Critique of The association between early career informal mentorship in academic collaborations and junior author performance

Biwei Zheng

22/12/2020

Keywords: Statistics, Survey, Sampling, Gender, Woman, Mentorship, Performance

Abstract

This paper is a critique of the following retracted article: The association between early career informal mentorship in academic collaborations and junior author performance, by AlShebli, Makovi, and Rahwan. The article was first published on nature on 17 November 2020, and retracted on 21 December 2020 'in response to criticisms about the assumptions underpinning the article in terms of the identification of mentorship relationship and the measure of mentorhship quality, challenging the interpretation of the conclusions'. I started by summarizing the article and interpreting the result. Then I raised various challenges towards its assumption, methodology, data unbiasedness, and interpretation of the result. After that, I cited other critiques on the article such as 'Jessica Hullman, Are female scientists worse mentors? This study pretends to know'. In the end, I provided with my own discussion and proposed the next steps to find effects of mentor-protege relationships in academia.

Introduction

Mentorship is the glue that connects academics. It is crucial in helping junior scientists to grow and learn under experienced senior members, and it's also exchange that's beneficial to the mentors. In the process of debating with mentees, mentors can be stimulated or provoked with perspectives that have never been thought of. Whether it's formal or informal, mentorship is the linkage to build connections within a field. It can also ease the barrier to entry for novice scientists, set up role models for minorities, people of color, and women.

BUT, are female scientists perform worse than male peers in a mentorship? Three academics from NYU Adu Dhabi studied 3 million mentor-portege pairs and conducted analysis on the impact of the mentor-protege relationship to the protege and mentor. Their interpretation of result suggests that, and I quote, '... having more female mentors is associated with a decrease in the mentorship outcome' and 'by mentoring female instead of male proteges, the female mentors compromise their gain from mentorship".

This paper is intended to do a indepth examination of the creation process of such result, raise challenges to their methodology and interpretations, and provide with my suggestions on this matter. It starts by summerizing the article, then raised various challenges about the assumptions, data source and selection process. Then I laid out critiques on their interpretation of results, especially on the gender inequality part. In the end,

Summary of the article

The article 'The association between early career informal mentorship in academic collaborations and junior author performance' studies informal mentor-protege relationship and their outcome to both proteges and mentors. They define informal mentorship as a published paper co-authored by someone with less than 7 years of experience and someone with over 7 years of experience in the field, without holding a formal

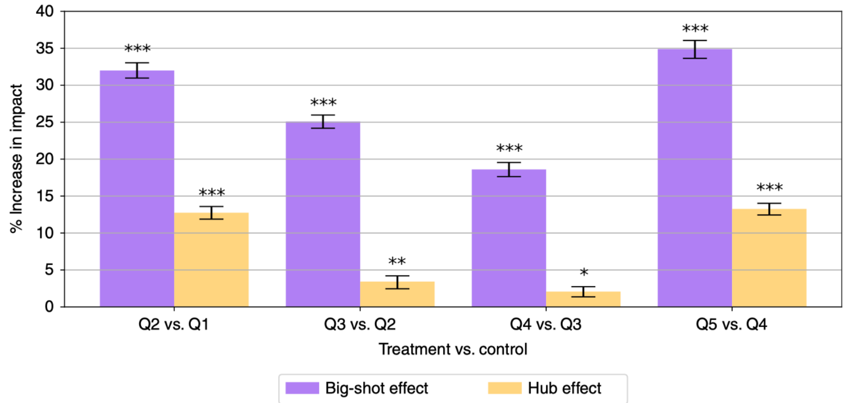

Fig. 2 The big-shot experience and hub experience of 3 million mentor–protégé pairs. For every independent variable, be it big-shot experience or hub experience, Q_i denotes the ith quintile of the distribution of that variable. For $i \in \{1, 2, 3, 4\}$, we consider Q_{i+1} and Q_i to be the treatment and control groups, respectively, and write Q_{i+1} vs. Q_i when referring to the CEM used to compare these two groups. The color of the bar indicates whether the independent variable is the big-shot experience (purple) or the hub experience (yellow), whereas the height of the bar equals δ, which is the increase in the average post-mentorship impact of the treatment group relative to that of the control group. A t-test shows that the values of δ are all statistically significant; see the corresponding p-values in Supplementary Tables 2 and 3. Since scientific impact is sensitive to external values, we bootstrap a 95% confidence interval. The error bars represent the 95% confidence interval. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$. Source data are provided as a Source Data file.

