**Editor comments:**

**While there is nothing inherently flawed in the study design, there are a number of addressable comments and suggestions for improvement. A major point of emphasis was not the data itself, but the representation of it. Further, revisions for brevity and clarity would be most helpful to not mask any of the important points made here.**

**Reviewer #1:**

**The study by Hagan et al. uses a retrospective approach to assess gender disparities in the context of publications in the broad field of microbiology. Using the ASM family of journals, the authors analyzed and compared data for male and female editors, reviewers, and authors (first, corresponding, and middle) of papers published in ASM journals. Overall, this study demonstrates a persistent disadvantage for women microbiologists throughout the publication process and includes suggestions for how to work towards gender equality. The authors are commended for taking on this large body of data and making several novel and important findings that should have a direct impact on publishing at ASM and other society journals. Specific recommendations are below.**

**Major:**

1. **Statistical analyses. Neither quantitative nor qualitative statistics are used to explain if any of the gender differences are statistically significant. We can debate the philosophical usefulness of statistics, but it would help the reader to know, for example, whether reviewers really are significantly more likely to turn down a request from a female than a male editor (e.g., Fig 3C). Furthermore, use of proportional analyses for the data presented in Figures 2 and 4 would assist the reader with interpreting the reported gender-biased differences.**

Statistical analyses (such as the proportional analysis suggested) are not appropriate because we are not reporting the results of randomized samples to generalize to the ASM journal population. Instead, we are reporting the results of the population and are not attempting to generalize them to other journals or fields.

**In Figure 4D, as a result of not using statistics, it is confusing that the data show that there is no difference in the proportion of papers submitted and published if the corresponding author is a female, but the corresponding text in the Results section (lines 166-176) does not state this. In fact, the corresponding text focuses more on people not classified which isn't even shown in the Figure. The lack of statistics would help with interpretation throughout the manuscript.**

We thank the reviewer for this comment. We added a statement indicating that the difference between women’s submitted and published were near equal and clarified the portion of the analyses that referred to published papers.

**2. Graphics. The graphs in Figures 5, 6, 7, S6 and S7 were very difficult to interpret, with figure captions that did not further assist the reader. Some examples:**

We thank the reviewer for noticing this gap in our communication and bringing it to our attention. Responses are below each subpoint.

**1) in the % points difference heat map is blue indicative of a 'male bias' and red/orange indicative of a 'female bias;**

Blue indicates a bias against men (e.g., men receive that decision/outcome more often), while orange indicates a bias against women (e.g., women receive that decision/outcome more often). We have expanded figure captions to explain the calculation and its interpretation.

**2) how were the differences calculated and why are the x-axes ranges different for the different panels;**

We have added a description of how the difference in percentage points were calculated to the results. Regarding the x-axis scales, There was so much variation between plots that we used the range that improved the interpretation of each graph. We did, however, standardize the gradients (head maps) across plots to improve comparability.

**3) are the differences really ratios of males to females thus explaining the negative and positive numbers on the x-axes?**

No, the differences are the male proportion minus the female proportion, so negative values indicate that the female proportion was greater than the male proportion. We have clarified this in the text and figure captions.

**Also, Figures 7, S6, and S7 have no heat map.**

We have added heat maps to the figures that are missing them.

**The data presented in Figure S1 seems very important and should be a main figure with the findings emphasized as is shows meaningful trends that should be fixed in different ways at different ASM journals.**

We have moved Figure S1 to Figure 2.

**The data presented in Figure S4 is confusing and a bit weak. It was not a strong way to end such an important manuscript. My recommendation would be to omit as the analysis and groupings seemed subjective and confusing.**

We have greatly expanded the caption for Figure S4 to explain the methodology and indicate key observations.

**In Figure S5, there is no indication of gender-maybe the slight blue and orange areas are indicating gender; because there is no difference, does this really require a figure?**

We have added a gender legend to this figure and quantified the results in the text.

1. **Unknown gender. A large portion of the population under study, especially from outside of the US, was of "unknown gender". In the Discussion section, the authors acknowledge that a large portion of this group may be from Asian countries as their algorithm was not as good at categorizing Asian names. Could the authors consider finding a way, even working with a colleague familiar with Asian names, to try and classify more of these names. This is important not only for ethnic and racial sensitivity but because some findings of gender biases were greatest at non-US institutions and categorizing these unknown gendered individuals could improve interpretation from these countries.**

We agree with the reviewer that there are likely important data lost within the “unknown” category. However, a large reason why this category exists is because many languages do not have strictly gendered naming conventions. Additionally, the algorithms that have improved prediction of Asian names require logographic character input1. Our data was collected using the Latin alphabet, which loses the subtleties of a logogram that might help indicate gender2. Of the available gender inference algorithms, genderize.io (the one used here) has significantly higher confidence values for inferring gender from Asian names3.

1. <https://arxiv.org/pdf/1906.05769.pdf>

2. <https://www.aclweb.org/anthology/P15-2062.pdf>

3. <https://peerj.com/articles/cs-156/>

**4. Discussion. The Discussion section is excessively long (7 pages) and as a result, the primary points are lost.**

**There should not be introduction of new data and graphs in the Discussion section (e.g., Fig S8). In fact, the data from Fig S8 could be eliminated from the manuscript without loss as this could be briefly summarized in a sentence in the Results section.**

We feel that these data are important for understanding the full validation process of genderize.io and thus understanding its limitations We have moved Figure S8 to Figure S1 and now introduce it early in the results section during our description of the genderize.io validation.

