30 June 2023

Dear Dr Rodgers,

Following email correspondence since our previous submission (submitted 19 June), we attach a final version and track changes of our manuscript “Meta-research: The effect of the COVID-19 pandemic on the gender gap in research productivity within academia” for consideration for publication in *eLife.*

Again, we would like to thank the editors and the reviewers for their constructive and positive comments which have much improved our manuscript. Please find below a detailed description of our responses and alterations in light of these comments, **in bold,** underneath Author action/response, with text from the revised manuscript written in ***bold, blue italics***.

Yours sincerely,

Kiran Lee (on behalf of all co-authors)

Comment:

1. I actually think the differences in how individual studies estimate a gender gap matters a lot. It might be that, as the authors argue, there is no set standard for conducting this type of study, but this does not mean we should treat all study designs as equally telling about the effect of the COVID-19 pandemic. Studies comparing publication or submission rates before and after the pandemic onset should, all else equal, give us more precise estimates of gender gaps than retrospective survey responses, etc. I think the authors still underappreciate this in their manuscript. It is not only a question of "Such a change might reflect lower submission and acceptance rates of articles by women compared to their male colleagues or an increased drop-out of women from academia caused by the pandemic." , but also problems with human recall, self-serving bias, over- or underreporting in surveys, etc.  
   I agree that the gender gap estimated entails a multitude of different causal mechanisms, but each study still differs in which rung on the causal ladder it is situated. The amount of evidence each study provides differs, and I think the authors should acknowledge this more up front and clearly.

**Author response:**

***We now include in the Limitations section L657-663, “For survey-studies, only 18 effect sizes were used. These had a large heterogeneity in effect sizes, possibly reflecting subtle differences in the measure of research productivity asked in the survey. Surveys sample limited numbers of respondents, potentially biased towards sampling those holding extreme opinions of the pandemic. Subjectivity in survey responses could skew the estimate because of recall limitations and self-serving bias.” We also summarise this in the Discussion section L390-392, “Survey-studies may unintentionally sample respondents with extreme opinions and can suffer from recall limitations and self-serving bias.”***

Comment:  
2) I appriciate that the authors now speak of gender gaps and not bias, but I think this would have been a good point for the discussion section. Because the meta-analysis gives us an idea of the average effect of the pandemic, I think the authors should reflect on what type of mechanisms could be at work here. Could this inflate gender bias or is it a more indirect effect on womens' productivity (e.g. increased teaching load, more work at home such as homeschooling, etc.). It would give a much needed context for interpreting the effects.

**Author response:**

***We now include in the Discussion section L377-383, “These mechanisms are likely to be a combination of gender inequalities that affect all women during the pandemic such as changes in carer roles and financial stability (Collins et al., 2021; Fisher and Ryan, 2021; Flor et al., 2022), and those specifically affecting women working in academia such as changes in the potential to start new projects (Malisch et al., 2020; Herman et al., 2021; Pereira, 2021) and changes in research topics and publication processes to cover new topics (Viglione, 2020; Clark, 2023).”***

Other changes made:

1. We make our Abstract more concise to accommodate a sentence summary, L24-30, ***“Measuring research productivity to gauge academic performance is flawed because it is biased against certain groups of researchers. Women researchers are disadvantaged because gender roles and unconscious biases operating both at home and in academia can affect research productivity. These biases may have strengthened during lockdown conditions of the COVID-19 pandemic. Correspondingly, we bring together existing studies to show the gender gap has increased during the pandemic, with the relative research productivity of women declining compared to the pre-pandemic period.”***

1. We add in the “Has the pandemic increased the gender gap in research productivity? subsection of the Results section L131-138, ***“Across the full dataset (N = 130), after controlling for multiple effect sizes from the same study, we found the relative productivity of women to men decreased during the pandemic by -0.071 compared to before the pandemic ( 95% CI = -0.099 to -0.043, SE = 0.0144, p = <0.001, Fig. 1). This indicates that the relative productivity of women compared to men is 7% lower than what it was prior to the pandemic, meaning that in cases where men and women were estimated to be equally productive, the productivity of women now is only 93% that of men, or in cases where men were estimated to be twice as productive as women, such that 33% of the measured productivity outcomes came from women, the contribution of women declined to 30%.*** This is to help readers interpret our effect size.
2. In Figure 3, investigating the effect of research field, our samples (and also analyses) are smaller because we excluded survey-studies, considering reviewer comments suggesting it is difficult to compare between survey-studies. This was changed in our second submission already, but we report it here, for transparency.
3. We change in the “Does the gender gap differ across authorship roles?” subsection of the Results section, L249-257, ***“Based on these data, we found evidence of a significant differential impact of authorship position on effect sizes (QM(df = 5) = 13.190, p = 0.022, Fig. 5). The pandemic had a significant effect on first authorship roles (-0.052, 95% CI = -0.073 to -0.007, SE = 0.017, p = 0.019) but not for any authorship roles (-0.045, 95% CI = -0.107 to -0.017, SE = 0.320, p = 0.154), corresponding authorship roles (-0.058, 95% CI = -0.123 to 0.007, SE = 0.033, p = 0.080), middle authorship roles (-0.045, 95% CI = -0.173 to 0.820, SE = 0.065, p = 0.485) or last authorship roles (-0.040, 95% CI = -0.094 to 0.015, SE = 0.028, p = 0.152).”*** This is a correction to new results made after addition of new articles upon review, which were correctly added, but for which we missed adding the correct interpretation. We subsequently change in the Discussion section L369-374, ***“We found that the pandemic increased the gender gap particularly among first authors, potentially suggesting that women were particularly more restricted in the time they had available to write papers. However, we cannot exclude that other authorship positions underwent a similar increase in the gender gap because the samples were uneven, with half of effect sizes focussed on first authorship roles. Additionally, not all fields have the same authorship order norms making comparisons difficult.”***
4. We clarify the statistic in the “Is there evidence of publication bias?” subsection of the Results sections, L282-285, ***“This model correlates standard error with effect size and a negative slope suggests small studies do not have large effect sizes, with the negative slope among survey studies indicating the large heterogeneity that exists among these kinds of studies (see Discussion).”***