4 NATURE COMMUNICATIONS | (2020)11:5855 | https://doi.org/10.1038/s41467-020-19723-8 | www.nature.com/naturecommunications

Figure 1: Screen Shot 2020-12-22 at 5.53.56 PM.png

supervisory role. First, they conducted a survey asking protege how they feel about the support given by their informal mentors. As expected, the majority of proteges gave positive responses to the mentors' help in areas like research studying and social networking.

Then they use the Microsoft Academic Graph (MAG) dataset to analyze 215 million scientists and 222 million papers, covering ten major disciplines and over a century of research. Their results consist of two parts: the impact of 'big-shot' and 'hub' of their mentors to proteges after the mentorship, and the relationship between gender and gain from mentorship. The 'big-shot' effect is defined as the impact of mentor prior to the mentorship, measured by their average number of citations per annum up to the year of their first publication with the protege. The 'hub' effect is the network of scientific collaborations of the mentors up to the year of their first publication with the protege. The measure the post-mentorship result of proteges by selecting the papers they publish when their academic age exceeds 7 years and the author includes the protege but not their mentors, and measure the papers' citation number within 5 years of publication. For each of their independent variables (big-shot and hub), they group them into 5 quantiles, and compare each quantile's impact with the group one quantile lower, as the image below shows. These findings imply that the scientific impact (big-shot) of the mentors matters more than their connection network in the field when measuring the post-mentorship effect of proteges.

The second part of their result is to analyze the relationship between gender and mentorship gain for both proteges and mentors. They first measured female mentor's effect on both male and female proteges, by assigning Fi (F0, F1,. . . , F5) as the set of proteges with i number of female mentors, compare F1 to F5 with their baseline F0 (no female mentors), then generate heatmaps to show female mentors' existence have a negative effect on protege's post-mentorship outcome. It further shows that female mentor and female protege generates up to -30% of protege's impact compared to F0, the set of female protege with no female mentor at all, while female mentor and male protege generates a less worse outcome but still worse than male mentors. Other than the gain of proteges, the article further analyzed the gain of mentors after the mentorship. It shows that both female and male mentors have less gain with female proteges compared to male proteges. Female mentors have -9% gain mentoring female proteges compared to male proteges, while male mentors have -3%. They concluded that 'having more female mentors is associated with a decrease in the mentorship outcome' and 'by mentoring female instead of male proteges, the female mentors compromise their gain from mentorship'.

