The citation disadvantage of female-as-corresponding-author research teams in biology

Jinyu Gao[1][0009-0005-4744-9416] and Yi Bu[1,2][0000-0003-2549-4580]

[1] Center for Informationalization and Information Management, Peking University, Beijing 100871, China
[2] Department of Information Management, Peking University, Beijing 100871, China
buyi@pku.edu.cn

Abstract. First and corresponding authors are often regarded as leading members in a research team. Distinguishing the gender composition of leading authors paints a nuanced picture on the gender differences in scientific performance at the research team level. This poster selects scientific publications in 2010 in the field of biology and finds that, in mixed-gender teams, while female authors hold a higher rank trend in publications' bylines than male authors, female-as-corresponding-author research teams tend to receive less citations.

Keywords: Gender Inequality, Team Science, Citation, Bibliometrics.

1 Introduction

Gender differences have always existed in academia (Lerman et al., 2022; Ross et al., 2023; Vasarhelyi et al., 2021). This has been revealed by the extant findings that the total number of papers published by women in their careers is significantly lower than that of male authors and that the length of female scientists' academic career is lower than that of men (Huang et al., 2020). Meanwhile, the gender composition of teams is starting to get more attention. For instance, Shen et al. (2022) distinguished the gender compositions of collaboration pairs and suggested that inter-gender collaboration has a positive effect on academic performance. Yang et al. (2022) divided research teams into two groups, namely same- and mixed-gender teams, and found that the mixed-gender teams are more likely to create novel and impactful publications than same-gender teams. Bradley et al. (2021) explored the relationships between gender majority in a team and its performance. Teich et al. (2022) concluded that papers led by women are significantly under-cited and that those by men are significantly over-cited.

In gender-related studies, a couple of bibliometric indicators, such as citations and novelty, have been used to evaluate the scientific performance of a scientific publication that contains joint efforts of multiple authors. That being said, the gender differences in scientific performance should be explored from the team level. In many natural science disciplines such as biology, first and last authors are often regarded as leading authors that make the most significant contribution of a scientific publication. Considering the different contributions of team members, this poster focuses on team gender

composition of the first and the last authors and explores the relationships between the team gender composition and citations. This poster aims to investigate potential differences in the contribution patterns of male and female scientists to scientific publications via differences in author sequence and to assess the impact of the gender of teams' main contributors on the recognition of research outputs.

2 Data

The sample data was retrieved from SciSciNet (Lin et al., 2023), including 350,438 journal papers published in 2010 in biology. For each paper, SciSciNet calculated a range of indicators, containing the number of citations five and ten years after a paper was published (C5 and C10, respectively), as well as the disruption score (Funk & Owen-Smith, 2017; Wu et al., 2019); for each name disambiguated author, the SciSciNet applied statistical models (Van Buskirk, 2022) based on the author's name to infer the probability the author being a female. The probability over greater than 0.5 is considered female, while others are considered male in this poster.

3 Methods

Author Sequence. The author sequence of each publication's byline shows the rank order of its authors, i.e., the sequence number of the first author equals one, the sequence number of the second author equals two, and the sequence number of the last author equals, in numeral, the team size. Considering that team size is quite various, the author sequence number is divided by the team size to recalculate the *author sequence*, that is, the percentile value of each author's ranking in the article. In biology, the two most important contributors to an article are usually the first and the last authors (Kassis, 2017). The first author is most likely a junior scientific researcher who is responsible for the implementation and presentation of the article (Li et al., 2022), while the last author oftentimes plays the roles of supervisory and be corresponding author. Given the importance of corresponding authors (usually the last author) in a research article, the author sequence of the last author is particularly set to be equal to the first author.

Independent variables. (1) *Team Type (TT)*. Teams were classified into three categories based on the gender composition, namely all-male (128,678 papers), all-female (22,116 papers), and mixed-gender teams (199,644 papers). Mixed-gender teams were further divided into four groups according to the gender of the first and the last authors, namely, teams with both first and last authors female (20,500 papers), annotated as *ff*; teams with first author female and last author male (72,960 paper, *fm*); teams with first author male and last author female (36,318 papers, *mf*); and teams with both first and last authors male (69,866 papers, *mm*). We generated three dummy variables for the mixed-gender team type. (2) *Team Size (TS)*. This poster aggregated all teams equal to or larger than seven as seven-author teams, and generated six dummy variables for the team size.

