reflect the ATE, E[Y1−Y0], for instance (in case that’s the target of inference).

Relatedly, the authors simply mention they used the default priors of brms, but it would be more transparent to explicilty mention what these priors are, and perhaps some discussion. Finally, some discussion regarding the choice of the region of practical equivalence would be useful, instead of simply following Kruschke.

1. Sensitivity analysis and comparison to observables

Sensitivity analyses in observational studies are extremely important, since we rarely measure all relevant confounders. The authors are to be applauded for their effort in this direction, by directly quantifying the robustness of their estimates to the threat of unobserved confounding.

I have two technical comments on the sensitivity of the effects on the two count outcomes, frequency of vaginal intercourse and masturbation: (1) The sensitivity by Cinelli and Hazlett (2020) targets OLS regression coefficients. Since your treatment is binary, regular OLS still works for estimatings a conditional expectation, even if the outcome is a count. Thus, you should use OLS estimates for the sensitivity analysis, instead of the Poisson GLM estimates (it is unlikely this will change the main results, but it still should be corrected). (2) The comparison of relative strength of unobservables vs observables cannot be done by comparing changes in the robustness values directly. Instead, you should use the formal bounds on the strength of confounding as in Cinelli and Hazlett (2020), section 4.4.

Regarding the use of OLS for estimating conditional expecations with binary treatments with non-continuous outcomes, see Mostly Harmless Econometrics (Chapter 3), or, for instance, the recent discussion in Psychology provided by Gomila (2020). There the focus is logistic regression, but similar reasoning can be applied to Poisson regression:

Gomila R. Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. Journal of Experimental Psychology: General. 2020 Sep 24.

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

**Author Response**
Oct 4, 2021

Title: Hormonal Contraception and Sexuality: Causal Effects, Unobserved Selection, or Reverse Causality?

September xth, 2021

Dear Dr. Donnellan,

Thank you very much for the helpful and constructive feedback on our manuscript on hormonal contraception and sexuality and for the opportunity to revise and resubmit the manuscript to *Collabra Psychology*.

We were happy to hear that you and two reviewers found our study of interest. We greatly appreciate yours and the reviewers’ insightful comments and valuable feedback on ways to further improve the manuscript. Thus, we have carefully considered and responded to all points raised by you and the reviewers.

We have uploaded two versions of the manuscript, one in which all changes from the initially submitted version of the manuscript are marked in blue to ease the review process and another one without marked changes.

We think that these changes have further strengthened our manuscript and would be grateful if you consider it for publication.

We have responded to each suggestion below and numbered comments to make them easier to identify.

**Editor’s remarks:**

**Thank you for submitting your work to Collabra: Psychology. I sent the paper to two experts who provided outstanding feedback as you will soon discover. These individuals are extraordinarily well qualified to review this paper and I thank them for their service to this journal. I independently read the paper and then consulted the comments from the two reviewers.**

**As you will read below, the reviewers were positive about the work. I concurred with their positive sentiments. The paper was straightforward, clear, and interesting. My reading is that the Reviewers raised a few questions and offered some constructive suggestions but I believe all of their comments and concerns are addressable. Thus, I would like to move forward with a revise decision. I do not plan to send a revision back out for review and thus the time lag to the final decision should be greatly reduced if/when you submit the revised manuscript. My plan would be to review the new version and the letter of response and then make a final up or down decision. I still reserve the right to consult with one or both of the reviewers if anything drastically changes.**

**The reviewers did an outstanding job and you should address each of their concerns either in the revised text or in the response letter. Addressing their concerns will make this a stronger paper. I will highlight some issues that came up as I read your paper. I acknowledge that you might disagree with some (or all) of these points so feel free to pushback against any suggestions you believe will harm your work. Just describe your counterpoints in the letter.**

**E.1: I understand the value of analyzing existing data and I believe the paper is straightforward about this issue. I think it would be valuable to detail existing papers from this dataset in a bit more detail and to be clear about all of the measures in the dataset (see Reviewer #1, esp. points #4 and #7).**

We agree that we should have been more transparent in the manuscript about already published papers based on this dataset. Three papers have been submitted on this dataset so far, but only two are published. One is exclusively about measurement arcana (Arslan et al., 2020) and one studies fertile window effects on attractiveness (Schleifenbaum et al., 2021). The unpublished manuscript focuses on fertile window effects on inbreeding avoidance (Holzleitner et al., major revision). Both papers focus on fertile window effects on a within-subject level, but make no attempt to causally identify the effect of hormonal contraception. These studies use hormonal contraceptive users as a quasi-control group who do not experience ovulation. However, these studies do not attempt to causally identify the effect of being on hormonal contraceptives and analyze different outcomes.

