Some considerations for studying gender, mentorship, and scientific impact: commentary on AlShebli, Makovi, and Rahwan (2020)

Daniel E. Acuna^{1,2}

School of Information Studies
Syracuse University

Abstract

Students, mentors, educational institutions, and funders want to make the best decision possible to maximize mentees' success. Previous work has revealed the critical long-term impact of mentor-mentee relationships, opening the door for policy-relevant recommendations that could improve mentor-mentee matching. The article authored by AlShebli, Makovi, and Rahwan (2020)—the AMR article from now on—focuses on the little-known effects of gender dynamics on these relationships. The most controversial recommendation in AMR is that women mentees would do better with men mentors than women mentors. At best, we think this claim relies on weak evidence; at worst, this claim could be misleading and problematic. In this commentary, we describe several issues regarding definitions, methodology, and data quality in AMR. We offer alternative hypotheses rooted in past research regarding women's lack of access to dissemination, training, and funding. If anything, these factors could significantly reduce or reverse the findings reported by AMR. We offer suggestions for controlling these effects and a research path forward for studying this important subject.

Issues with the definition of mentorship

AMR uses co-authorship as a proxy for mentorship. They surveyed a sample of co-authorship pairs to justify that junior co-authors felt mentored by senior ones. However, focusing on co-authorship alone biases the analysis substantially: graduate students receive mentorship from postdoctoral researchers, teachers, staff, and university officials without necessarily producing a publication (Byars-Winston & Dahlberg, 2020). Also, some interactions are not career-related but rather general support (Mullen, 2010; Rose et al., 2005). Another issue is that AMR uses the term “informal mentorship,” perhaps wrongly suggesting that the article is about non-publication-related mentorship opportunities.

¹ E-mail: deacuna@syr.edu Website: <https://acuna.io>

² Daniel E. Acuna was Reviewer #1 for AlShebli et al. The reviews are openly available. For more information about the process, please refer to https://twitter.com/daniel_akuna/status/1329475493440139268

Unfortunately, getting access to real “informal mentorship” is challenging and probably not captured by any openly available dataset yet. Future work should perhaps attempt to capture these aspects of mentorship. One possibility is to focus on crowdsourced mentorship networks such as AcademicTree.org, which contains crowdsourced data on advisor-advisee and data from Ph.D. dissertations (David & Hayden, 2012; Ma et al., 2020).

Issues with the definition of impact

Scientific impact is essential to many aspects of academic careers. Hiring, promotion, and funding decisions depend on it. The AMR study uses the average citations for five years as a measure of impact. However, citations do not capture all impact dimensions. If we take the de-facto impact assessment to grant tenure, committees consider the number of publications, funding, service, and mentorship (McKiernan et al., 2019; Rice et al., 2020). The full measure of impact, however, can go much beyond these traditional factors and it is an active area of inquiry (Lindner et al., 2018; Moed & Halevi, 2015; Cronin & Sugimoto, 2014; Smith et al., 2014; Thelwall et al., 2013; “How to Improve the Use of Metrics,” 2010; Bornmann & Daniel, 2009).

Issues with claiming causal effects of gender on impact using observational data

One of the most controversial claims in AMR states that women mentees should prefer to work with men mentors, and vice versa. We now explain why this recommendation might just be the result of spurious correlations rather than causation. Although AMR uses causal inference, the technique they employed requires correct assumptions. We describe assumptions rooted in research on gender and scientific careers that could have dramatically changed the results.

We can broadly classify techniques for making causal inferences into three categories. One traditional way of doing it, and widely considered the gold standard, is randomized control trials (RCT). In this setting, analysis units are randomly assigned to either a control condition or an experimental condition. If everything is followed correctly, this breaks any potential external influences between the treatment and outcome, allowing us to establish clean causal relationships. In AMR, it would be unethical or impractical because we would need to assign—and force—mentees to have a specific number of women mentors.