Dependent variables. (1) Citation. Because citations vary over time, the citation windows were set to improve comparability (Wang, 2013). The number of citations a paper received within five years of publication (C5) is considered short-term citations, while ten-year citations (C10) are considered long-term citations. (2) Disruption. Whether the subsequent researches cited a focal paper's references reflect the disruptive or developing nature of the focal paper contributed to the existing scientific thoughts. As citation counts fail to illustrate this characteristic, we added another indicator, namely disruption score calculated through citation networks to quantify the extent to which a paper disrupts or develops existing works (Funk & Owen-Smith, 2017; Wu et al., 2019). The value ranges between -1 and 1. Considering its uneven distribution, we define *Disruption_positive* equals one if the value of disruption score over greater than zero and zero otherwise (Lin et al., 2022).

Control variables. (1) *Team age mean (TA)*, refers to the average career age of authors per paper. Research output can be affected by potential life-cycle effects (Jones & Weinberg, 2011). This study uses the number of years from the year 2010 to the scientist's first publication year recorded in the database to represent his/her career age. (2) *Team C5 Mean (TC)* refers to the average C5 of each author in the previous five years before the year 2010. The better the past scientific research performance, the more likely it is to have better scientific research output in the future.

Regressions. The empirical models based on ordinary least squares regressions and the logit regression are as below:

$$\text{Log}(C5+1)= \alpha_1 + \beta_1 TT + \delta_1 TS + \mu_1 TT*TS + \tau_1 TA + \Theta_1 TC + \epsilon_1 \quad (1)$$

$$\text{Log}(C10+1)= \alpha_2 + \beta_2 TT + \delta_2 TS + \mu_2 TT*TS + \tau_2 TA + \Theta_2 TC + \epsilon_2 \quad (2)$$

$$\text{Disruption\_positive}= \alpha_3 + \beta_3 TT + \delta_3 TS + \mu_3 TT*TS + \tau_3 TA + \Theta_3 TC + \epsilon_3 \quad (3)$$

4 Preliminary results

Fig.1 shows a higher rank trend of women in the mixed-gender teams' author sequence in the biology field. A null model was designed to estimate the expected distribution of female authors. Particularly, the null model randomly reshuffles the author sequence for each paper and held a constant number of female and male authors. This reflects gender discrepancies in author sequences. Thus, we next intend to explore the gender gap in a citation from the team level.

Table 1 presents the results of the regression analysis. VIF values all less than ten indicate no significant collinearity. Firstly, the regression results illustrate that the larger the team size, the more citations but the less disruptive, which echoes Wu et al. (2019). Secondly, there is a significant negative relationship between the number of citations and teams with male first authors and female last authors. Models (1) and (2) suggest that female-as-last-author teams are less likely to be cited. Thirdly, there is a significant positive relationship between the disruption and teams with males as both first and last authors. Model (3) indicates that teams with males as both first and last authors may publish more disruptive papers.

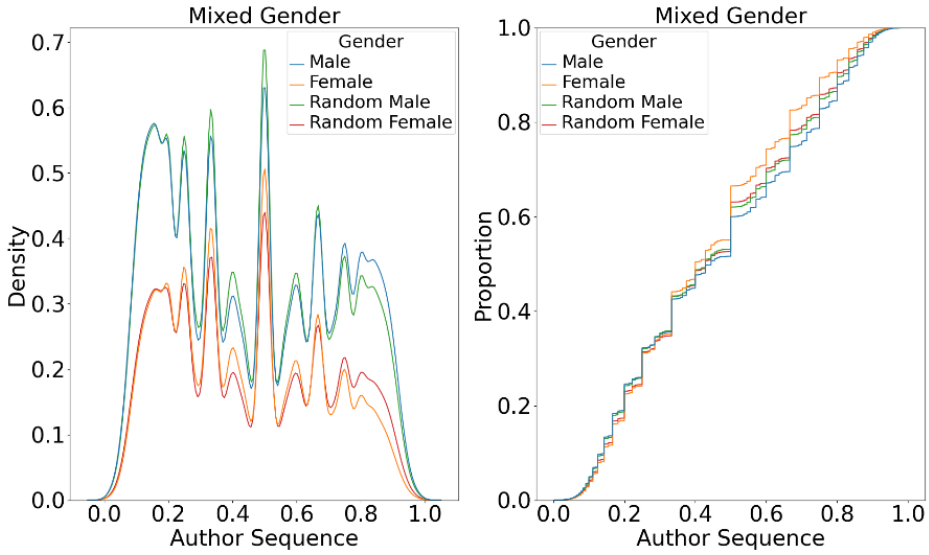

Fig1. The kernel density (left) and cumulative distribution (right) of author sequence.