We have included a reference to the already published manuscripts on page 15 in the methods section:

*Arslan, Reitz, et al. (2020) published a paper based on the data collection focusing on measurement arcana investigating the benefits of a planned missingness design in diary studies and Schleifenbaum et al. (2021) studied fertile window effects on attractiveness on a within-subject level using hormonal contraceptive users as a quasi-control group who do not experience ovulation.*

The website at <https://rubenarslan.github.io/gocd2/> that is referenced in the methods section documents the data and measures. It also links to several blog posts published with the data. We adjusted the reference on page 15 in the methods section slightly to communicate more clearly that this website documents all measures included in the Goettingen Ovulatory Cycle Diaries 2:

*A codebook for the full dataset including all measures is available at* [*https://rubenarslan.github.io/gocd2*](https://rubenarslan.github.io/gocd2) *(Arslan, Driebe, et al., 2020).*

We have responded to the excellent point R1.7 by reviewer #1 below.

Arslan, R. C., Reitz, A. C., Driebe J. C., Gerlach T. M., & Penke L. (2020). Routinely randomize potential sources of measurement reactivity to estimate and adjust for biases in subjective reports. *Psychological Methods*, *26*(2), 175‒185. <https://doi.org/10.1037/met0000294>

Holzleitner, I. J., Driebe, J. C., Arslan, R. C., Hahn, A. C., Lee, A. J., O’Shea, K. J., Gerlach, T. M., Penke, L., Jones, B. C., & DeBruine, L. M. (being revised). No evidence that inbreeding avoidance is up-regulated during the ovulatory phase of the menstrual cycle. *bioRxiv*. <https://doi.org/10.1101/192054v4>

Schleifenbaum, L., Driebe, J. C., Gerlach, T. M., Penke, L., & Arslan, R. C. (2021). Women feel more attractive before ovulation: evidence from a large-scale online diary study. *Evolutionary Human Sciences,* 1‒34. <https://doi.org/10.1017/ehs.2021.44>

**E.2: Reviewer #2 had several technical questions that deserve your attention.**

We thank reviewer #2 very much for their elaborate and really valuable feedback and we addressed all their concerns below.

**E.3: Most of my reactions were minor comments on style and phrasing.**

**E.3a: Can you define selection effects, attrition effects, and reverse causality on page 5 and perhaps ground the brief explanations with reference to this substantive area. This does not need to be extensive but just outlined for readers right away to anticipate the material on page 10.**

Thank you very much for this comment. We agree and included a few additional sentences describing selection effects, attrition effects, and reverse causality with respect to effects of hormonal contraceptives in the introduction section on page 4.

*Selection effects describe the possibility that interindividual differences in women affect both the choice of contraceptive method and psychological outcomes of interest here. Attrition effects occur when women who experience negative effects of hormonal contraceptives discontinue using them, so that the remaining hormonal contraceptive users are more likely to experience no effects or positive effects (Vitzthum & Ringheim, 2005). Selection and attrition could both exaggerate or mask possible causal effects of hormonal contraception. The possibility that relationships between psychological outcomes and hormonal contraceptive use in correlational studies might exist because the outcome influences the contraceptive choice (e.g., higher frequency of vaginal intercourse might lead to the decision to start using hormonal contraceptives) is called reverse causality.*

**E.3b: I agree with Reviewer #1 about the need to provide a bit more support and justification for the hypotheses about hormonal contraceptives and sexual behavior given the outlet.**

We thank you for this suggestion and have responded to it in our comment to R1.1 below.

**E.3c: I would refer to personality traits rather than to the broader concept of personality on page 10.**

We agree and updated the introduction (page 10) as well as Figure 1 accordingly:

*Our first goal was to investigate the degree to which contraceptive use and congruent contraceptive use could be explained by selection effects of demographic variables (age, education, and income), personality traits (openness, conscientiousness, extraversion, agreeableness, neuroticism, religiosity), and relationship duration.*

**E.3d: I think some reviewers might need a bit more orientation to DAGs. Again, this does not need to extensive but just something a bit more detailed. I suspect readers here might be better versed with old school path diagrams so a few sentences about the similarities and differences would be helpful.**

Thank you for your opinion on this matter. We added a few sentences further describing DAGs and their differences compared to path diagrams in the introduction (page 10):

*Figure 1 shows a directed acyclic graph (e.g., Pearl, 1995) of the suggested causal network. Directed acyclic graphs visually represent causal assumptions. They offer an intuitive approach for thinking about causal structures and help to answer questions about potential third variables (i.e., confounders). Even though they look a lot like structural equation models, directed acyclic graphs differ in two important points: While structural equations models allow for bidirectional relationships between variables, directed acyclic graphs include only one-headed arrows and whereas structural equation models assume linear, additive relationships (unless indicated otherwise) the arrows included in directed acyclic graphs might reflect any form of relationship. For a primer on directed acyclic graphs see Rohrer (2018).*

**E.3e: I had a similar reaction to Point #2 from Reviewer #1 about the causal language. I think I understand the various arguments about causal language so I am not insisting on anything but I think using associative terms would be cleanest and mitigate any reactance in readers.**

We understand the caution concerning causal language and have responded to the points raised in R1.2 below.

