**Faulty evidence, specious logic, dangerous conclusions: Challenging gendered ideas of
mentorship**

Robert G. Alexander¹

Katherine E. Alexander²

Jena Barchas-Lichtenstein³

Stacey R. Finkelstein⁴

Seungyeon Lee⁵

Michelle L. Litchman⁶

Claudio R. Nigg⁷

Kathryn M. Ross⁸

Lisa R. Starr⁹

Xiaomeng Xu^{10*}

*Corresponding author, xuxiao@isu.edu

¹SUNY Downstate Health Sciences University

²The College of Mount Saint Vincent

³Knology

⁴Stony Brook University

⁵University of Arkansas at Monticello

⁶University of Utah

⁷University of Bern

⁸University of Florida

⁹University of Rochester

¹⁰Idaho State University

In their *Nature Communications* paper, AlShebli et al.¹ showed that women scientists face under-citation relative to men, and that this effect was compounded when multiple women co-authored manuscripts together. None of this is surprising; it is consistent with a wide body of research documenting pernicious systematic and structural gender biases. What is shocking is their conclusion that *women students should avoid female mentors and vice versa* so they may benefit, by proxy, from male privilege. They further suggest policy makers revisit the consequences of diversity policies and take this conclusion into account. The corrosive effects of this advice would only serve to exacerbate gender differences in science. Worse still, the study is so replete with flaws that their conclusions are as unmerited as they are offensive. Here, we discuss several of these significant flaws.

Lack of Causal Evidence. Policy decisions must be based primarily on causal evidence. However, the paper does not establish any causal effect related to same-gender mentoring. Regardless of analysis methods, results from a single research study using observational data collected at one time point cannot be used to establish causality.² During peer-review, reviewers appropriately requested that authors tone down causal claims, and specifically asked authors to indicate that the coarsened exact matching technique used in the paper results “cannot be claimed to establish causality”. In their responses to reviewers, the authors noted that their “gender analysis is only meant to be exploratory”; however, in the final publication the authors continued to infer causality from their observational data, over-extending results to support proposed policy implications. Policy recommendations should not be made based on initial exploratory analyses. Worse still, these initial exploratory analyses are flawed and ignore known gender biases in academia, especially academic publishing.³

Outdated Operationalization and Measurement of Gender. The authors' binary definition of gender is harmful and non-inclusive, and the algorithmic gender inference software used (Genderize.io) produces inaccuracies as it reduces gender to a binary construct. Gender is a complex social, interactional, and relational property that encompasses both identity and ascription by others, which intersects in complex ways with other social identities such as race, class, nationality, and disability status. Names are similarly related to gender in a complex way: names and naming practices often do not map cleanly onto either gender identity or gender presentation. Indeed, the authors pointed out that Genderize.io failed to classify almost half of all names tested (see also¹⁴). This software likely misgendered many people in the dataset, especially those with names outside the culture of the algorithm's developers.¹⁵ Additionally, the authors state they "analyze 215 million scientists" but supplemental materials show they excluded 125.66 million scientists (58.44%) for not meeting their gender classification criteria. The authors' gender-based analyses are clearly not representative of the population.

Flawed Operationalization of "Mentorship". A fundamental methodological flaw was the mentorship definition: scientists were considered "mentors" to junior scientists whenever they shared discipline, US-based affiliation, and co-authored a paper with 20 or fewer co-authors. AlShebli et al. justified this narrow definition in peer-review as falling under broader definitions of mentorship as the activity of giving help or advice to junior collaborators. That is, co-authorship implies support was given, and therefore co-authorship is mentorship. This was backed by an unrepresentative survey (see below) where fewer than 60% of respondents indicated they received advice from co-authors (Figures 1 and S1 suggest that 20-25% of respondents disagreed or strongly disagreed that they had received advice).

Flawed Operationalization of Protege “Success.” Other central flaws in this paper are the authors’ conceptualization of *quality* mentorship, defined in terms of “big shot” and “hub” experiences, and their definition of protege “success” as the number of citations received for published manuscripts within the first 5 years after their publication. These definitions assume that publications and citations are the sole metric of scholarly success and that they directly correspond to quality mentorship. This operationalization of success as citations is methodologically questionable given self-citations were not excluded and research has demonstrated that men are more likely to self-cite.⁴ Moreover, papers published by women are under-cited due to systemic issues such as implicit bias, discrimination, and structural inequality, not because women are unproductive.⁵ Women introduce more novel links between concepts than men, indicating more innovative work, yet these contributions are discounted and less likely to lead to faculty positions and continued careers.⁶ Third, “quality mentorship” cannot be boiled down to being highly-cited and a “hub”. For example, quality mentorship may help retain talented scientists. This is particularly true for same-gender mentors, who improve the likelihood of women scientists staying in the field.⁷⁻⁸ In contrast, harmful mentorship can produce highly-cited publications—including from trainees facing excessive workloads or sexual harassment from mentors—that might drive scholars out of their field.

