



Comparison of the Proportions of Female and Male Corresponding Authors in Preprint Research Repositories Before and During the COVID-19 Pandemic

Mackenzie R. Wehner, MD, MPhil; Yao Li, MS; Kevin T. Nead, MD, MPhil

Introduction

There is substantial interest regarding the effects of the coronavirus disease 2019 (COVID-19) pandemic on academic productivity. Many researchers are working from home, and schools and childcare centers are closed, leading to childcare disruptions. Research has demonstrated that women in academia perform a greater proportion of domestic work than men, including in dual academic career partnerships.^{1,2} We hypothesized that remote working and disruptions to family support structures secondary to the pandemic may have been disproportionately associated with the academic productivity of women. We examined changes in the proportion of female corresponding authors in bioRxiv ([biorxiv.org](https://www.biorxiv.org)) and medRxiv ([medrxiv.org](https://www.medrxiv.org)), which are online archive and distribution services for unpublished preprint research in the life and health sciences, respectively, before vs during the COVID-19 pandemic.³

+ Invited Commentary

Author affiliations and article information are listed at the end of this article.

Methods

We conducted a cross-sectional study using web-scraped metadata from medRxiv from June 6, 2019 (inception), to May 5, 2020, and bioRxiv from January 1, 2019, to May 20, 2020. This study was determined to be exempt from institutional review board approval by MD Anderson Cancer Center. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.

We determined corresponding authors' gender for 49 924 of 51 249 articles (97%) using first name via Gender API, which includes data on more than 2 million names across 177 countries and has demonstrated superior performance vs similar gender inference services.⁴ We determined corresponding authors' institution country of origin using a combination of listed country, the Nature Index, and manual entry. We analyzed the most recent version of each article and included data from months with at least 100 unique submissions.

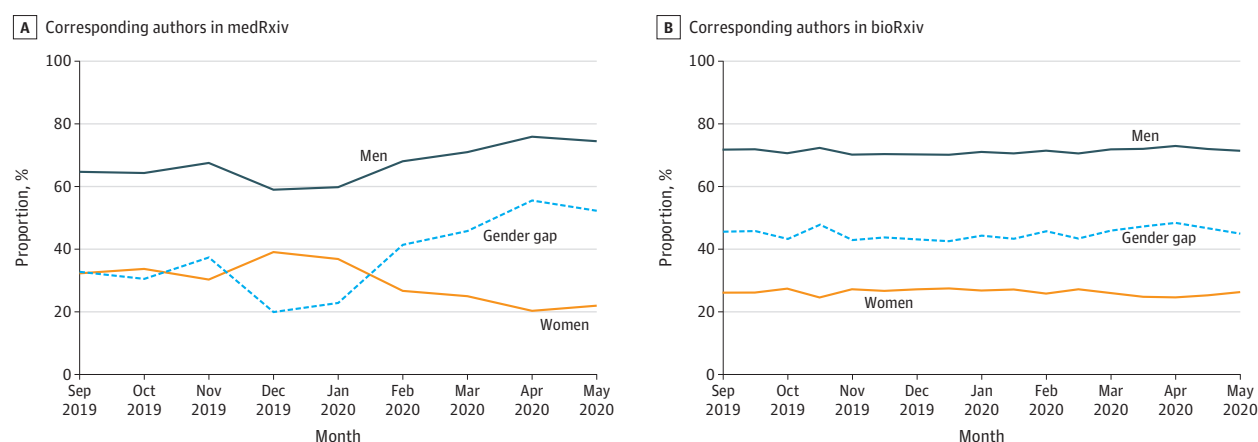
We conducted a test for trend using the Somers Delta (Somers D; *somersd*) in Stata, version 16.1 (StataCorp) to assess whether the absolute difference in the percentages of publications between male and female corresponding authors (ie, *gender gap*) changed over time. Tests were considered to be significant with a 2-sided $P < .05$.