Table 1. Regression results on citations and disruption.

	(1)	(2)	(3)
	Log(C5+1)	*Log(C10+1)*	*Disruption_positive*
Regression	OLS	OLS	Logit
Team Size			
Team Size=3	0.041***	0.047***	-0.173***
	(0.013)	(0.015)	(0.034)
Team Size=4	0.079***	0.092***	-0.183***
	(0.013)	(0.015)	(0.033)
Team Size=5	0.133***	0.140***	-0.236***
	(0.013)	(0.015)	(0.034)
Team Size=6	0.185***	0.191***	-0.215***
	(0.013)	(0.015)	(0.034)
Team Size>=7	0.469***	0.480***	-0.294***
	(0.013)	(0.014)	(0.031)
Team Type			
fm	0.014	0.011	0.014
	(0.010)	(0.011)	(0.027)
mf	-0.058***	-0.074***	0.046
	(0.011)	(0.012)	(0.030)
mm	0.006	-0.003	0.079***
	(0.010)	(0.011)	(0.027)
Team Age Mean	0.037***	0.040***	-0.013***

	(0.001)	(0.001)	(0.001)
Team C5 Mean	0.020***	0.020***	-0.013***
	(0.001)	(0.001)	(0.001)
_cons	1.208***	1.592***	-1.608***
	(0.018)	(0.019)	(0.041)
N	191612	191612	191612
R^2	0.184	0.164	
adj. R^2	0.184	0.164	
Mean VIF	2.47	2.47	

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We also add the interaction term between team size and team type to reevaluate the aforementioned relationship. From Table 2 and Fig. 2, we see that teams with first author male and last author female have the least marginal effect on the citation, revealing that the female corresponding author has a disadvantage in gaining recognition from peers. The gender disparities of corresponding authors may reduce the career opportunities in obtaining grants and recruiting students for female team leaders. This finding echoes previous studies on gender inequality that female scientists are less likely to be promoted (Huang et al., 2020) and are underrepresented among journal reviewers and editors (Berenbaum, 2019).

Table 2. Regression results involving interaction terms on citations and disruption.

	(4)	(5)	(6)
	Log(C5+1)	*Log(C10+1)*	*Disruption_positive*
Regression	OLS	OLS	Logit
Team Size			
Team Size=3	0.023	0.024	-0.237**
	(0.035)	(0.040)	(0.096)
Team Size=4	0.085***	0.088**	-0.197**
	(0.033)	(0.037)	(0.088)
Team Size=5	0.139***	0.143***	-0.208**
	(0.033)	(0.037)	(0.088)
Team Size=6	0.190***	0.190***	-0.185**
	(0.033)	(0.037)	(0.090)
Team Size>=7	0.454***	0.459***	-0.393***
	(0.022)	(0.024)	(0.054)
Team Type			
fm	-0.002	-0.016	-0.076
	(0.029)	(0.032)	(0.076)
mf	-0.045**	-0.059***	0.093*
	(0.019)	(0.021)	(0.054)
mm	0.023	0.016	0.154***
	(0.015)	(0.017)	(0.045)

