PEER REVIEW HISTORY

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

(This paper received three reviews from its previous journal and three reviewers agreed to published their review.)

ARTICLE DETAILS

TITLE (PROVISIONAL)	A cross-sectional study of the relationship between women's representation amongst editors and peer reviewers in journals of the British Medical Journal Publishing Group
AUTHORS	Pinho-Gomes, Ana-Catarina; Vassallo, Amy; Woodward, Mark; Peters, Sanne

VERSION 1 – REVIEW

REVIEWER	Michael Mensah Yale University
REVIEW RETURNED	01-Jan-2022

GENERAL COMMENTS	Summary: This research article examines gender representation among editors and peer reviewers for 46 of the 74 British Medical Journal publishing group in 2020 and BMJ from 2009, 2010 and 2013-17 and using publicly available data—a scarce resource for this parameter. Overall, women are ~30% of peer reviewers, and a progressively smaller percentage of editors and editors in chief. There is a correlation between journals with women editor-in-chief and greater women representation among peer reviewers, although this association does not consider impact factor, age of journal, specialty of the journal, or several other factors that may influence peer reviewer and editor decision. The authors then recommend 6 interventions for gender equity among journals. Impression: This manuscript engages a very important topic in medicine today. It makes a strong case that the BMJ-PG has ~30% women peer reviewers and fewer editors and editors-in-chief. The authors should clarify their methods, especially with regard to the years of analysis for all journals besides BMJ. The paper could go further by putting their results in greater context earlier with the JAMA and NEJM data points they present, especially since they have similar data for JAMA and NEJM as they do the 46 BMJ-PG journals (besides BMJ). More context could include baseline gender representation among academic physicians in US and UK, although this assumes that the peer reviewer and editor pool has negligible contributions from academicians from other countries. Additionally, the discussion's solutions to gender inequity should start with the insightful point

	about technology selecting for already established authors as reviewers, and incorporate evidence when suggesting diversity trainings, which have drawn ire in the United States and in some literature. Specific comments: Page 6 line 7: Equally distributed in comparison to which pool of people? Among all reviewers and editors? All academic physicians in the UK? US and UK? Page 6 line 15-17: I'm not sure what the authors mean in this important sentence. Do they mean that they correlated percentage of women as peer reviewers to percentage of women as editors overall? I'm not sure what this would mean--how closely the peer reviewer sample matches the editor sample by gender? Page 7 line 48-52: Can the authors credibly claim there was a slight increase in peer reviewers if the confidence intervals overlap? The data suggest they should instead claim that there was no statistically significant increase. Line 53-54: At significance level $p<0.05$, these results (correlation between citescore/impact factor and woman representation among reviewers) are not significant, so the authors cannot claim a weak negative association. Page 8 line 3-5: 1) the authors should specify again that NEJM and JAMA percentages were in different years (2020 compared to 2010-2017 in the BMJ sample) 2) these results are noteworthy and should probably get higher priority as a comparison point in other sections of the paper (women as editors for example) Page 9 line 8: the authors should give the exact time period investigated, since some readers might assume this includes 2021, 2020 (which is does not). Line 46: In general, the authors should stay away from saying "both genders" as if there are only two genders. As a social construct, gender has evolved beyond the man/woman binary for many readers Line 52: "peer reviewers are usually senior researchers or leaders in their field" should be substantiated with a citation. Page 10 line 1: The authors should refrain from language like "less able", since it connotes an essentialism upon which gender bias thrives. Line 45-47: The authors are right to advocate for more gender inclusion, but using explicit quotas may draw ire from anti-affirmative action advocates in the united states, where supreme court case law (Bakke v. University of California Davis medical school) supports the claim that this suggestion is not the way forward. Line 52-54: The authors should advocate for longitudinal, multimodal evidence based training, since a large meta-analysis of 260 studies of diversity training found a significant overall impact from training (hedge's g (similar to Cohen's d)=0.39) especially from a cognitive standpoint, and that multimodal approaches were most effective (Bezrukova, K. , Spell, C. , Perry, J. & Jehn, K. (2016). A Meta-Analytical Integration of Over 40 Years of Research on Diversity Training Evaluation. Psychological Bulletin, 142 (11), 1227-1274. doi: 10.1037/bul0000067) Page 11 line 18-27: The authors should discuss this cause earlier since it is not well known to those unfamiliar with journal

	infrastructure and likely has a huge impact on who gets to review papers. Line 51: The authors should be more cautious about saying the findings are likely generalizable, possibly by saying that the findings may be generalizable to large journals

REVIEWER	Paula Chatterjee University of Pennsylvania
REVIEW RETURNED	01-Jan-2022

GENERAL COMMENTS	Comments: Thank you for the opportunity to review this manuscript. Using data collected from journals in the BMJ Publishing group, the authors have examined gender disparities in representation among peer reviewers. They find that women comprise approximately 30% of peer reviewers and 30% of editors. They also suggest that women editors are associated with higher percentages of women as peer reviewers. Taken together, they argue that these findings further the need for gender equity at all levels of the academic publishing process. As the authors acknowledge, several of their findings are consistent with prior work including the underrepresentation of women as editors and peer reviewers (refs #6, 9-10, 13-15). The potential novelty of this study lies in the correlations between gender representation of peer reviewers and editors. These correlations are limited by the few female editors in the sample, but are still somewhat interesting. Most of the story can be told in Figure 1, with some edits recommended below. Since the contributions of the other figures are minimal in terms of providing novel evidence of gender disparities in academia, this study may be better suited as a shorter article with a tighter focus on the potential relationship between the gender of editors and representation among peer reviewers. Minor points: -Introduction: It is unclear how including more women as peer reviewers will “ensur[e] [that] scientific publications adhere to reporting standards and guidelines, including those for the incorporation of sex and gender analyses.” A stronger argument for improving the gender gap among peer reviewers might be centered on the importance of reviewer service as a stepping stone to journal leadership positions, in academic promotion, and in signaling professional expertise in a given field. -Introduction: Are there plausible reasons to suspect that more women editors would not be correlated with more women peer reviewers? This seems strongly supported by network theories of homophily. If not, then this study may be better framed as testing a one-tailed hypothesis. -Methods: Were data only available or only collected for 2020 for NEJM and JAMA? They appear to be publicly available in other years (https://jamanetwork.com/journals/jama/fullarticle/1667096), in which case, is there a reason those other years were not included? This would help improve generalizability beyond the BMJ family, though I recognize this is a non-trivial data collection effort. -Methods: Additional details are needed on the 4% of peer reviewers