There are multiple reasons for co-authorship that do not involve mentoring, and co-authorship norms vary dramatically across disciplines. Operationalizing mentorship as co-authorship also does not represent the academic mentoring literature which recommends that mentor-protégé relationships include regular meetings, realistic expectations, agreement on goals and methods for achieving these goals, mutual respect, and maintaining reciprocity such that both members benefit from the relationship.⁹⁻¹³ Mentorship is a process, the outcomes of which

are certainly not limited to citations. A more comprehensive and accurate mentorship definition is required for research to carry any policy significance regarding mentoring decisions.

Use of an Unrepresentative Sample. Of the 3 million mentor-protégé pairs in the dataset, 167 participants completed surveys. These *self-selected* responses were misrepresented in Figure 1's caption as responses of "randomly-chosen scientists". This subsample survey was not representative, since (1) all 2000 invited scientists were selected because they were junior scientists in 2014, although the other analyses spanned decades and did not assess whether the patterns changed over time; and (2) <10% of invited scientists actually responded. The demographics of respondents were not reported, thus gender differences in participation cannot be evaluated. Relatedly, analyses focused on 10 disciplines but the authors do not address gender inequities in these disciplines, nor discipline differences in gender distributions such as the prevalence of women mentors and mentees (e.g., nursing vs engineering). Citation norms across fields and institution types were also not addressed. Vast differences in citation counts exist across fields as a result of differences in publication practices. These factors introduce substantial confounds that render conclusions non-interpretable.

We strongly recommend that diversity policies continue to advocate against all forms of discrimination. AlShebli et al.'s conclusions begin with a deficit model, positioning gender itself as the cause of under-citation, rather than considering the role of structural and environmental factors. Therefore, while AlShebli et al.'s results demonstrate some effects of systematic gender bias, the solution they proposed would result in further biases against women scientists.

Scientists should NOT accept the suggestions that:

- 1) working with women mentors is harmful,

2) women should focus on mentoring men, and

3) policies should be based around limiting women mentors or prioritizing men as mentors.

Decreasing diversity in science slows down progress and harms us all. Instead of relying on flawed exploratory analyses and overgeneralized conclusions, diversity policy decisions must continue to be based on the best evidence available.

Acknowledgements

We extend our gratitude to the women mentors and mentees who make our fields, our lives, and our work better. We also extend appreciation to everyone who has joined this discussion across private and public platforms, offering solidarity, perspectives, and important lived experiences.

Author Contributions

All authors contributed to writing and editing the paper.

Competing Interests

The authors declare no competing interests.

References

1. AlShebli, B., Makovi, K. & Rahwan, T. *Nat Commun* **11**, 5855 (2020).
2. Hill, A. B. The environment and disease: association or causation. *J R Soc Med* **58**, 295-300 (1965).
3. Lundine, J., Bourgeault, I. L., Clark, J., Heidari, S., & Balabanova, D. The gendered system of academic publishing. *The Lancet* **391**(10132), 1754-1756 (2018).
4. King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J., & West, J. D. Men set their own cites high: Gender and self-citation across fields and over time. *Socius* **3**, 1-22 (2017).
5. Dworkin, J. D., Linn, K. A., Teich, E. G., Zurn, P., Shinohara, R. T., & Bassett, D. S. The extent and drivers of gender imbalance in neuroscience reference lists. *Nat Neurosci* **23**, 918–926 (2020).
6. Hofstra, B., Kulkarni, V. V., Galvez, S. M. N., He, B., Jurafsky, D., & McFarland, D. A. The diversity–innovation paradox in science. *Proc Natl Acad Sci* **117**(17), 9284-9291 (2020).
7. Dennehy, T. C., & Dasgupta, N. Female peer mentors early in college increase women’s positive academic experiences and retention in engineering. *Proc Natl Acad Sci* **114**(23), 5964-5969 (2017).
8. Hernandez, P. R. et al. Role modeling is a viable retention strategy for undergraduate women in the geosciences. *Geosphere* **14**(6), 2585-2593 (2018).
9. Binkley, P. F. & Brod, H. C. Mentorship in an academic medical center. *Am J Med* **126**, 1022–1025 (2013).
10. Kashiwagi, D. T., Varkey, P. & Cook, D. A. Mentoring programs for physicians in academic medicine: A systematic review. *Acad Med* **88**, 1029–1037 (2013).

11. Rustgi, A. K. & Hecht, G .A. Mentorship in academic medicine. *Gastroenterology* **141**, 789–792 (2011).
12. Straus S. E. et al. Characteristics of successful and failed mentoring relationships: A qualitative study across two academic health centers. *Acad Med* **88**, 82–89 (2013).
13. Xu, X., Schneider, M., DeSorbo-Quinn, A. L., King, A., Allegrante, J. P., & Nigg, C. R. Successful distance mentoring of health researchers: Three case studies from across the career development trajectory. *Health Psychol Open* **4**(2), 1-6 (2017).
14. Wais, K. Gender prediction methods based on first names with genderizer. *R Journal* **8**, 17-37 (2016).
15. Santamaría, L., & Mihaljević, H. Comparison and benchmark of name-to-gender inference services. *PeerJ Comput Sci* **4**, e156 (2018).