**E.3f: I think a few passages defining weakly informed priors for non-Bayesians on page 20 would be helpful.**

We should have been much clearer here. We went with the defaults mainly for pragmatic reasons and our summary of the defaults was misleading (priors on the estimated parameters are improper flat priors, which are usually labelled "non-informative" though some consider that misleading too). We corrected this on page 22 in the methods section and added further information concerning the comment R2.3 by reviewer #2:

*Due to a lack of research especially on some of the outcomes (e.g., masturbation) and the fact that we switched between Bayesian and frequentist approaches for sensitivity analyses improper flat priors (which ensure consistency with maximum likelihood) for all parameters were used.*

**E.3g: Would including percentages for income and relationship duration in Table 2 make sense?**

We agree that this information would help the reader and included percentages for income and relationship duration in Table 2.

**E.3h: Please check the first sentence of the Conclusion. I think the restatement of the results is incorrect (or perhaps my caffeine levels waned).**

This mistake was probably due to our low caffeine levels and we are very thankful to you for pointing it out. We adjusted the conclusion on page 55 accordingly:

*We found evidence that the use of hormonal contraceptives positively predicts frequency of vaginal intercourse and negatively predicts frequency of masturbation.*

**E.4: I like to see formal statements of constraints on generalizability in papers at this outlet. Would you consider formally including such a section and citing the Simons et al. paper? This would not need to be extensive and could pull from some of the material already in the Limitations section. I also wonder if it would be useful to note that other “controls” not measured in this study could result in different conclusions.**

**Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. Perspectives on Psychological Science, 12(6), 1123-1128.**

Thank you very much for this suggestion. We added a passage about generalizability based on Simons et al. (2017) on page 53 in the discussion section.

***Constraints on Generality***

*Following the guidelines on constraints on generality (Simons et al., 2017) the following four factors reduce the broad generalizability of the current results: First, the sample studied consisted of heterosexual WEIRD (Henrich et al., 2010) women with a high proportion of undergraduate psychology students. Although our main explanation assumes a universal biological mechanism for the results, different absolute hormone levels in less prosperous and well-nourished populations (Vitzthum, 2009) might affect the frequency of ovulation and hence the observable effect size. Moreover, some of our alternative explanations (such as reverse causality) may be much more dependent on circumstances. Second, even though the gold standard for measuring sexual frequency and masturbation frequency by using diary reports was applied, all results are purely based on self-reports and generalizability to other measures might be limited. Third, from a temporal perspective the composition of hormonal contraceptives (e.g., the dosage of estrogen and progestin) has changed over the decades and will change further in the future. If the reported effects are due to certain estrogen or progestin dosages, results might not be replicable based on samples from different times.*

**E.5: The paper is long. That might be unavoidable but I would just ask you to consider anyways to make this work more streamlined for readers.**

We agree that the paper is very long and have tried to make it more streamlined for readers while implementing all reviews. Nevertheless, the paper is still about the same length because we included additional information answering the important issues raised by the reviewers and you.

**Reviewer 1 (R1)**

**General Comments:**

**This manuscript investigates three core questions: selection effects regarding choice of contraceptive method (i.e., no/non-hormonal vs hormonal), potential effects of contraceptive method (and contraceptive congruency) on relationship quality and sexual functioning, and sensitivity of these effects when unobserved confounders are considered.**

**The authors report that a complex model of selection effects did not substantially improve on the simple model when predicting hormonal contraceptive use or change in contraceptive method. They also find no evidence for an effect of hormonal contraceptive method, congruent contraceptive use, and their interaction, regarding partner attractiveness, relationship satisfaction, sexual satisfaction and libido. They do find evidence for positive effect of hormonal contraceptives on vaginal intercourse, and a negative effect of hormonal contraceptives on masturbation frequency. Potential observed and unobserved effects were considered.**

**I think the paper is well written overall, and the results (although they are not what the authors predicted) demonstrate commitment to good science. I also want to note that I really appreciated the authors providing access to the analysis and codebook. The contraceptive and ovulatory literature has a need for large scale studies that utilise conservative statistical techniques and I enjoyed being able to see the exact measures and confirm what was described in text. I think this manuscript could be published with some revisions.**

**Major Comments**

**Background and Literature Review**

**R1.1: The authors discuss relevant literature regarding hormonal contraceptives, the congruency hypothesis and cycle shifts (e.g., partner preferences, attractiveness ratings, etc). However, more detail could be provided regarding why the authors hypothesize that hormonal contraceptives may impact sexual functioning. I recommend this issue is addressed because the audience of this journal may not be familiar with hormonal contraceptives and theories relevant to the research aims.**