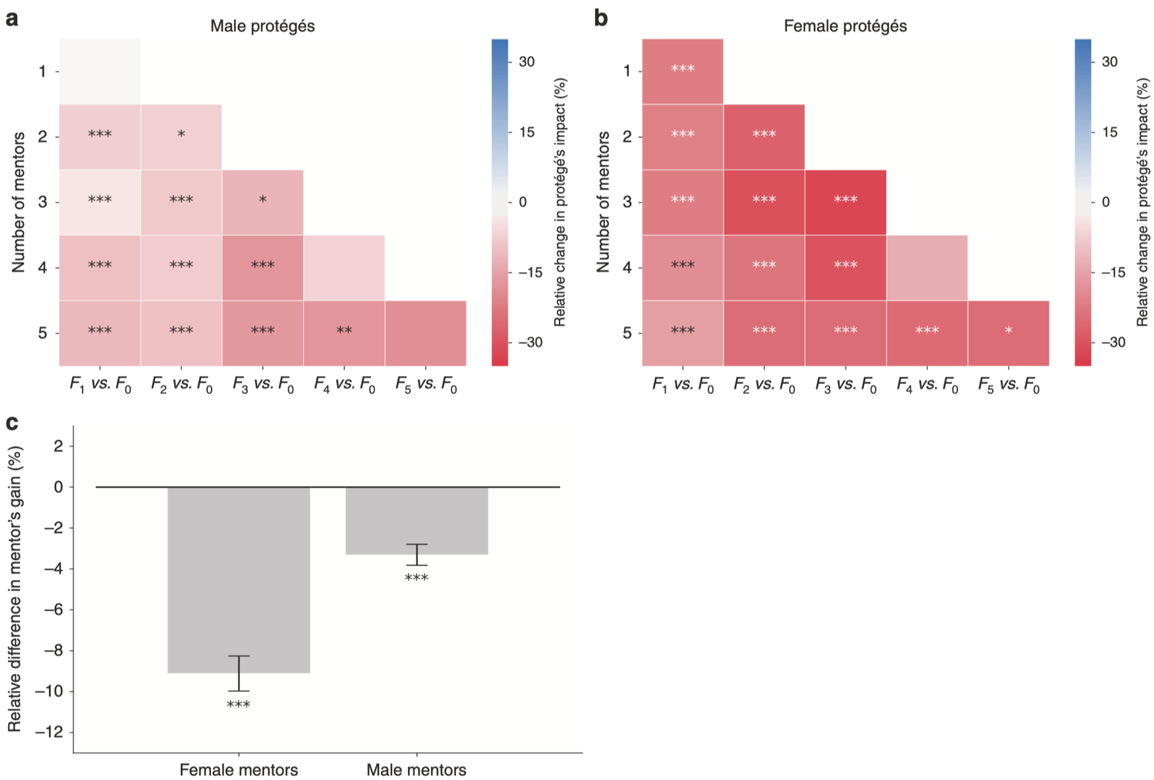

Fig. 3 The relationship between gender and the gain from mentorship. a F_i denotes the set of protégés from our 3 million pairs that have exactly i female mentors. Focusing on male protégés, F_i vs. F_0: $i = 1, \ldots, 5$ refers to the change in the average post-mentorship impact of protégés in F_i relative to the average post-mentorship impact of those in F_0 while controlling for the protégé's big-shot experience, number of mentors, discipline, affiliation rank, and the year in which they published their first mentored paper. A t-test is used to show the that values are all satistically significant; see the corresponding p-values in Supplementary Table 24. **b** The same as **a** but for female protégés instead of male protégés. **c** The gain of a mentor when mentoring a particular protégé is measured as the average impact ($\langle c_5 \rangle$) of the papers they authored with that protégé during the mentorship period. While controlling for the protégé's discipline, affiliation rank, number of mentors, and the year in which they published their first mentored paper, the figure depicts the change in the mentor's average gain when mentoring a female protégé relative to that when mentoring a male protégé; results are presented for female mentors and male mentors separately. A t-test shows that the values are all statistically significant. Since scientific impact is sensitive to external values, we bootstrap a 95% confidence interval. The error bars represent the 95% confidence interval. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Source data are provided as a Source Data file.

In the discussion session, the authors listed a few possible reasons for the under-performing female mentors: female mentors serve on more committees so reduce the time invested in their proteges, and women take on less recognized topics that their proteges emulate. As for mentors benefit more from male proteges, authors suggest that it could be due to 'male scientists had enjoyed more privileges' or it could be biaed in the sorting mechanisms.

Assumption, data source, and methods

The data used for the study are from all the papers in Microsoft Academic Graph (MAG) dataset up to December 31st, 2019. However, when measuring the post-mentorship result for proteges, they used 'number of citation within 5 years of publication', which can't be calculated for paper post-2015. Either they only include papers up until 2015, or all the post-2015 papers are been left out for less time to cite, making it unfair to compare with all the pre-2015 papers.

The duration of mentorship is also confusing to me. They consider 'the start of mentorship period to be the year of the first publication of the protege', but they also consider the start of a scientist's academic experience as the year of their first publication, which means that the mentorship starts automatically when the scientist enters the first year of academic age. They define the end of the mentorship period as when the junior became a senior scientist, which is exactly 7 years later. The fixture of mentorship period at 7 years makes no sense to me. Why would they assume the relationship continued for 7 years, for millions of mentor-protege pairs they analyzed. I understand that they need to make some assumptions for the data to make it analyzable, but this fixed 7-year duration regardless of the year protege and mentor published their co-authored paper should be improved to increase the accuracy of the results.