*TS*TT*			
*3*fm*	0.024	0.034	0.132
	(0.040)	(0.044)	(0.106)
*3*mf*	-0.000	-0.010	-0.045
	(0.035)	(0.039)	(0.097)
*3*mm*	0.010	0.013	-0.036
	(0.033)	(0.037)	(0.091)
*4*fm*	0.008	0.025	0.074
	(0.037)	(0.042)	(0.099)
*4*mf*	-0.023	-0.020	-0.064
	(0.032)	(0.036)	(0.089)
*4*mm*	-0.038	-0.036	-0.114
	(0.028)	(0.032)	(0.079)
*5*fm*	0.014	0.024	0.013
	(0.037)	(0.042)	(0.101)
*5*mf*	-0.030	-0.029	-0.058
	(0.032)	(0.037)	(0.091)
*5*mm*	-0.040	-0.047	-0.170**
	(0.028)	(0.031)	(0.079)
*6*fm*	0.033	0.049	0.051
	(0.038)	(0.043)	(0.103)
*6*mf*	-0.053	-0.065*	-0.244**
	(0.034)	(0.039)	(0.096)
*6*mm*	-0.041	-0.042	-0.132*
	(0.029)	(0.032)	(0.080)
*7*fm*	0.017	0.024	0.164**
	(0.027)	(0.030)	(0.068)
*7*mf*	0.000	0.000	0.000
	(.)	(.)	(.)
*7*mm*	0.000	0.000	0.000
	(.)	(.)	(.)
Team Age Mean	0.037***	0.040***	-0.013***
	(0.001)	(0.001)	(0.001)
Team C5 Mean	0.020***	0.020***	-0.013***
	(0.001)	(0.001)	(0.001)
_cons	1.213***	1.604***	-1.572***
	(0.028)	(0.031)	(0.070)
N	191612	191612	191612
R^2	0.184	0.165	
adj. R^2	0.184	0.164	
Mean VIF	8.87	8.87	

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

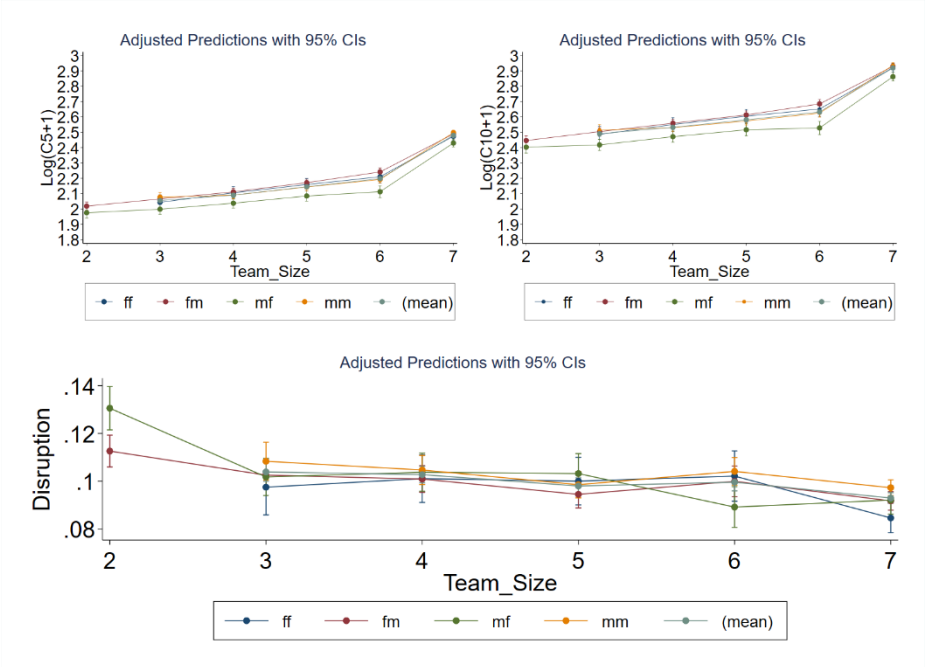

Fig2. The marginal effect of team size and team type on citations (top left and right) and disruption (bottom).

5 Summary

This poster explores whether male and female scientists contribute differently to publications in biology and finds that female scientists in mixed-gender teams tend to have a higher author sequence. We divide gender-mixed teams into four groups based on the gender of the first and the last authors and analyze the impact of different team types and team sizes on citations and disruption through regression models. We observe that female-as-corresponding-author teams are less likely to be recognized in terms of citation counts and disruption. This research contributes to a deeper understanding of gender dynamics in scientific collaborations by offering empirical support for gender inequity and fairness. Findings provide valuable insights for research team management and assessment and highlight potential influence of gender on academic recognition.

Yet, the current poster has several limitations. Only one year's data from one discipline is selected, and chronical data analyses can be included in the subsequent research to form a more comprehensive and rich understanding of the relationship between team gender composition and citations. For example, we will explore the influence of team network structure, knowledge base, and collaboration on the relations above mentioned. For the classification of team gender composition, this poster considers the gender of only two leading authors, and the gender of other team members can be added to the classification criteria in the future.