A second way of making causal inferences is to use observational data where a pseudo-experiment has occurred. The most straightforward such setting is when so-called *natural* experiments occur (Dunning, 2012). In natural experiments, there is an unintended event that randomly assigns units to conditions that are interesting to study. For example, to test the causal effect of different economic regimes such as Marxism-Stalinism vs. laissez-faire capitalism on individual behavior, we could study people who lived in East Germany vs. people who lived in West Germany after World War 2. People on both sides of Germany had the same

background information distribution but were pseudo-randomly assigned to one of the two sides. Many studies have used this natural experiment to study human behavior (Ockenfels & Weimann, 1999). In scientific analysis, for example, previous researchers have used similar random events to study the influence of prominent scientists—the sudden death of a scientist (Azoulay et al., 2019).

A third way of making causal inferences is by using purely observational data. These methods are usually the most attractive because observational data is cheap and abundant. However, these methods require significantly more assumptions, which are often impossible to test (Spirtes et al., 1993). In citation and publication datasets such as the Microsoft Academic Graph (MAG), we have to rely on these kinds of assumptions and causal inferences. The AMR study, in particular, uses relatively recent techniques to exploit natural variations in purely observational data. There is nothing intrinsically wrong with this approach, but it can lead to fragile causal inference results unless correct assumptions are made.

There are many techniques to do causal inference on observational data, some of them with better statistical properties than others. We cannot possibly review all of them—see (Cunningham, 2021; Hernan & Robins, 2019; Imbens & Rubin, 2015; Pearl et al., 2016). The technique employed in AMR, Coarsened Exact Matching (CEM), is among the newest ones. CEM promises to be very easily applicable—“If you understand the tradeoffs in drawing a histogram, you will understand how to use this method,” according to the authors (Iacus et al., 2012).

The basic premise of CEM (and similar techniques) is to find “experimental” units in the data (i.e., scientists with one or more women mentors) and compare them to “control” units (i.e., scientists with no women mentors). Importantly, a set of predefined covariates must be kept as close as possible between the experiment and control pairs. These pairs are chosen according to a matching mechanism where the covariates have been “coarsely” binned, improving the number of units we can have for any given combination of covariates. This binning improves the statistical power because we have more data points per “multidimensional” bin. CEM has desirable statistical properties (Iacus et al., 2012) and good software support (e.g., Bails, 2020).

CEM and other similar causal inference techniques based on observational data depend on assumptions about the statistical and causal relationships in the data. More specifically, to isolate the effects of causes from other confounding variables, we need to condition them and assume that there is no unmeasured confounding. Given a set of covariates, the assignment to the treatment group or groups is conditionally independent of the potential outcomes of the experiment. In our example at hand, knowing the potential impact of a researcher and his or her covariates should not allow us to predict how many women mentors he or she had. The notion of no unmeasured confounding has many other names, such as unconfoundedness and ignorability.

If the assumptions are incorrect, there could be three kinds of issues. The simplest one is that we have factors that we are not measuring but affect both the experimental variable and outcome. In the words of causal inference, we have *open* backdoors Pearl (2016), where confounding effects could be contributing to the relationship between experiment and outcome in unknown ways. A second issue is when we condition on the wrong variable. This conditioning might happen when we think we control for unmeasured confounding but are *opening* a backdoor, adding unknown dependencies between experiment and outcome. This phenomenon is sometimes called conditioning on a *collider*. The final kind of issue is when there is no direct causal relationship between a treatment and outcome. Rather, there is a path from treatment to a mediator and it is the mediator that has a causal relationship to the outcome. Strictly speaking, there *is* a causal relationship between the treatment and the outcome but only through the mediator (Pearl, 2014).

We believe there are reasons to suspect that at least two types of uncontrolled effects (mediators and controlling on a collider) are present in the AMR analysis, affecting the results' validity in unknown ways. We can at least think of three factors that affect the results' validity: popularity, training, and funding. Briefly, there are significant differences in how women disseminate their work, the life-long opportunities to accumulate experience before and during their careers, and access to funding.

Factor 1. Popularity. Women do not promote their work as much as men. Men tend to self-cite significantly more (King et al., 2017), use language that is more positive about their result (Jagsi & Silver, 2019), and have their work described more strongly in recommendation letters (Dutt et al., 2016). These factors affect how much others are aware of women's work: if fellow scientists have not heard of them, they will not get cited. This phenomenon might also explain why women do not get cited as much as men even when they might publish in higher-impact journals (Ghiasi et al., 2015). Because mentors and mentees share an intellectual space, this lack of "marketing" will affect the dissemination of mentees' ideas too. Therefore, there is no direct cause between gender and impact but rather between popularity and impact.