We agree with reviewer #1 that we should have provided more information about potential mechanisms behind the effects of hormonal contraceptives on sexuality. Nevertheless, our main goal was to point out that there is neither a clear picture about effects of hormonal contraceptives on all outcomes nor a theory that would explain the patterns observed. To communicate this more clearly and to provide a range of potential mechanisms discussed in the literature, we added a paragraph on page 7 in the introduction section:

*Beyond partner preferences, there is a growing body of research investigating effects of hormonal contraceptives on sexuality and satisfaction. Several mechanisms explaining how hormonal contraceptives might influence sexuality have been suggested. For example, the estrogen-induced increase in the production of sex hormone binding globulins based on hormonal contraceptives might lead to a decrease in libido by lowering the amount of free, biologically active testosterone (Pastor et al., 2013, Zimmermann et al., 2013). At the same time, other mechanisms like overcoming the fear of unwanted pregnancies, the resolution of gynecologic disorders (e.g., endometriosis, dysmenorrhea), and the reduction of body image concerns with an increase in self-esteem for women with clinical hyperandrogenism might be mechanisms behind potential positive effects of hormonal contraceptives on sexuality (Both et al., 2019). The recent review by Both et al. (2019) found that only a minority of women experienced changes in sexual functioning and concluded that the effects of hormonal contraceptives on sexual functioning - and sexual desire in particular - are understudied and poorly understood.*

**R1.2: The authors write that the terms “effects of hormonal contraceptives” and “effects of congruent contraceptive use” are used to remain constant with previous research (footnote 2, page 5). We recommend the use of language that is appropriate for correlational research, to help improve the accuracy and quality of the literature.**

We thank reviewer #2 for this comment and agree that one must be cautious when deciding which causal formulations to use when interpreting correlational research. Nevertheless, the aim of our study was to infer the causal effect based on correlational data as closely as possible. Instead of hiding assumptions behind correlative language and risking implicitly made inferences we wanted to communicate causal assumptions and causal effects openly (Grosz et al., 2020). We don't think we did a very good job explaining our point in footnote 2 on page 5 in the theory section. We updated the footnote accordingly:

*Although our study had a correlative nature and was therefore not able to directly determine causal effects, we controlled for potential selection variables in order to estimate potential causal effects. To make the proposed causal structure as transparent and comprehensible as possible, a directed acyclic graph was drawn, incorporating all included selection variables, predictors, and outcomes (see Figure 1). In addition, the sensitivity analyses quantitatively formulated the assumptions about unobserved confounders that must be taken into account when considering potential causal effects of hormonal contraceptives and congruent contraceptive use. As the aim of the study was to infer causal effects based on correlational data as closely as possible and to avoid hiding this causal aim behind correlational language (Grosz et al., 2020), the terms* effects of hormonal contraceptives *and* effects of congruent contraceptive use *are used throughout this manuscript.*

In addition, we reviewed the manuscript to ensure that whenever we use causal language, it is appropriately couched in uncertainty and that the assumptions made are clear in the context. We kept the phrase "effect size estimate" because it's a commonly understood statistical phrase and is in keeping with our research goal. For theory and hypotheses as well as the discussion about potential alternative explanations, we remained with the proposed causal language to communicate the aim of the paper (as well as its limitations) as clearly and transparently as possible.

**Method**

**R1.3: The authors provide sufficient evidence regarding participant demographics overall. However, more information regarding hormonal contraceptive (HC) type could be included in text. I recommend that the authors include (from the supplementary documentation) information about the type of HC (e.g., pill, patch, implant) and the proportion of each type in the current dataset. If available, information regarding estrogen dose (particularly regarding pill users) is also valuable.**

Thank you very much for this comment. We agree that more information about contraceptive methods would be very helpful for the reader. Therefore, we included sample size and percentage for each broad contraceptive method in the methods section of the main document (page 20) based on the supplemental material. As part of this process we updated Figure S2 in the supplemental material to correspond to the text in the main manuscript. In addition, we distinguished between the normal pill and the mini-pill and included means, standard deviations, and ranges for estrogen dosage in oral hormonal contraceptives.