	for whom gender was not identified, and the 24% of missing gender assignments at BMJ Case Reports. Genderize provides a probability of correct gender assignment for each name. Was there a probabilistic threshold applied to include/exclude assigned values? If so, how was that threshold determined? Were non-classified names treated differently from names with low probabilities of confident assignment? Along these same lines, why was Genderize used only to assign genders to reviewers, and not also used to assign genders to editors? -Methods: For what journals are Impact Factor and CiteScore not highly correlated? -Figure 1: The patterns would be better described as a 2-panel figure, stratifying journals by whether men or women were in EIC roles. It would also be helpful to scale the size of the points to reflect the number of peer reviewers or the number of editors at a journal. I would also recommend labeling the data points for the journals with the 10 highest impact factors, as the impact factor/CiteScore correlations are limited for the reasons described below. I would also add the correlation coefficient to the plot itself. -Figure 2: What is the hypothesis being tested to support the use of confidence intervals? Given that the entire universe of BMJ family journals with available data on peer reviewers was included, it would be helpful to specify the reason for including CI's on this plot. In their absence, it seems that there is a small, overall increase in the percent of women reviewers over time. -Figures 3&4: Other than highlighting the outlier journal with the highest impact factor, these figures do not add much information to the story. -Discussion: The authors recognize that women have non-professional responsibilities that may influence their ability to perform peer reviews. However, there is also literature to support the fact that women are also more likely to provide unpaid professional service in the form of departmental responsibilities (Guarino CM & Borden V. Research in Higher Education; 2017; 58:672–694). Given these prior findings, it would be useful to further discuss the ramifications of narrowing the gender gap in peer reviewership: how should reviewership, another form of essential but unpaid professional service, be weighed in the context of other unpaid professional service that women are already more likely to provide? -Discussion: Prior literature has raised concerns about the limited efficacy of implicit bias trainings in terms of improving representation or having durable effects on implicit bias, itself (Onyeador I, Hudson STJ, Lewis Jr NA. Policy Insights from the Behavioral & Brain Sciences; 2021; 8(1):19-26). While they have primarily evaluated training specific to racial bias, these limitations may also be relevant to training on gender bias. -Discussion: The authors might consider how their findings should be contextualized with other recent work suggesting multiple other sources of gender bias in the editorial process, beyond equitable representation among editors: (preprint) Malkinson TS et al. Gender Imbalance in the Editorial Activities of a Researcher-led Journal. doi: https://doi.org/10.1101/2021.11.09.467796.

REVIEWER	James Salazar UCSF
REVIEW RETURNED	01-Jan-2022

GENERAL COMMENTS	Comments: In this manuscript, Pinho-Gomes et al. use publicly available data on peer reviewers and editors at the British Medical Journal Publishing Group to infer their gender (for peer reviewers, using an algorithm genderize and for editors, using available photographs/pronouns) and evaluate the representation of women among peer reviewers/editors (and trends in representation over time at BMJ), and query if representation was associated with the representation of editors/differed by the inferred gender of the editor-in-chief. The investigated topic is highly important. Peer review informs the development and the selection of research and opinion articles that get published, and a lack of gender parity may cause significant inequities and bias in the research that is published (not to mention inequity in the career opportunities that are often afforded to peer reviewers - e.g. potential offers to write editorials). Thus, the authors' intent to characterize the gender breakdown of peer reviewers among such a large sample of impactful journals should be commended. The authors find that these journals have a long way to go towards gender parity - with low representation of women among peer reviewers and editors. As the authors acknowledge, these disparities are known and are similar to other previously published data. However, their analyses comparing representation of peer reviewers to representation of editors/gender of editor in chief are more novel. Overall, I think the studied population is significant and the findings would be helpful to continue to propagate change. That said, my major concern/reservation is that the authors infer the gender of peer reviewers/editors using a program/photo/pronoun. Research using these methods is increasingly commonplace and exists because the question of the diversity of peer reviewers/editorial teams (along with other populations) is important yet self-reported data on these populations and social identities is not readily available. In this case, even if we accept that the genderize program does well at classifying gender, it does not acknowledge non-binary genders and still assumes a person's identity. Furthermore, the use of photos in ascribing gender or other identities is problematic. These data may help further change and equity in scientific research and publishing, but I worry considerably about the methods/means. I've provided further comments below in case it may help the development. Overall, the BMJ (and their readership) may be best suited to conduct their own survey of peer reviewers/editors and report on inclusive, self-reported demographics of these important populations. Thank you to the authors and editors for the opportunity to review this work. James Salazar MD MAS Specific Comments: • Introduction o Would preview the study population to orient the reviewers. E.g. "We investigated the BMJ Publishing Group which publishes in a wide variety of (largely medical?) fields.”