**Some of the broader discussion about STEM could be eliminated. It is suggested that the authors condense the Discussion section to around 4 pages and stick to the primary take home messages, tangible recommendations, and acknowledgement of limitations.**

* Reassessed and condensed background areas in the discussion

**Minor:**

**1. This manuscript needs to be grammatically revised and proofread for clarity. In many places it**

We have revised and proofread the manuscript as suggested.

**2. Citations should be included for the following sentences:** -Citations have been added as suggested below.

**o Sentence beginning on line 52**

**o Sentence beginning on line 111**

**o Sentence ending on line 325**

**o Sentence ending on line 400**

**3. Please provide justification for eliminating commentaries and reviews from the study.**

Commentaries and review do not introduce new data (e.g., unpublished experiments or figures) and are often commissioned. As a result, these publications do not reflect the either assessment of original science by gatekeepers or its completion for submission by the corresponding author. We have added our rationale to the methods.

**Reviewer #2:**

**The manuscript entitled "Women are underrepresented and receive differential outcomes at ASM journals: A six-year retrospective analysis" describes gender disparity in manuscript submission and biases of reviewers and editorial members for manuscripts submitted to ASM journals between 2012 and 2018. The authors show less representation of women both as editors and as corresponding authors, despite similar levels of female first authors on published manuscripts. The study further revealed biases against women in multiple areas of publishing including lower acceptance rate to review a manuscript when requested by a female editor and higher rate of editorial rejections and rejections after review across multiple journals. Finally the authors show that manuscripts submitted by female corresponding authors in the US who are at medical schools or research institutions disproportionately experienced higher editorial rejections, whereas male corresponding authors at these institutions experienced the highest rate of acceptance.**

**The study is a powerful analysis that supports the continued lack of gender equity that exists in STEM from the perspective of publishing in the field of microbiology. Overall, the manuscript is well written, but some of the data are not clearly presented. There are areas where data described are not depicted in the figures presented and clarifications or revision of the presented data would improve the quality of the manuscript and strengthen the conclusions.**

**1. The skewed gender representation both among gatekeepers (Figure 1) and senior/corresponding authors (Figure 4) seems consistent with the current overall representation of women in the field at the level of faculty, roughly 25-30%. The authors should clarify and, if possible, include data to show whether the percentages of editor and senior authors are comparable to overall representation in each of the institution types or whether there is further diminishment of representation beyond the general population in the field. This is important for the longitudinal analysis where it is not clear whether the slight change or lack thereof in the proportion of female editors and reviewers over time is in spite of increases in female representation in at faculty ranks or is this another metric that supports a lack of progress in female representation overall, not just for editorial members.**

Thank you for your comment. We have clarified in the caption and the text that Figure 4A describes the proportions of senior (last/corresponding) authors while Figure 4B includes all author types (junior and senior).

**2. The authors should define "proportion of editor workload" for Figure 3. In the text, the authors describe "a slightly greater" proportion of workloads for male editors compared to female editors, but the actual data shows a large gap, 75% vs. 25%. Is this calculated as a proportion of total manuscripts handled or is this describing workload per individual editor?**

We have clarified our definition of editor workload in the text.

**3. The authors describe the number of manuscripts reviewed by men, women, and unknown gender between lines 135-139, but Figure 3B doesn't accurately represent this result. Both the text and figure should be modified to match the analysis.**

Thank you for pointing this out. The x-axis of Figure 3B is on a log10 scale, making the results appear to be inconsistent. To clarify this, we have added tick marks to identify the upper value of each box plot and noted the scale in the axis title.

**4. Based on analysis described between lines 197-209, the authors conclude that "women were half as likely to move to senior author or reviewer roles, and 30% as likely to be an editor than men." This analysis is not sufficient to show progression of women from junior roles to other roles such as corresponding authors and editors.**

We thank the reviewer for their comment, however we disagree with this conclusion. Each individual that submits to the ASM journals is assigned a unique identifier. We used this identifier to track individuals through their roles at the ASM journals. By restricting the evaluation to individuals who were junior authors at the ASM, we could identify how many authored manuscripts as a senior author, reviewed, or handled a manuscript. We have clarified the specificity of this analysis to the ASM journals and noted our use of identifiers to conduct the analysis in the text.

**However, the analysis does reveal that higher proportions of male junior scientists are being placed in these perceived higher positions. The potential interpretation for this is that there is mentoring bias in favor of junior scientists (postdocs and researchers) to be promoted by established faculty to become corresponding authors and members of editorial boards. Thus, the data provides potential support for the existence of mentoring bias, which is a key factor in gender disparity for academic success.**

We have added the possibility of mentoring bias to the discussion.

**5. The authors state the proportion of each institutional types for Figure 1 and Figure 4 in the text, but it would be helpful if they included a figure (like a venn diagram) to depict these percentages, instead of the numbers in the y-axis.**

The proportion of each gender at the institution is indicated by the x-axis while the number in parentheses on the y-axis are the total number of individuals (e.g., editor, reviewer, author) counted at the institution type. We have clarified this in the figure captions.

**6. The authors should indicate the figures that represent the AUROC metric and analysis result described in the text.**

We have added the regression analysis results in Figure S5, which includes a box plots indicating the range of AUROC values for each analysis as well as the logistic regression variables and their weights for the first two analyses. (The variables for the third analysis is described in Figure S6.)