*Participants were sorted into two large groups according to current contraception (non-HC users: n = 688, 58%; HC users: n = 491, 42%; see Figure S2 in the Supplemental Material for allocation process and number of women in each group). No/nonhormonal contraception included nonhormonal intrauterine devices (n = 85, 7%), fertility awareness methods
(n = 120, 10%), condoms (n = 380, 32%), and no contraception at all (n = 89, 8%), as well as miscellaneous other barrier-based methods (n = 14, 1%). Hormonal contraception included only the pill (n = 251, 21%), only other hormonal contraceptives (n = 61, 5%), and any combination of the pill and nonhormonal contraceptives (n = 161, 14%), other hormonal contraceptives and nonhormonal contraceptives (n = 17, 1%) as well as one combination of the pill, other hormonal contraceptives and nonhormonal contraceptives (n = 1, 0%). From all women using oral contraceptives (n = 413, 35%) only 25 women (2%) used progestin-only oral contraceptives (i.e., “minipill”). Information about estrogen dosage was available for 377 women using regular oral hormonal contraceptives (m = 27.54µg; sd = 7.24µg; min = 2µg;
max = 100µg) and 352 of all women using oral hormonal contraceptives reported the type of gestagen (Dienogest:* n *= 109, 31%; Levonorgestrel:* n *= 105, 30%; Chlormadinonacetat:* n *= 62, 18%; Desogestrel:* n *= 38, 11%; Drospirenon:* n *= 17, 5%; Cyproteronacetat:* n *= 14, 4%; Nomegestrolacetat:* n *= 5, 1%; Norgestimat:* n *= 2, 1%).*

**R1.4: The data within this manuscript was derived from an existing dataset. Can the authors confirm whether the measures analysed in this article been published elsewhere?**

We agree that we did not communicate transparently enough whether other articles based on the *Goettingen Ovulatory Cycle Diaries 2* had been published before. We responded to this concern above to the points raised in comment E1.

**Analyses**

**R1.5: Some measures (i.e., frequency of vaginal intercourse and masturbation) were collected daily but were aggregated to produce a single outcome value. Can you explain why the decision was made to aggregate the data, rather than using the daily measures?**

As noted on page 53 in the discussion section we used diary reports because they are more reliable than retrospective behavior measurements (McAuliffe et al., 2007) and described as the gold standard for measuring sexual frequency by Graham et al. (2003). We did not examine any predictors at the within-subject level, so aggregating the outcomes reduced complexity. To communicate this more clearly we added a sentence on page 21 of the methods section.

*Diary outcomes were all aggregated to reduce complexity and because we were not interested in any predictors at the within-woman level.*

**R1.6: Fertility is known to be associated with extra- and in-pair desire in naturally cycling women (Arslan et al, 2018; cited in the manuscript). I wonder if fertility should be considered in the sensitivity analyses.**

We consider suppressed ovulation (fertility) to be one potential mechanism for differences between hormonal contraceptive and non-hormonal contraceptive users in the introduction section on page 7, but we decided that investigation of fertility as a mechanism is beyond the scope of the present manuscript which, as the editor rightly noted in comment E5 above, is already very long.

**Results**

**R1.7: In the section regarding ‘Further Unobserved Confounders’ (page 49), the authors write that some variables (i.e., sociosexual orientation (SOI), satisfaction with body image and life satisfaction) were not included due to the potential for colliders or mediators. The original codebook includes SOI, however it was unclear whether satisfaction with body image and life satisfaction were recorded in the original diary study. Can the authors clarify whether these variables were available in the dataset, or if their exclusion was justified retrospectively?**

We thank reviewer #1 for this comment and agree that this section wasn't clear about whether potential further unobserved confounders would have been available in the current dataset. As some of them were and others weren't we added a sentence on page 51 in the discussion section to communicate this more transparently.

*While some of these potential unobserved confounders were not measured in the available dataset (in particular orgasm frequency, importance of sex, anxiety, depression, and contentment with sexual frequency), others would have been available (in particular sociosexuality, general life satisfaction, and satisfaction with own body image) but we decided not to include them in the current study to prevent controlling for potential colliders or mediators (Rohrer, 2018).*

**R1.8: Footnote 3, page 9, justifies the inclusion of all HC users as one group, while highlighting that previous research has often included only oral contraceptive users. Follow up analyses, contrasting data from oral contraceptive users and naturally cycling women, would be a valuable addition to the analyses.**

We agree with the reviewer that these robustness analyses would be very informative and helpful. Therefore, we have added the results based on these analyses to the website. The patterns between oral contraceptive users and naturally cycling women did not differ substantially from the patterns reported in our main analysis.

We refer to these results in footnote 3 on page 9 in the introduction section

*Results from robustness analyses including only pill users compared to naturally cycling women are available on the supplementary website* [*https://laurabotzet.github.io/effects\_of\_contraception/14\_analyses\_robust*](https://laurabotzet.github.io/effects_of_contraception/14_analyses_robust)*. The results did not differ substantially from the results based on our main analyses including all hormonal contraceptive users.*

In addition, we mention the robustness analyses in the methods section on page 24:

*As robustness analyses, we performed all analyses of effects of hormonal contraceptives including only pill users compared to naturally cycling women. The results are reported on the supportive website*[*https://laurabotzet.github.io/effects\_of\_contraception/14\_analyses\_robust*](https://laurabotzet.github.io/effects_of_contraception/14_analyses_robust)*. The patterns between oral contraceptive users and naturally cycling women did not differ substantially from the patterns reported in our main analysis including all forms of hormonal contraceptives.*

**Minor Comments and Suggestions**

**R1.9: On page 6, the authors write “Hormonal contraceptives contain synthetic progestin, which suppresses the natural production of estrogen and progesterone.”. It would be more accurate to write that hormonal contraceptives contain synthetic progesterone (i.e., progestin), as all progestins are synthetic progestogens.**

We agree with reviewer #1 and changed the sentence on page 5 in the introduction as suggested.