	o Would clarify and elaborate the aims. The first and most significant contribution (to me) is characterizing the percentage of women among peer reviewers at these impactful journals. A secondary aim would be to investigate the association between representation as editors and peer reviewers. • Methods o Is it true (or relevant in the methods) that the BMJ PG is the only group that reports peer reviewers annually? Perhaps the authors could just comment that the BMJ publishing group was selected as the study population and then comment in the discussion about issues of generalizability/the lack of available data from other publishing families in the discussion. o Was this a recent decision to report peer reviewers by BMJ PG? Is that the reason why data is only available in 2020? If so I'd mention that explicitly in the methods. o I'd state “for comparison” rather than “to compare whether findings were generalizable to other journals.”NEJM/JAMA are distinct from many other journals for a variety of reasons and the authors can delve into the limitations of this comparison in the discussion. o Were assistant editors also included (would suspect that other editors were included since there were only 4% of editors excluded)? How many editors had available photographs/pronouns? • Results o Could some general characteristics of the included journals be provided? Summary CiteScore or ImpactFactor for those available? Type of field (e.g. cardiology, general medicine) - this would also allow for further sub-analysis as disparities may vary by field (acknowledging journals often don't fit neatly into categories)? Average number of peer reviewers per journal? Overall, this would be helpful orientation for readers in interpreting the findings. o Pg 7 line 15 ISAKSOS should be instantiated. o Median and interquartile range of % women as peer reviewers and editors by journal may also be informative data to include to demonstrate alluded "marked variation” o Is data for 2020 not available at BMJ? May want to explain why not since it is available for many of the others • Discussion o Consider re-write: although gender parity in the editorial board was achieved in twelve journals □ Women accounted for greater than 50% of editors in only 12 of 47 journals (~25%) - I think this is what is intended? And to me is notable as fairly few journals. o The journal editorial board often is different than the decision-making editors included in the study so would avoid use of the word editorial board o Re: limitations, would elaborate on the limitations of using genderize/ photos/pronouns in predicting gender o Would soften the utility of NEJM/JAMA as comparators to the sampled population as they are distinct from many of the journals in the included population and journals at large. o May be helpful to provide data for the state of gender parity in potential reference populations (e.g. population as a whole, European physicians/scientists (both non-academic and academic) o Consider citing Gender, Race, Ethnicity, and Sexual Orientation of Editors at Leading Medical and Scientific Journals JAMA Intern Med. 2021;181(9):1248-1251. doi:10.1001/jamainternmed.2021.2363 as another reference point on contemporary self-reported data re: gender identity of editors at comparable journals • Conclusion o It should be clarified that“No evidence of improvement over the

	past decade" applies only to BMJ since that was the only journal there was temporal data for. o Re: the last line, would contextualize this as only one of many factors needed to be change in the pursuit of equity.

VERSION 1 – AUTHOR RESPONSE

Answers to reviewers' comments

Reviewer: 1

We thank Reviewer 1 for their comments.

Summary:
This research article examines gender representation among editors and peer reviewers for 46 of the 74 British Medical Journal publishing group in 2020 and BMJ from 2009, 2010 and 2013-17 and using publicly available data—a scarce resource for this parameter. Overall, women are ~30% of peer reviewers, and a progressively smaller percentage of editors and editors in chief. There is a correlation between journals with women editor-in-chief and greater women representation among peer reviewers, although this association does not consider impact factor, age of journal, specialty of the journal, or several other factors that may influence peer reviewer and editor decision. The authors then recommend 6 interventions for gender equity among journals.
Impression:
This manuscript engages a very important topic in medicine today. It makes a strong case that the BMJ-PG has ~30% women peer reviewers and fewer editors and editors-in-chief. The authors should clarify their methods, especially with regard to the years of analysis for all journals besides BMJ. The paper could go further by putting their results in greater context earlier with the JAMA and NEJM data points they present, especially since they have similar data for JAMA and NEJM as they do the 46 BMJ-PG journals (besides BMJ). More context could include baseline gender representation among academic physicians in US and UK, although this assumes that the peer reviewer and editor pool has negligible contributions from academicians from other countries. Additionally, the discussion's solutions to gender inequity should start with the insightful point about technology selecting for already established authors as reviewers, and incorporate evidence when suggesting diversity trainings, which have drawn ire in the United States and in some literature.

Specific comments:
Page 6 line 7: Equally distributed in comparison to which pool of people? Among all reviewers and editors? All academic physicians in the UK? US and UK?
That sentence in the methods means equally distributed between genders in comparison to the global population, among whom about 50% are women and 50% are men. We do not think it would be appropriate compare with physicians in the US and UK or with clinicians in the US and the UK for several reasons. First, it is impossible to reliably determine the proportion of women among academic clinicians, which is in itself an ill-defined concept. Second, peer reviewers do not necessarily come from the UK and US or clinicians/physicians, which means such comparison would be reductionist and invalid.

Page 6 line 15-17: I'm not sure what the authors mean in this important sentence. Do they mean that they correlated percentage of women as peer reviewers to percentage of women as editors overall? I'm not sure what this would mean--how closely the peer reviewer sample matches the editor sample by gender?

We tested whether there was a correlation between women's representation among editors and peer reviewers for each journal. We amended the paragraph in the methods explaining the data analysis as it was not very clear.

Page 6, line 8
We computed the percentage of women among peer reviewers and editors overall and for each journal. We plotted the association between the percentage of women as peer reviewers and editors, stratified by gender of the editor-in-chief. We computed the Spearman correlation coefficient between the percentage of women as editors and the percentage of women as peer reviewers. We also calculated the Spearman correlation coefficient between the percentage of women as peer reviewers and the journal impact factor and Citescore. We compared the percentage of women among peer reviewers and editors according to the gender of the editor-in-chief using Fisher's exact test.

Page 7 line 48-52: Can the authors credibly claim there was a slight increase in peer reviewers if the confidence intervals overlap? The data suggest they should instead claim that there was no statistically significant increase.
Following the comments of another reviewer, we decided to remove all confidence intervals. This is because we used all the reviewers rather than a random sample of reviewers. Therefore, estimating 95% confidence intervals did not make sense. We amended the manuscript as follows:

Page 7, line 30
The percentage of women as peer reviewers increased slightly from 27.3% in 2010 to 29.7% in 2017 in the BMJ, from 23.9% in 2010 to 28.1% in 2020 in JAMA, and from 16.9% in 2010 to 18.9% in 2020 in the NEJM (Figure 2 and Table S2).