Results

The percentages of male and female corresponding authors and the gender gap in bioRxiv (46 101 articles) and medRxiv (5148 articles) over time are presented in **Figure 1**. We observed a statistically significant increase over time in the gender gap in medRxiv (Somers D, 0.14; 95% CI, 0.03 to 0.24) but not in bioRxiv (Somers D, 0.06; 95% CI, -0.01 to 0.12). During the pandemic, the gender gap in medRxiv increased from 23% in January 2020 to 55% in April 2020 and in bioRxiv changed from 46% in January 2020 to 47% in April 2020. We identified corresponding authors' institution country for 89% of articles, and the numbers of US- and non-US-based analyses were consistent (medRxiv US: $n = 1377$; Somers D, 0.04 [95% CI, -0.07 to 0.15]; medRxiv non-US: $n = 3804$; Somers D, 0.17

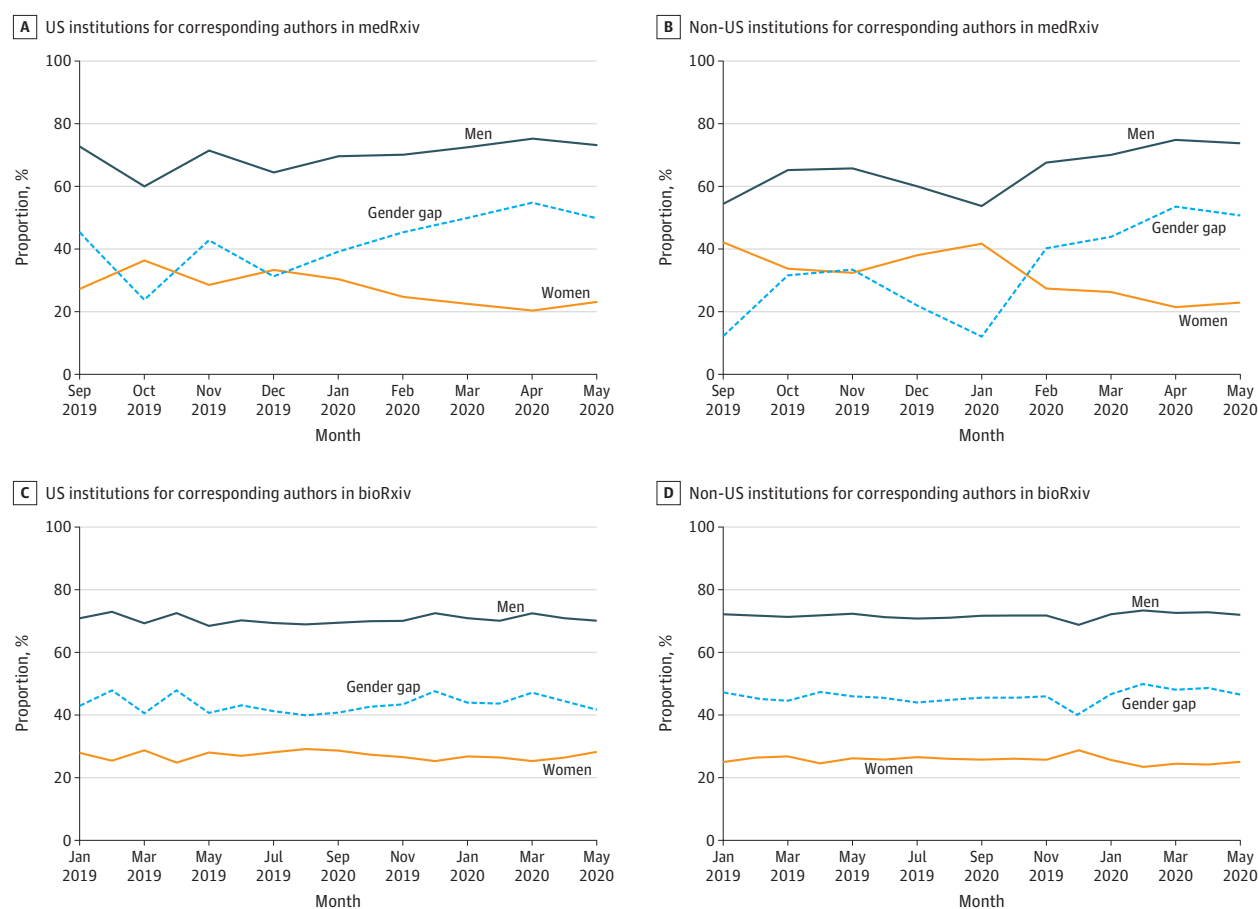
Open Access. This is an open access article distributed under the terms of the CC-BY License.