*Hormonal contraceptives contain synthetic progesterone (i.e., progestin), which suppresses the natural production of estrogen and progesterone.*

**R1.10: In addition to Wiegratz et al (2003), page 6, a useful reference for testosterone and oral contraceptive use: Zimmerman, Y., Eijkemans, M. J. C., Coelingh Bennink, H. J. T., Blankenstein, M. A., & Fauser, B. C. J. M. (2014). The effect of combined oral contraception on testosterone levels in healthy women: A systematic review and meta-analysis. Human Reproduction Update, 20(1), 76–105.**

Thank you very much for this really helpful literature suggestion. We have included the reference on page 5 of the introduction.

*A systematic review and meta-analysis by Zimmermann et al. (2013) concluded that oral hormonal contraceptives reduce levels of total as well as free testosterone and increase levels of sex hormone-binding globulin.*

**R1.11: For clarity, on page 20, when referring to the outcome measures, clarity would be enhanced if the authors explicitly state which measures were from the baseline (as they have for the diary measures).**

We agree and have added information about outcome measures based on the initial survey on page 21 of the methods section.

*Perceived partner attractiveness (two items; one measuring facial attractiveness and one on body attractiveness), relationship satisfaction (five items measuring aspects including satisfaction, fulfilling of needs, and—reverse-scored—conflicts), and sexual satisfaction (one item) were reported by partnered participants in the initial survey (1 =* does not apply at all*, 5 =* fully applies*).*

**Reviewer 2 (R2)**

**Summary**

**This paper aims to assess the causal effect of hormonal contraceptives on several behavioral and psychological outcomes, such as perceived partner attractiveness, relationship satisfaction, frequency of vaginal intercourse and masturbation. The paper uses a survey data of 1,179 women, regression analyses to adjust for observed confounders, and sensitivity analyses to assess robustness to unobserved confounders. For most outcomes, it concludes that it does not find much evidence for an effect. The two exceptions are the effect of contraceptives on sexual and masturbation frequency, which are robust to the inclusion of observed confounders, and appear to be somewhat robust to potential unobserved confounders. However, reverse causation and attrition biases cannot be ruled out.**

**Comments**

**This paper is well written and very transparent in its methodological choices. Causal inference with observational data is difficult, and most social sciences, in particular psychology, have been doing it casually, either by naively avoiding causal language (but interpreting the results causally) or simply by making causal claims ignoring the role of causal assumptions altogether.**

**This paper is refreshing in its transparency. It clearly states its causal goals, it explains the rationale for controlling for certain covariates and avoiding others (eg., potential colliders and mediators), and it is not afraid to point out limitations of what can be inferred from the data. The supporting information provided is also very through and complete.**

**There are some technical issues that still need to be addressed as I detail below, but overall I am positive about the paper.**

**R2.1: Be more specific about the estimand**

**Although the paper is clear regarding its causal goal, it still does not explicitly mention which causal effect it is trying to estimate. For instance, is it the average treatment effect (ATE) ? This is not a particular limitation of this paper, but a general limitation of most papers in the social sciences. Only recently this has been gradually changing due to advances in causal inference research and education. I suggest the authors be even more specific about their target causal quantity. As a reading suggestion, the following may be useful:**

**Lundberg I, Johnson R, Stewart BM. What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory. American Sociological Review. 2021;86(3):532-565. doi:10.1177/00031224211004187**

We thank reviewer #2 for this comment and agree that we were not explicit enough about our theoretical and empirical estimands. We are indeed interested in the average treatment effect and have included an additional paragraph in the methods section on page 24.

*For effects of hormonal contraceptives our theoretical estimand of interest (Lundberg et al. 2021) was the average treatment effect of hormonal contraception on sexual satisfaction, frequency, and so on. We strove to identify this causal effect by adjusting for confounding variables. The so-adjusted estimated effect size was our empirical estimand. For linear regression models, the effect size of interest was the estimated coefficient. For Poisson regression models, we used the difference in percentage frequencies across the diary. Therefore, we based our ROPE for these outcomes on percentage frequencies and reported differences in percentage frequencies in addition to estimated effect sizes.*

In addition, we added more information about the average treatment effect of hormonal contraception for Poisson regression models in the results section on page 32.