Line 53-54: At significance level $p<0.05$, these results (correlation between citescore/impact factor and woman representation among reviewers) are not significant, so the authors cannot claim a weak negative association.
We amended the manuscripts as follows:

Page 7, line 29
There was a non-significant negative correlation between the impact of the journal and the percentage of women as peer reviewers (Figure 3).

Page 8 line 3-5:
1) the authors should specify again that NEJM and JAMA percentages were in different years (2020 compared to 2010-2017 in the BMJ sample)
The only journal for which data refer to 2010 to 2017 is the BMJ. For all the other journals in the BMJ Publishing Group, data are for 2020 as for the NEJM and JAMA. We removed the sentence referred to by the reviewer as it no longer made sense after amending the previous sentence on the lack of a significant correlation between journal impact and women's representation as peer reviewers. We clarified the dates available for each journal in the Methods.

Page 5, line 9
For all journals of the BMJ-PG, apart the BMJ, data were available only for 2020, and we used those data to investigate current representation of women in the BMJ-PG overall.
In addition, we investigated trends over time in women's representation using data available for peer reviewers in the BMJ for 2009, 2010, and 2013 to 2017. Data were not available for the BMJ after 2017. For comparison, the list of peer reviewers in 2010, 2012, 2014, 2016, 2018, and 2020 for two leading medical journals (The New England Journal of Medicine (NEJM) and Journal of the American Medical Association (JAMA)) were also reviewed and given names of reviewers extracted. These two journals were not included in the analysis of BMJ-PG journals.

2) these results are noteworthy and should probably get higher priority as a comparison point in other sections of the paper (women as editors for example)
We have emphasised the comparison with the journals used as external comparators in the Results.

Page 7, line 10
Women's representation among peer reviewers in the BMJ-PG was higher than in the JAMA (28.1%) and the NEJM (18.9%).

Page 9 line 8: the authors should give the exact time period investigated, since some readers might assume this includes 2021, 2020 (which is does not).
We agree and have amended the text as follows:

Page 9, line 4
(…) no evidence of a meaningful change between 2009 and 2017 in the BMJ.

Line 46: In general, the authors should stay away from saying "both genders" as if there are only two genders. As a social construct, gender has evolved beyond the man/woman binary for many readers
We discuss in greater detail the issue of not accounting for non-binary gender identification in the limitations. We clarified this sentence as follows:

Page 9, line 27
Indeed, a previous study demonstrated that editors have substantial same-gender preference when selecting peer reviewers irrespective of whether they are women or men.

Line 52: "peer reviewers are usually senior researchers or leaders in their field" should be substantiated with a citation.
We added the following reference:

Garisto D. Diversifying peer review by adding junior scientists. Nature index 2019. https://www.natureindex.com/news-blog/diversifying-peer-review-by-adding-junior-scientists (accessed 1 January 2022).

Page 10 line 1: The authors should refrain from language like "less able", since it connotes an essentialism upon which gender bias thrives.
This is an important point as what we meant was that women experience barriers not that they inherently lack ability. We amended the text as follows:

Page 9, line 34
Third, it is possible women face barriers that prevent them from accepting invitations to take part in the peer review process due to competing demands.

Line 45-47: The authors are right to advocate for more gender inclusion, but using explicit quotas may draw ire from anti-affirmative action advocates in the united states, where supreme court case law (Bakke v. University of California Davis medical school) supports the claim that this suggestion is not the way forward.
We acknowledge that quotas are not the perfect solution and have potential shortcomings. However, in many areas, quotas have improved women's representation, thus suggesting they may be a necessary step whilst societal values are changing and gender bias gradually fading away. We amended the manuscript to reflect the fact that gender quotas are unlikely by themselves to fix the gender gap.

Page 11, line 10
However, evidence from a researcher-led journal suggests improving women's representation (e.g., by gender quotas) may not be enough to stem deep-rooted gender bias observed along the editorial process.[33] For instance, senior editors and authors were more likely to select men than women as reviewing editors, even after correcting for the gender imbalance in the pool of reviewing editors available.[33]

Line 52-54: The authors should advocate for longitudinal, multimodal evidence based training, since a large meta-analysis of 260 studies of diversity training found a significant overall impact from training (hedge's g (similar to Cohen's d)=0.39) especially from a cognitive standpoint, and that multimodal approaches were most effective (Bezrukova, K. , Spell, C. , Perry, J. & Jehn, K. (2016). A Meta-Analytical Integration of Over 40 Years of Research on Diversity Training Evaluation. Psychological Bulletin, 142 (11), 1227-1274. doi: 10.1037/bul0000067)
We thank the Reviewer for this very useful suggestion. We have amended the text as follows:

Page 11, line 16
Although equality and diversity training is not a silver bullet to address longstanding gender inequalities,[34] it may have benefits on cognitive, behavioural and attitudinal/affective learning, especially when complemented by other initiatives targeted to both awareness and skills development, and conducted over a significant period of time.[35]

Page 11 line 18-27: The authors should discuss this cause earlier since it is not well known to those unfamiliar with journal infrastructure and likely has a huge impact on who gets to review papers.
We have moved this cause to the beginning of the paragraph:

Page 10, line 34
First, editors should be made aware of the inherent properties of software tools available to help them find suitable peer reviewers (…)

Line 51: The authors should be more cautious about saying the findings are likely generalizable, possibly by saying that the findings may be generalizable to large journals
The 47 journals of the BMJ-PG included in this study encompass a broad range of journals in terms of 'size' and 'impact'. For instance, the impact factor ranges from 1.6 to 38.9 and Citescore from 1 to 35.6. This together with the comparable (if not slightly worse) findings for the NEJM and JAMA make as believe women's under-representation among peer reviewers is not an issue specific to the BMJ-PG but a widespread problem among medical journals. This is also corroborated by previous studies, as we explain in the discussion. We softened that sentence as follows:

Page 12, line 16
Third, we used journals from a single publishing family, which might not be representative of all medical journals. However, comparable results from two leading journals from different publishers together with previous reports from other journal families suggest our findings might reflect the magnitude of women's under-representation among peer reviewers of medical journals.[15 16]

Reviewer: 2

We thank Reviewer 2 for their comments.