For gender identification, they used an API called genderized.io to identify whether a name is male or female. But only in the supplementary note, they mentioned that it has a 47.7% of unclassified rate, while only half the names are identified, with a 7% error rate. For the unidentified half, I believe they include a lot of non-traditional western names, like names of the minorities and people of color. If you put in 'James Bond', sure it will identify as a male, no problem. If you put in my name 'Biwei Zheng', it will most likely to be unclassifiable because 'Biwei Zheng' is the roman script of my Chinese name. Of course it won't identify my name because average readers like you can't tell I'm male or female either. This is just an example of minority scientist's names been left out by their system, and the study based on a data set excluding half of data due to unable to identify gender should at least mention the possible left out and bias in the article.

Interpreting the result

They sampled 2000 scientists out of their millions of mentor-protege pairs, without specifying their sampling method. I'd assume there is some bias within the selection process. For example, if they only select scientists with personal a webpage so they can get their email address, then these scientists with personal pages are most likely be the successful ones who remained in academia. Those who remain unknown or ceased pursuing academics often don't have a active personal page with contact info. Thus it is possible that the survey results are biased towards proteges with successful outcomes, so the response of large percentage of positive effect of mentorship from the selected proteges are far from representing the population. In addition, out of the 2000 surveys they sent out, only 167 responsded, an 8.35% response rate. This is a relatively low response rate and triggers me to question the credibility of the survey again. But lets put aside all the questioning and look at their conclusion on this: 'Altogether, these findings indicate that the relationship between our identified proteges and mentors indeed involved some form of mentorship'. This conclusion makes me question the necessity of existence for the survey, as it does not help to conclude any meaningful results nor contributes to analyzing the data set.

The first part of their result shows that the mentorship quality (measured by big-shot effect or hub effect) is positively associated with the post-mentorship outcome of proteges. Their CEM mechanism compares the treatment group of a higher level mentorship quality to a control group of a lower level mentorship quality, and resemble each other in terms of characteristics of proteges. I agree with their methods here to prevent post-mentorship outcomes affected by proteges' characteristics, but I still have some questions. In my point of view, the big-shot effect is inherently an independent variable of the hub effect. If the mentor is prestigious, he/she will most likely have a lot of connections in the field. But the authors seem not to recognize that and list both big-shot effect and hub effect as independent variables of the post-mentorship outcome of proteges, which could lead to biases of their result.

Most importantly, their interpretation of the relationship between gender and mentorship is not only statistically invalid but also an insult to women scientists around the world. They used F0 (set with no female mentors) as a baseline and compare F1 - F5 (set with 1 - 5 female mentors in the group) against it, making the heatmap as shown in the summary of the article. They thus claim that 'having more female mentors is associated with a decrease in the mentorship outcome'. This is absurd! For one, gender is known to have a robust effect on citations. So defining mentorship outcome based on citation number is inherently biased against female scientists. Other than that, the source table in their supplementary for the heatmaps does not suggest more female mentors lead to worse result for proteges, even though it indicates an overall negative protege impact compared to all-male mentor groups. As we recognize the inequality of male and female scientists in academia, we should compare the mentorship outcome with sets of similar levels of male and female mentors. this would've been fairer to female mentors as they are competing against similar-level male mentors, and would make the result more meaningful. Other than that, if they are comparing F1 - F5