*Based on the uncontrolled models this corresponded to a difference of 3.5 [90% HDI: 3.0; 4.1] percentage points in estimated probabilities of penetrative intercourse per day (HC users: 15.9%; non-HC users: 12.4%) and a difference of 5.0 [4.5; 5.5] percentage points in estimated probabilities of masturbation per day (HC users: 10.3%; non-HC users: 15.3%). For the controlled models we computed average marginal effects (documented at* [*https://laurabotzet.github.io/effects\_of\_contraception/18\_marginal\_effects.html*](https://laurabotzet.github.io/effects_of_contraception/18_marginal_effects.html)*) by setting number of diary days filled out as one, using all levels of categorical variables (income, relationship duration), using five categorical levels for the continuous variables age and years of education and assuming empirical mean values for all other continuous variables (Big Five personality traits, religiosity). Differences in average marginal effects were 2.8 [2.0; 3.6] percentage points in estimated probabilities of penetrative intercourse per day (HC users: 18.0%; non-HC users: 15.2%) and 3.1 [2.7; 3.6] percentage points in estimated probabilities of masturbation per day (HC users: 9.0%; non-HC users: 12.2%). Therefore, the observed differences were all larger than the difference in 1 percentage point that our predefined ROPE for analyses assuming Poisson distributions was based on.*

**R2.2: Choice of effect size, priors and ROPE**

**I personally like the choice of the authors of not focusing on statistical significance, and instead focusing on estimating effect sizes and regions of practical equivalence. I may have missed, however, which effect size is being reported: for instance, is it a regular or standardized regression coefficient? This is particularly relevant for a Poisson or logistic model, as I’m not sure the default “effect size” would directly reflect the ATE, for instance (in case that’s the target of inference).**

**Relatedly, the authors simply mention they used the default priors of brms, but it would be more transparent to explicilty mention what these priors are, and perhaps some discussion. Finally, some discussion regarding the choice of the region of practical equivalence would be useful, instead of simply following Kruschke.**

We reported unstandardized regression coefficients throughout the whole manuscript. To communicate that more transparently to the reader we added *unstandardized* whenever we reported effect size estimates (page 29, 32, 35, 39, 40; Figure 3, 4, 5; Table 3, 4, 5, 6). Concerning the estimand we agree with reviewer #2 and have answered their concern R2.1 above.

As noted in our answer to comment E3.f we should have been much clearer about the priors used and our summary of the defaults was misleading. We corrected this on page 22 in the methods section and added a short reason for our decision to use these priors instead of more informative priors:

*Due to a lack of research especially on some of the outcomes (e.g., masturbation) and the fact that we switched between Bayesian and frequentist approaches for sensitivity analyses improper flat priors (which ensure consistency with maximum likelihood) for all parameters were used.*

Our ROPEs similarly reflect a lack of strong prior research that could inform our notion of practical equivalence, especially given widespread heterogeneity in outcome measures. An advantage of the ROPE+CI decision rule is that it is trivially easy to apply different ROPEs as more information about practical equivalence bounds becomes available. To communicate this to the reader we added some discussion on page 22 of the methods section:

*The decision to follow Kruschke’s (2018) suggestion reflects a lack of strong prior research that could inform our notion of practical equivalence, especially given widespread heterogeneity in outcome measures.*

**R2.3: Sensitivity analysis and comparison to observables**

**Sensitivity analyses in observational studies are extremely important, since we rarely measure all relevant confounders. The authors are to be applauded for their effort in this direction, by directly quantifying the robustness of their estimates to the threat of unobserved confounding.**

**I have two technical comments on the sensitivity of the effects on the two count outcomes, frequency of vaginal intercourse and masturbation: (1) The sensitivity by Cinelli and Hazlett (2020) targets OLS regression coefficients. Since your treatment is binary, regular OLS still works for estimatings a conditional expectation, even if the outcome is a count. Thus, you should use OLS estimates for the sensitivity analysis, instead of the Poisson GLM estimates (it is unlikely this will change the main results, but it still should be corrected). (2) The comparison of relative strength of unobservables vs observables cannot be done by comparing changes in the robustness values directly. Instead, you should use the formal bounds on the strength of confounding as in Cinelli and Hazlett (2020), section 4.4.**

**Regarding the use of OLS for estimating conditional expecations with binary treatments with non-continuous outcomes, see Mostly Harmless Econometrics (Chapter 3), or, for instance, the recent discussion in Psychology provided by Gomila (2020). There the focus is logistic regression, but similar reasoning can be applied to Poisson regression:**

**Gomila R. Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. Journal of Experimental Psychology: General. 2020 Sep 24.**

We are really thankful for the thoughtful and elaborated review from reviewer #2. Thank you very much especially for the very helpful comments on the performed sensitivity analyses based on Cinelli and Hazlett (2020) approach. We are grateful to receive feedback from - what seems to be to us - an expert on this topic. Therefore, we agree with all proposed changes and have implemented them in the main manuscript.

First, for the sensitivity analyses, we substituted the Poisson models with Gaussian regression on the daily frequency of vaginal intercourse/masturbation. To make this transparent to the reader we updated a paragraph on page 25 in the methods section.