Comments:

Thank you for the opportunity to review this manuscript.
Using data collected from journals in the BMJ Publishing group, the authors have examined gender disparities in representation among peer reviewers. They find that women comprise approximately 30% of peer reviewers and 30% of editors. They also suggest that women editors are associated with higher percentages of women as peer reviewers. Taken together, they argue that these findings further the need for gender equity at all levels of the academic publishing process
As the authors acknowledge, several of their findings are consistent with prior work including the underrepresentation of women as editors and peer reviewers (refs #6, 9-10, 13-15). The potential novelty of this study lies in the correlations between gender representation of peer reviewers and editors. These correlations are limited by the few female editors in the sample, but are still somewhat interesting.
Most of the story can be told in Figure 1, with some edits recommended below. Since the contributions of the other figures are minimal in terms of providing novel evidence of gender disparities in academia, this study may be better suited as a shorter article with a tighter focus on the potential relationship between the gender of editors and representation among peer reviewers.

Minor points:
-Introduction: It is unclear how including more women as peer reviewers will “ensur[e] [that] scientific publications adhere to reporting standards and guidelines, including those for the incorporation of sex and gender analyses.” A stronger argument for improving the gender gap among peer reviewers might be centered on the importance of reviewer service as a stepping stone to journal leadership positions, in academic promotion, and in signaling professional expertise in a given field.
There is evidence that having women as authors of papers increases the likelihood of adhering to standards and guidelines related to sex and gender disaggregated analysis as well as women’s representation among participants in RCTs. This is a benefit for medicine and ultimately those served by medical research, e.g., patients. We understand the reviewer’s point about the advantages for women’s careers and we added it to the discussion, where we also expand on the advantages of equality and diversity at all levels in the scientific community.

Page 4, line 9
Peer reviewers also have an important role in ensuring scientific publications adhere to reporting standards and guidelines, particularly those for the incorporation of sex and gender analyses.[8] Since women as authors are more likely to report sex and gender analyses, women as peer reviewers may also be more likely than men to ensure sex and gender are adequately handled in medical papers.[9]

Page 10, line 15
The wider benefits of gender equality for science and medicine have been compelling demonstrated for men as well as women.[24 25] Indeed, a research community that is more inclusive, diverse, and representative, and works to ensure that everyone counts, is more likely to generate research that is universally beneficial and not limited by inequalities.[26]

Page 10, line 28
On the other hand, women’s under-representation as peer reviewers may be both a symptom and a cause of broader under-representation in senior positions in academia and journals as taking part in the peer review process can be a career milestone and a stepping stone to leadership roles.[27 28]

-Introduction: Are there plausible reasons to suspect that more women editors would not be correlated with more women peer reviewers? This seems strongly supported by network theories of homophily. If not, then this study may be better framed as testing a one-tailed hypothesis.
We understand the reviewer's point. There is evidence that editors have a predisposition to make same-gender choices for peer reviewers as we explain and cite in the discussion. We amended the aim of this study as follows:

Page 4, line 23
Therefore, this study aimed to determine women's representation among peer reviewers and editors of medical journals and investigate whether greater women's representation among editors correlated with greater representation as peer reviewers.

-Methods: Were data only available or only collected for 2020 for NEJM and JAMA? They appear to be publicly available in other years (https://jamanetwork.com/journals/jama/fullarticle/1667096), in which case, is there a reason those other years were not included? This would help improve generalizability beyond the BMJ family, though I recognize this is a non-trivial data collection effort.
We included data for the NEJM and JAMA for 2020 as the aim was to compare with the average women's representation in the BMJ-PG, for which data were available for 2020 only. However, we understand the reviewer's point and hence added data for previous years for the NEJM and JAMA to compare with trends over time for the BMJ.

-Methods: Additional details are needed on the 4% of peer reviewers for whom gender was not identified, and the 24% of missing gender assignments at BMJ Case Reports. Genderize provides a probability of correct gender assignment for each name. Was there a probabilistic threshold applied to include/exclude assigned values? If so, how was that threshold determined? Were non-classified names treated differently from names with low probabilities of confident assignment? Along these same lines, why was Genderize used only to assign genders to reviewers, and not also used to assign genders to editors?
We apologise for the typo in the table as the percentage of missing gender assignments in BMJ case reports was 1% not 24%. Therefore, the overall missing gender assignments was also 1% rather than 4%. We corrected this in the table and throughout the manuscript.
The number of editors was relatively small compared with the number of authors and further details were available to identify them, such as affiliation. This allowed checking pronouns used on institutions websites. A similar approach was infeasible for the almost 43,000 peer reviewers. Therefore, we had to use the GenderizeR package. Despite its limitations, which we describe in the discussion, it is the best available approach to determine gender based on given names. Although this package uses the genderize.io database, it functions slightly differently. First, we extract given names from the full names using a specific command and this output is then fed into the command that determines the gender. This improves the accuracy of the gender determination, which explains why we have less than 1% of missing genders. However, this command classifies given names and woman or man without giving a 'probability' as happens when the genderize.io website is used. All given names that the package was unable to classify as woman or men were reported as "missing" and excluded from the analysis. We clarified this in the methods and also in the discussion.

Page 5, line 18
We used the "genderizeR" package for R to predict the gender of the peer reviewers based on their given names. This software collects data from the Internet and includes 38,659 given names from 242 countries across the globe.[13] A two-step approach was used to determine gender based on given names.[14] First, given names were extracted from full names using a specific command of the GenderizeR package. Second, the gender of the vector of given names was classified as either woman or man using another command. When given names could not be recognised and extracted

from full names by the software, those reviewers were considered as “missing” and excluded from all analyses.