Gender of protégé	No. of mentors		$imp(T')$	$imp(C')$	$\delta(\%)$	p
Male	1	F_1 vs. F_0	8.14	7.96	-2.2	0.06
	2	F_1 vs. F_0	10.0	9.29	-7.1	1.8e-13
	2	F_2 vs. F_0	9.7	9.0	-6.8	0.01
	3	F_1 vs. F_0	10.6	10.2	-3.3	0.0002
	3	F_2 vs. F_0	10.6	9.7	-8.1	1.8e-06
	3	F_3 vs. F_0	10.4	9.2	-11.7	0.01
	4	F_1 vs. F_0	11.4	10.3	-9.8	2.9e-17
	4	F_2 vs. F_0	11.3	10.4	-7.8	2.7e-05
	4	F_3 vs. F_0	11.6	9.6	-17.1	7.6e-08
	4	F_4 vs. F_0	10.6	9.9	-6.7	0.4
	5	F_1 vs. F_0	12.0	10.6	-10.9	1.3e-11
	5	F_2 vs. F_0	12.0	10.8	-9.6	2.9e-06
	5	F_3 vs. F_0	12.1	10.1	-16.4	4.2e-09
	5	F_4 vs. F_0	10.9	9.0	-17.5	0.006
	5	F_5 vs. F_0	7.3	6.0	-18.2	0.3
Female	1	F_1 vs. F_0	9.8	7.7	-21.3	2.7e-66
	2	F_1 vs. F_0	11.8	9.3	-21.0	5.5e-73
	2	F_2 vs. F_0	11.2	8.2	-26.6	5.1e-32
	3	F_1 vs. F_0	13.0	10.2	-21.3	2.9e-45
	3	F_2 vs. F_0	12.8	8.9	-30.6	4.1e-48
	3	F_3 vs. F_0	11.4	7.7	-32.0	3.9e-16
	4	F_1 vs. F_0	12.9	10.4	-18.6	5.1e-33
	4	F_2 vs. F_0	12.7	9.8	-22.7	1.0e-30
	4	F_3 vs. F_0	11.8	8.3	-29.6	2.3e-21
	4	F_4 vs. F_0	10.3	9.0	-12.9	0.09
	5	F_1 vs. F_0	13.0	10.9	-15.5	1.8e-16
	5	F_2 vs. F_0	13.0	9.8	-24.6	2.4e-31
	5	F_3 vs. F_0	12.5	9.4	-24.7	7.4e-16
	5	F_4 vs. F_0	10.8	8.1	-24.9	1.4e-07
	5	F_5 vs. F_0	11.6	8.7	-24.6	0.03

Figure 2: Screen Shot 2020-12-22 at 10.23.08 PM.png

against F0, they should also compare M1 - M5 against M0, which compares groups with 1 - 5 male mentors against groups with zero male mentors.

Next step

In conclusion, the article is biased in its gender identification, survey sampling, and interpretation of results regarding female scientists in the mentorship. However, there are a few steps I'd suggest.

One crucial step is to improve the survey to proteges. We can start by doing a proper sampling within the data set, like stratified sampling, to select a larger sample size (like 50,000) for the population size of millions of mentor-protege pairs. Then design a better survey to quantify the results. For example, we can ask questions like 'From 1 to 100, what do you think your level is pre-mentorship and post-mentorship in writing, research study, data analyzing/modeling …..'. We can also ask how them to rate from 1 to 100 of their mentor's assistance in academics, career planning, networking, etc. After that, we should send surveys to the respective mentors, asking them how they feel about their protege's ability pre and post mentorship, how much they think they helped the protege, and cross-reference the results from proteges and mentors. In addition, we can also ask how much the mentors gained themselves in the relationship. After gathering all the self-reported information, we can then compare it against the results from the data set and measurement from citation numbers. Another thing we should ask in the survey is their gender, and use this result to compare the API classified result to see how accurate it is, or if it misses out the names of the minorities as unclassified.

Reference

1. AlShebli, B., Makovi, K. & Rahwan, T. RETRACTED ARTICLE: The association between early career informal mentorship in academic collaborations and junior author performance. Nat Commun 11, 5855 (2020). https://doi.org/10.1038/s41467-020-19723-8
2. AlShebli, B., Makovi, K. & Rahwan, T. Retraction Note: The association between early career informal mentorship in academic collaborations and junior author performance. Nat Commun 11, 6446 (2020). https://doi.org/10.1038/s41467-020-20617-y
3. Jessica Hullman, Are female scientists worse mentors? This study pretends to know https://statmodeling.stat.columbia.edu/2020/11/19/are-female-scientists-worse-mentors-this-study-pretends-to-know/
4. Emmanuel Olabayo, Academic Mentorship. https://medium.com/@olabayoemmanuel/academic-mentorship-eda142fe7ac(
5. Average survey response rate - what you need to know https://www.customerthermometer.com/customer-surveys/average-survey-response-rate/

A copy of this paper can be found on my github at https://github.com/StefanZheng/STA304