Figure 1. Gender of Corresponding Authors in medRxiv and bioRxiv Over Time



The absolute difference of the percentages of male and female corresponding authors are presented as the *gender gap*. Percentages do not always add to 100% because the gender of some corresponding authors was unknown. The time scale for medRxiv is shorter than that of bioRxiv owing to the more recent inception of medRxiv.

Figure 2. Gender of Corresponding Authors in medRxiv and bioRxiv Over Time Stratified by Institution Country



The absolute difference of the percentages of male and female corresponding authors are presented as the *gender gap*.

[95% CI, 0.01 to 0.23]; bioRxiv US: n = 17 996; Somers D, 0.01 [95% CI, -0.05 to 0.07]; bioRxiv non-US: n = 23 209; Somers D, 0.04 [95% CI, -0.02 to 0.10]) (Figure 2).

Discussion

The present analysis focused on corresponding authors to investigate the consequences of the COVID-19 pandemic for established investigators because this group may encounter unique challenges requiring specialized solutions (eg, tenure-clock extensions).

We observed a statistically significant increase in the corresponding author gender gap in medRxiv (23% in January 2020 to 55% in April 2020), which was unlikely to reflect seasonal variation alone. We did not observe a statistically significant change in the corresponding author gender gap in bioRxiv (46% in January 2020 to 47% in April 2020). The results were consistent in US- and non-US-based analyses. The difference in our findings between bioRxiv and medRxiv may reflect distinct populations of researchers contributing to each service.

Limitations of this study include the inference of gender based on first name and that findings among individuals who submit to preprint services may not be generalizable. Further research is needed to determine whether trends in preprint services translate to concrete metrics of academic achievement.

As the COVID-19 pandemic continues to progress, preprint archiving services may be a useful metric to evaluate real-time academic trends. If a clear gender gap develops, as it appears to be doing in health sciences research, academic institutions and funding agencies may need to address the COVID-19 pandemic-related gender disparities.

ARTICLE INFORMATION

Accepted for Publication: July 29, 2020.

Published: September 17, 2020. doi:10.1001/jamanetworkopen.2020.20335

Open Access: This is an open access article distributed under the terms of the [CC-BY License](#). © 2020 Wehner MR et al. *JAMA Network Open*.

Corresponding Author: Mackenzie R. Wehner, MD, MPhil, Department of Health Services Research, University of Texas MD Anderson Cancer Center, 1400 Pressler St, Unit 1444, Office FCT9.5036, Houston, TX 77030 (mwehner@mdanderson.org).

Author Affiliations: Department of Health Services Research, University of Texas MD Anderson Cancer Center (Wehner, Li); Department of Dermatology, University of Texas MD Anderson Cancer Center, Houston, Texas (Wehner); Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, Texas (Nead); Department of Radiation Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas (Nead).

Author Contributions: Drs Wehner and Nead had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Wehner, Nead.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: All authors.

Critical revision of the manuscript for important intellectual content: Wehner, Nead.

Statistical analysis: Li, Nead.

Administrative, technical, or material support: Wehner.

Supervision: Wehner, Nead.

Conflict of Interest Disclosures: None reported.

Funding/Support: This research was supported, in part, by grant CCSG P30 CA016672 from the National Institutes of Health. Drs Wehner and Nead are Cancer Prevention Research Institute of Texas Scholars in Cancer Research.

Role of the Funder/Sponsor: The National Institutes of Health had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Jolly S, Griffith KA, DeCastro R, Stewart A, Ubel P, Jagsi R. Gender differences in time spent on parenting and domestic responsibilities by high-achieving young physician-researchers. *Ann Intern Med*. 2014;160(5):344-353. doi:[10.7326/M13-0974](https://doi.org/10.7326/M13-0974)
2. Derrick GE, Jaeger A, Chen PY, Sugimoto CR, van Leeuwen T, Lariviere V. *Models of parenting and its effect on academic productivity: preliminary results from an international survey*. Pro Int Conf Sci Inf. 2019:1670-1676.
3. Callaway E. BioRxiv preprint server gets cash boost from Chan Zuckerberg Initiative. *Nature*. 2017;545(7652):18. doi:[10.1038/nature.2017.21894](https://doi.org/10.1038/nature.2017.21894)
4. Santamaria L, Mihaljevic H. Comparison and benchmark of name-to-gender inference services. *Peerj Comput Sci*. 2018;4:e156. doi:[10.7717/peerj-cs.156](https://doi.org/10.7717/peerj-cs.156)