Page 12, line 11
Second, the genderizeR package could not assign a gender to less than 1% of the peer reviewers because the given name could not be classified as belonging to a woman or a man. However, we adopted a two-step approach to maximise the efficiency of the package, and hence the minimal percentage of missing data is unlikely to have had a material impact on our key findings.[14]

-Methods: For what journals are Impact Factor and CiteScore not highly correlated?
The reason we used Citescore and Impact factor was because many journals are not indexed in Web of Science and, hence, don’t have an impact factor. Using only impact factor would have meant losing data for many journals, some of which have a Citescore. Therefore, we decided to use both metrics of ‘impact’ to test for a correlation between journal impact and women’s representation. So, the point is not whether they are correlated. They are complementary as they rely on different data as explained in the methods. We amended the manuscript as follows:

Page 6, line 4
We used these two metrics to assess impact because impact factor was not available for 21 journals, of which 15 had a Citescore available.

-Figure 1: The patterns would be better described as a 2-panel figure, stratifying journals by whether men or women were in EIC roles. It would also be helpful to scale the size of the points to reflect the number of peer reviewers or the number of editors at a journal. I would also recommend labeling the data points for the journals with the 10 highest impact factors, as the impact factor/CiteScore correlations are limited for the reasons described below. I would also add the correlation coefficient to the plot itself.
We scaled the size of the dots according to the number of peer reviewers as this is the main outcome of this study. We added the correlation coefficient to the figure legend. Please see the revised Figure 1. Although we understand the reviewer’s points, we feel that a 2-panel figure would be appropriate since there are very few journals with a woman as editor-in-chief. Moreover, the correlation was tested for all journals, so it makes sense to have all journals in the same plot. We did not label the points as it was too confusing due to the high number of points, and it didn’t seem appropriate to us to label some of them using an arbitrary criterion.

-Figure 2: What is the hypothesis being tested to support the use of confidence intervals? Given that the entire universe of BMJ family journals with available data on peer reviewers was included, it would be helpful to specify the reason for including CI’s on this plot. In their absence, it seems that there is a small, overall increase in the percent of women reviewers over time.
We agree with the reviewer. We have revised our methodological approach and removed all the confidence intervals from the manuscript as we included the entire population of peer reviewers. We have amended this throughout the manuscript. We added JAMA and the NEJM to the new Figure 2, as requested by the reviewer above.

-Figures 3&4: Other than highlighting the outlier journal with the highest impact factor, these figures do not add much information to the story.
We agree with the reviewer and have moved those figures to supplementary data.
-Discussion: The authors recognize that women have non-professional responsibilities that may influence their ability to perform peer reviews. However, there is also literature to support the fact that women are also more likely to provide unpaid professional service in the form of departmental responsibilities (Guarino CM & Borden V. Research in Higher Education; 2017; 58:672–694). Given these prior findings, it would be useful to further discuss the ramifications of narrowing the gender gap

in peer reviewership: how should reviewership, another form of essential but unpaid professional service, be weighed in the context of other unpaid professional service that women are already more likely to provide?
We added this to the discussion on possible causes for the gender gap among peer reviewers.

Page 10, line 1
Deeply entrenched gendered roles in our contemporary societies mean women still bear the brunt of homemaking, childcare, other unpaid care roles.[20 21] Furthermore, women undertake a greater share of internal service in academic institutions (e.g., activities related to faculty governance, faculty recruitment, evaluation and promotion, student admissions and scholarships, programme supervision, development and marketing, internal awards) in comparison to men.[22] Taken together, these additional commitments reduce women's availability to engage with scholarly activities with unscheduled and tight deadlines, such as peer review.

-Discussion: Prior literature has raised concerns about the limited efficacy of implicit bias trainings in terms of improving representation or having durable effects on implicit bias, itself (Onyeador I, Hudson STJ, Lewis Jr NA. Policy Insights from the Behavioral & Brain Sciences; 2021; 8(1):19-26). While they have primarily evaluated training specific to racial bias, these limitations may also be relevant to training on gender bias.
Although we understand reviewer's point, a large meta-analysis showed that EDI training can have positive effects. Therefore, we amended the text as follows:

Page 11, line 16
Although equality and diversity training is not a silver bullet to address longstanding gender inequalities,[34] it may have benefits on cognitive, behavioural and attitudinal/affective learning, especially when complemented by other initiatives targeted to both awareness and skills development, and conducted over a significant period of time.[35]

-Discussion: The authors might consider how their findings should be contextualized with other recent work suggesting multiple other sources of gender bias in the editorial process, beyond equitable representation among editors: (preprint) Malkinson TS et al. Gender Imbalance in the Editorial Activities of a Researcher-led Journal. doi: https://doi.org/10.1101/2021.11.09.467796.
We added this to the discussion:

Page 11, line 11
However, evidence from a researcher-led journal suggests improving women's representation (e.g., by gender quotas) may not be enough to stem deep-rooted gender bias observed along the editorial process.[33] For instance, senior editors and authors were more likely to select men than women as reviewing editors, even after correcting for the gender imbalance in the pool of reviewing editors available.[33]

Reviewer: 3

We thank Reviewer 3 for their comments.

Comments:
In this manuscript, Pinho-Gomes et al. use publicly available data on peer reviewers and editors at the British Medical Journal Publishing Group to infer their gender (for peer reviewers, using an algorithm genderize and for editors, using available photographs/pronouns) and evaluate the representation of women among peer reviewers/editors (and trends in representation over time at BMJ), and query if representation was associated with the representation of editors/differed by the inferred gender of the editor-in-chief.
The investigated topic is highly important. Peer review informs the development and the selection of research and opinion articles that get published, and a lack of gender parity may cause significant inequities and bias in the research that is published (not to mention inequity in the career opportunities that are often afforded to peer reviewers - e.g. potential offers to write editorials). Thus, the authors' intent to characterize the gender breakdown of peer reviewers among such a large sample of impactful journals should be commended.
The authors find that these journals have a long way to go towards gender parity - with low representation of women among peer reviewers and editors. As the authors acknowledge, these disparities are known and are similar to other previously published data. However, their analyses comparing representation of peer reviewers to representation of editors/gender of editor in chief are more novel. Overall, I think the studied population is significant and the findings would be helpful to continue to propagate change.
That said, my major concern/reservation is that the authors infer the gender of peer reviewers/editors using a program/photo/pronoun. Research using these methods is increasingly commonplace and exists because the question of the diversity of peer reviewers/editorial teams (along with other populations) is important yet self-reported data on these populations and social identities is not readily available. In this case, even if we accept that the genderize program does well at classifying gender, it does not acknowledge non-binary genders and still assumes a person's identity. Furthermore, the use of photos in ascribing gender or other identities is problematic.
These data may help further change and equity in scientific research and publishing, but I worry considerably about the methods/means. I've provided further comments below in case it may help the development. Overall, the BMJ (and their readership) may be best suited to conduct their own survey of peer reviewers/editors and report on inclusive, self-reported demographics of these important populations.
Thank you to the authors and editors for the opportunity to review this work.
Specific Comments:

• Introduction
o Would preview the study population to orient the reviewers. E.g. "We investigated the BMJ Publishing Group which publishes in a wide variety of (largely medical?) fields."
We clarified this in the methods as it didn't not fit with the narrative in the introduction.

Page 5, line 3
Among the major families of journals, only the British Medical Journal (BMJ) Publishing Group (BMJ-PG) requires their journals to report annually a list of their contributing peer reviewers.[12] The BMJ-PG is a large family of journals, which cover most medical specialties as well as other fields of research related to health services (e.g., quality improvement and safety).

o Would clarify and elaborate the aims. The first and most significant contribution (to me) is characterizing the percentage of women among peer reviewers at these impactful journals. A

secondary aim would be to investigate the association between representation as editors and peer reviewers.
We amended the manuscript as follows:

Page 4, line 23
Therefore, this study aimed to determine women's representation among peer reviewers and editors of medical journals and investigate whether greater women's representation among editors correlated with greater representation as peer reviewers.

• Methods
o Is it true (or relevant in the methods) that the BMJ PG is the only group that reports peer reviewers annually? Perhaps the authors could just comment that the BMJ publishing group was selected as the study population and then comment in the discussion about issues of generalizability/the lack of available data from other publishing families in the discussion.
To the best of our knowledge, only the BMJ-PG publishes data on their peer reviewers for most of their journals. Other journals do it but not in a systematic way and hence it was not possible to include other families. We believe this explanation is important for readers to understand why we selected the BMJ-PG and excluded other families (e.g., JAMA, BMC, Frontiers, etc.). We explored the uncertainty about generalisability of our findings to other journals in the limitations.

Page 12, line 16
Third, we used journals from a single publishing family, which might not be representative of all medical journals. However, comparable results from two leading journals from different publishers together with previous reports from other journal families suggest our findings might reflect the magnitude of women's under-representation among peer reviewers of medical journals.[15 16]

o Was this a recent decision to report peer reviewers by BMJ PG? Is that the reason why data is only available in 2020? If so I'd mention that explicitly in the methods.
We don't know exactly when the decision was taken but data are only available for 2020 as we state in the methods.

o I'd state "for comparison" rather than "to compare whether findings were generalizable to other journals."NEJM/JAMA are distinct from many other journals for a variety of reasons and the authors can delve into the limitations of this comparison in the discussion.
We amended the text as follows:

Page 5, line 14
For comparison, the list of peer reviewers for 2010, 2012, 2014, 2016, 2018, and 2020 for two leading medical journals (The New England Journal of Medicine (NEJM) and Journal of the American Medical Association (JAMA)) were also reviewed and given names of reviewers extracted.

o Were assistant editors also included (would suspect that other editors were included since there were only 4% of editors excluded)? How many editors had available photographs/pronouns?
Gender was based on pronouns and photos for all editors using the journal website or the websites of their affiliations. We clarified this in the methods as below. The reference to Table 1 was not correct and we removed it (it referred to missing gender for peer reviewers). We apologise for this typo.

Page 5, line 26
For each journal, data for editor-in-chief, deputy editors, assistant editors, and associate editors were extracted. These are defined as "editors" throughout the manuscript. Their gender was determined based on pronouns and photographs available on the journal website or professional affiliations.

Other members of editorial boards (e.g., advisory editors, statistical advisors, emeritus editors) were excluded.

• Results
o Could some general characteristics of the included journals be provided? Summary CiteScore or ImpactFactor for those available? Type of field (e.g. cardiology, general medicine) - this would also allow for further sub-analysis as disparities may vary by field (acknowledging journals often don't fit neatly into categories)? Average number of peer reviewers per journal? Overall, this would be helpful orientation for readers in interpreting the findings.
There is only one journal or a maximum of two journals per field, so analysis stratified by field/specialty are not meaningful as most fields would have one journal only. We added a summary of the impact of journals and added this to Table 1, too. The number of peer reviewers and editors per journal is already included in Table 1.

Page 7, line 32
The impact factor of the journals varied between 1.7 for the European Journal of Hospital Pharmacy and 38.8 for the BMJ, and the Citescore ranged from 1 for BMJ Leader to 35.6 for Gut (Table 1).

o Pg 7 line 15 ISAKSOS should be instantiated.
We have corrected this.

o Median and interquartile range of % women as peer reviewers and editors by journal may also be informative data to include to demonstrate alluded "marked variation"
We added this to the results.

Page 7, line 7
There was marked variation in women's representation across journals (median 31.3%, interquartile range 24.5% to 38.5%) (…)

o Is data for 2020 not available at BMJ? May want to explain why not since it is available for many of the others.
There are no data available for the BMJ since 2017. We relied on publicly available data so we don't have explanations for which data are or are not available. We clarified this in the methods

Page 5, line 9
For all journals of the BMJ-PG, apart the BMJ, data were available only for 2020, and we used those data to investigate current representation of women in the BMJ-PG overall.
In addition, we investigated trends over time in women's representation using data available for peer reviewers in the BMJ for 2009, 2010, and 2013 to 2017. Data were not available for the BMJ after 2017. For comparison, the list of peer reviewers in 2010, 2012, 2014, 2016, 2018, and 2020 for two leading medical journals (The New England Journal of Medicine (NEJM) and Journal of the American Medical Association (JAMA)) were also reviewed and given names of reviewers extracted. These two journals were not included in the analysis of BMJ-PG journals.