Acknowledgements

The authors acknowledge the financial support by the National Natural Science Foundation of China (No. 72104007 and No. 72174016) and by the Fundamental Research Funds for the Central Universities, Peking University. The authors thank Hongkan Chen, Yifan Tian and other members of the Knowledge Discovery lab at the Department of Information Management, Peking University for helpful discussions and suggestions.

References

1. Berenbaum, M. R.: Speaking of gender bias. *Proceedings of the National Academy of Sciences of the United States of America, 116*(17), 8086-8088 (2019)
2. Bradley, B., Henry, S., & Blake, B.: When Can Negativity Mean Success? Gender Composition, Negative Relationships and Team Performance. *Small Group Research, 52*(4), 457-480 (2021)
3. Funk, R. J., & Owen-Smith, J.: A Dynamic Network Measure of Technological Change. *Management Science, 63*(3), 791-817 (2017)
4. Huang, J. M., Gates, A. J., Sinatra, R., & Barabasi, A. L.: Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences of the United States of America, 117*(9), 4609-4616 (2020)
5. Jones, B. F., & Weinberg, B. A.: Age dynamics in scientific creativity. *Proceedings of the National Academy of Sciences of the United States of America, 108*(47), 18910-18914 (2011)
6. Kassis, T.: How do research faculty in the biosciences evaluate paper authorship criteria? *Plos One, 12*(8), e0183632 (2017)
7. Lerman, K., Yu, Y., Morstatter, F., & Pujara, J.: Gendered citation patterns among the scientific elite. *Proceedings of the National Academy of Sciences of the United States of America, 119*(40), e2206070119 (2022)
8. Li, W. H., Zhang, S., Zheng, Z. M., Cranmer, S. J., & Clauset, A.: Untangling the network effects of productivity and prominence among scientists. *Nature Communications, 13*(1), 11 (2022)
9. Lin, Y. L., Evans, J. A., & Wu, L. F.: New directions in science emerge from disconnection and discord. *Journal of Informetrics, 16*(1) (2022)
10. Lin, Z. H., Yin, Y., Liu, L., & Wang, D. S.: SciSciNet: A large-scale open data lake for the science of science research. *Scientific Data, 10*(1), 22 (2023)
11. Ross, M. B., Glennon, B. M., Murciano-Goroff, R., Berkes, E. G., Weinberg, B. A., & Lane, J. I.: Author Correction: Women are credited less in science than men. *Nature* (2023)
12. Shen, H. Q., Cheng, Y., Ju, X. F., & Xie, J.: Rethinking the effect of inter-gender collaboration on research performance for scholars. *Journal of Informetrics, 16*(4), 101352 (2022)
13. Teich, E. G., Kim, J. Z., Lynn, C. W., Simon, S. C., Klishin, A. A., Szymula, K. P., Srivastava, P., Bassett, L. C., Zurn, P., Dworkin, J. D., & Bassett, D. S.: Citation inequity and gendered citation practices in contemporary physics. *Nature Physics, 18*(10), 1161-1170 (2022)
14. Van Buskirk, I., Clauset, A. & Larremore, D. B. An Open-Source Cultural Consensus Approach to Name-Based Gender Classifcation. *arXiv preprint, arXiv:2208.01714* (2022)

15. Vasarhelyi, O., Zakhlebin, I., Milojevic, S., & Horvat, E.-A.: Gender inequities in the online dissemination of scholars' work. *Proceedings of the National Academy of Sciences of the United States of America, 118*(39), e2102945118 (2021)
16. Wang, J.: Citation time window choice for research impact evaluation. *Scientometrics, 94*(3), 851-872 (2013)
17. Wu, L. F., Wang, D. S., & Evans, J. A.: Large teams develop and small teams disrupt science and technology. *Nature, 566*(7744), 378-382 (2019)
18. Yang, Y., Tian, T. Y., Woodruff, T. K., Jones, B. F., & Uzzi, B.: Gender-diverse teams produce more novel and higher-impact scientific ideas. *Proceedings of the National Academy of Sciences of the United States of America, 119*(36), e2200841119 (2022)