*The R-package sensemakr v0.1.3 (Cinelli et al., 2020) was used for analysis. As this package does not implement sensitivity analyses for Bayesian analyses and only targets linear regressions, we performed frequentist linear regressions for all outcomes. For our scenario, linear regression is robust to this violation of assumptions and the effects closely approximated the conditional effects in Bayesian Poisson regressions.*

In addition, we included the information that linear regressions were used for all outcomes on page 39 in the results section.

*Results are based on frequentist linear models for all outcomes.*

Performing sensitivity analyses based on linear regressions indeed changed the values for the sensitivity analyses substantially, perhaps because Poisson models do not estimate the residual variance. We updated Table 5 and Table 6 with the new values based on sensitivity analyses with linear regression models and adjusted all values on page 39/40 in the results section.

Since unobserved confounders would have to explain *RVq = 1* = 9.4% / *RVq = 1* = 11.1% (compared to *RVq = 1* = 19.1% / *RVq = 1* = 29.4% based on generalized linear models based on Poisson distributions) of the residual variance of both the predictor and the outcome to bring the point estimate for hormonal contraceptives on frequency of vaginal intercourse / frequency of masturbation to zero, we adjusted the paragraph in the discussion section on page 44.

*These links were robust to the inclusion of observed confounders and sensitivity analyses suggested that unobserved confounders would need to strongly influence outcomes (about as strong as all observed confounders taken together) in order to substantially alter conclusions.*

Second, we removed all comparisons of relative strengths between observed and unobserved selection effects based on changes in robustness values. We are very thankful for reviewer #2 for pointing out section 4.4 in Chinelli & Hazelett (2020). Instead of using robustness values we focused on comparing the total R2 of all covariates with the total R2 of potential unobserved confounders that would be needed to fully account for the effect of hormonal contraceptives on the outcomes. We therefore adjusted the result section on page 40:

*To bring the point estimate for hormonal contraceptives on frequency of vaginal intercourse to zero, unobserved confounders would have to explain RVq = 1 = 9.4% of the residual variance of both the predictor and the outcome. This means that compared to observed confounders which explained ΔR2Y~D|X = 0.7% of the residual variance of the outcome unobserved confounders would need to explain R2Y~D|X = 1% of the residual variance of frequency of vaginal intercourse to fully account for the effect of hormonal contraceptives. [...]*

*To bring the point estimate for hormonal contraceptives on frequency of masturbation to zero, unobserved confounders would have to explain RVq = 1 = 11.1% of the residual variance of both the predictor and the outcome. This implies that compared to observed confounders which explained ΔR2Y~D|X = 1.5% of the residual variance of the outcome unobserved confounders would need to explain R2Y~D|X = 1.4% of the residual variance of frequency of masturbation to fully account for the effect of hormonal contraceptives.*

In addition, we updated the discussion section page 49.

*We estimated the robustness of the effects of hormonal contraceptives and congruent contraceptive use in light of potential unobserved confounders. Sensitivity analysis suggested that the influence of unobserved confounders would need to be nearly 1.5 times as strong as the influence of observed confounders to fully account for the effect of hormonal contraceptives on frequency of vaginal intercourse, and nearly as strong as the influence of observed confounders to fully account for the effect of hormonal contraceptives on frequency of masturbation. Even when taking into account the broad range of included observed confounders (demography, personality, and romantic relationship information) it seems plausible that unobserved confounders might exist that would fully explain the reported effects of hormonal contraceptives on frequency of vaginal intercourse and frequency of masturbation. Besides potential unobserved selection effects, we now consider three additional possible challenges: reverse causality, attrition effects, and further unobserved confounders.*

This summarizes the ways in which we have responded to your and the reviewers’ constructive and helpful comments. We believe we have addressed all issues that were raised in a satisfactory manner and hope that you will find this revised manuscript suitable for publication in *Collabra Psychology*.

We appreciate your continued consideration of our manuscript. All authors have approved the submission of the revised manuscript.

Sincerely,
Laura J. Botzet, Tanja M. Gerlach, Julie C. Driebe, Lars Penke, and Ruben C. Arslan

**Editor Final Decision: Accept**

Oct 4

Dear Laura J Botzet,

I have now had a chance to read your revised manuscript “Hormonal Contraception and Sexuality: Causal Effects, Unobserved Selection, or Reverse Causality?”, along with the letter describing the changes you made in response to the prior decision. Thank you for your responsiveness to the concerns raised in the prior reviews. Based on this outcome, your paper is now officially accepted for publication in Collabra: Psychology. Congratulations on this excellent work! I think it will make an important contribution to the literature and I look forward to seeing it published. I also hope your experiences with Collabra: Psychology have been positive and that you will continue to consider it as an outlet for your work.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point. Our managing editor will contact you in case there are any pre-prodution file related questions. You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Sincerely,
Brent Donnellan