• Discussion
o Consider re-write: although gender parity in the editorial board was achieved in twelve journals. Women accounted for greater than 50% of editors in only 12 of 47 journals (~25%) - I think this is what is intended? And to me is notable as fairly few journals.
Thank you for highlighting this important point. We amended the text as follows:

Page 8, line 6
Twelve journals (25%) had 50% or more women editors, and five journals had no women editors.

o The journal editorial board often is different than the decision-making editors included in the study so would avoid use of the word editorial board
We corrected this throughout the manuscript.

o Re: limitations, would elaborate on the limitations of using genderize/ photos/pronouns in predicting gender
We amended the manuscript as follows:

Page 12, line 5
First, we used a binary definition of gender of peer reviewers, which relied on predicting and assigning gender based on given names. Therefore, we did not account for non-binary gender or gender identities that did not match that of the given name and acknowledge that this method does not reflect the true diversity of the medical research community. Pronouns were used to determine gender of editors, and no they/them pronouns were present. However, it is still possible that non-binary gender identification was not reflected by the pronouns used on public websites. Ideally, future research should aim to investigate gender gaps based on self-identified gender, as has been done elsewhere.[39] Second, the genderizeR package could not assign a gender to less than 1% of the peer reviewers because the given name could not be classified as belonging to a woman or a man. However, we adopted a two-step approach to maximise the efficiency of the package, and hence the minimal percentage of missing data is unlikely to have had a material impact on our key findings.[14]

o Would soften the utility of NEJM/JAMA as comparators to the sampled population as they are distinct from many of the journals in the included population and journals at large.
We amended the manuscript as follows:

Page 12, line 16
Third, we used journals from a single publishing family, which might not be representative of all medical journals. However, comparable results from two leading journals from different publishers together with previous reports from other journal families suggest our findings might reflect the magnitude of women's under-representation among peer reviewers of medical journals.[15 16]

o May be helpful to provide data for the state of gender parity in potential reference populations (e.g. population as a whole, European physicians/scientists (both non-academic and academic)
We understand the reviewer's point. However, it is not clear which population to consider as reference for the pool of potential peer reviewers. Using European or American clinicians or scientists is not an adequate population. The gender gap in medicine, particularly in academic medicine, is well-known. Women's under-representation is more severe at senior than junior level, a phenomenon known as the 'leaky pipeline'. We mention this at several stages in the manuscript, but we avoided drawing comparisons with women's representation in other populations as we don't think this is relevant. We commented that women's under-representation among peer reviewers was similar to their under-representation among authors and editors.

o Consider citing Gender, Race, Ethnicity, and Sexual Orientation of Editors at Leading Medical and Scientific Journals JAMA Intern Med. 2021;181(9):1248-1251. doi:10.1001/jamainternmed.2021.2363 as another reference point on contemporary self-reported data re: gender identity of editors at comparable journals
We thank the reviewer for highlighting this paper, which we have now included in the limitations.

Page 12, line 5
First, we used a binary definition of gender of peer reviewers, which relied on predicting and assigning gender based on given names. Therefore, we did not account for non-binary gender or gender

identities that did not match that of the given name and acknowledge that this method does not reflect the true diversity of the medical research community. Pronouns were used to determine gender of editors, and no they/them pronouns were present. However, it is still possible that non-binary gender identification was not reflected by the pronouns used on public websites. Ideally, future research should aim to investigate gender gaps based on self-identified gender, as has been done elsewhere.[39]

• Conclusion
o It should be clarified that"No evidence of improvement over the past decade" applies only to BMJ since that was the only journal there was temporal data for.
We amended the manuscript as follows:

Page 12, line 29
Women account for less than one in three peer reviewers in BMJ-PG journals with no evidence of improvement between 2009 and 2017 in the BMJ.

o Re: the last line, would contextualize this as only one of many factors needed to be change in the pursuit of equity.
We amended the manuscript as follows:
Page 12, line 32
Therefore, promoting gender equality among editors may improve women's representation as peer reviewers, which is one among many necessary steps in the pursuit of equity in editorial and publishing systems.

VERSION 2 – REVIEW

REVIEWER	Mensah, Michael Yale University, NCSP
REVIEW RETURNED	07-Feb-2022

GENERAL COMMENTS	This is a much improved manuscript. I have only minor comments, described below. Page 8 line 55-60: It would be helpful if the authors could give citescores and impact factors for JAMA and NEJM, for reference. page 10 line 11-13: Women authors or women editors? Line 19-20: its worth noting--here or elsewhere--that the authors also used newer data, making the lack of change more disappointing and noteworthy.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Dr. Michael Mensah, Yale University
Comments to the Author:
This is a much improved manuscript. I have only minor comments, described below.

We thank Reviewer 1 for his comments.

Page 8 line 55-60: It would be helpful if the authors could give citescores and impact factors for JAMA

and NEJM, for reference.
The impact factor and Citescore for JAMA and the NEJM are provided in Table 1. We added this to the text, too.

Page 7, line 34
The impact factors of the NEJM and JAMA were 91.2 and 56.3, respectively. The Citescore of the NEJM was 80.6, and there was no Citescore for JAMA.

page 10 line 11-13: Women authors or women editors?
This was a typo. We have corrected this.

Line 19-20: its worth noting--here or elsewhere--that the authors also used newer data, making the lack of change more disappointing and noteworthy.
We have reinforced that point as follows:

Page 9, line 21
Furthermore, as we included more recent data, the lack of progress towards gender equity is disappointing.

VERSION 3 – REVIEW

REVIEWER	Mensah, Michael Yale University, NCSP
REVIEW RETURNED	03-Apr-2022

GENERAL COMMENTS	Thank you for writing this paper and responding to reviewer concerns.