Factor 2. Training opportunities. Becoming a scientist is a process that starts much earlier than getting a Ph.D. degree. Scientists often describe positive childhood experiences that fuel a lifelong passion for science (Funk & Hefferon, 2016). However, because of differences in society's perceptions of what men and women should do (Makarova et al., 2019), women end up taking less advantage of training opportunities in science. During adulthood, these negative reinforcements continue to compound. Many graduate trainees leave academia because of expectations not aligned with a scientific career (Jadidi et al., 2017; Makarova et al., 2016), mostly related to raising children (Ceci et al., 2014; Ginther & Kahn, 2014; Wolfinger et al., 2008). Structural factors also affect the acceptance and support for women (Curtin et al., 2016; Fried et al., 1996; Levine et al., 2013; Noy & Ray, 2012; Seagram et al., 1998; Varkey et al., 2012). The randomness of the COVID-19 pandemic has shed light on the difficulty of women to

work from home, diminishing their accumulated opportunities for impact (Myers et al., 2020; Squazzoni et al., 2020). And these factors can be more pronounced in some countries and cultures (Gupta et al., 2005). Therefore, there is a direct cause between training and impact but not gender and impact.

Fig. 1. Different assumptions might not be able to capture the causal effect of women mentors on mentee impact. **A.** Simplified Directed Acyclic Graph (DAG) assumed for (AIShebli et al., 2020). We do not know what the structure they considered was. Still, their analysis suggested that controlling for the following would remove confounders: average protege big-shot experience (B), number of mentors (M), the gender of the mentee (G), discipline (D), affiliation rank (A), and year of first mentored paper (Y). This conditioning would have allowed us to measure the direct causal effect between the number of female mentors (F) on the impact (I). **B.** In our alternative formulation, we assume, based on previous research, that gender does not have a different effect on impact but rather indirect effects through popularity (of the institution and the authors, P), training (T), and funding ($\$$). In causal inference terminology, these are factors that *mediate* the effect of mentors' gender on mentee's impact. We also add the assumption that funding and popularity of faculty affect the rank of an affiliation. Therefore, by conditioning on the affiliation rank (A), the authors have inadvertently created an *open backdoor* between income ($\$$) and popularity, further confounding the effect of the mentors' gender and the impact of mentees' publications.

Factor 3. Funding. Women scientists might have more difficulties attracting and obtaining funding (Jagsi & Silver, 2019; Kaatz et al., 2016), diminishing their ability to produce impact for their field. Additionally, because perceived ranks of institutions are subjectively affected by funding and the broad dissemination of the institution's publications and work (Harvey, 2008), controlling for ranking also correlates funding and dissemination factors. Therefore, to solve

both issues, we should control for funding and popularity simultaneously, as we hypothesize that they cause higher impact rather than gender.

We illustrate how these factors play out using graphical representations known as Directed Acyclic Graphs (DAGs) (Hernan & Robins, 2019; Pearl et al., 2016). There are multiple factors and choices that the AMR paper made that affect the causal inferences they are making. The treatment variable is the number of female mentors F_0 through F_5 , representing 0 to 5 female mentors. The AMR controls the protege’s average big-shot experience, the number of mentors, gender, discipline, affiliation rank, and the year of the first mentored paper (Fig. 1A). However, we suggest that the three factors explained above can have unaccounted effects between the number of women mentors and the mentee’s impact. In this alternative view, the variables used in AMR do not sufficiently control for them (Fig 1B).

A solution to improve the causal inference is to control for these three factors. We hypothesize that it could be possible to measure the degree of dissemination of a mentor by controlling for the language they use and how much they “advertise” their work. Does their work appear in the popular press, interviews, commentaries? As for the training opportunities, it could be challenging to control because we would need to measure the lifelong exposure to training opportunities and the amount of time spent on these activities. Finally, there are plenty of datasets that could be used to control for the funding of the mentor and the academic institution.

Issues with the data

The Microsoft Academic Graph (MAG) is a stimulating attempt by the Microsoft Academic Search unit to provide researchers with data to study science. This dataset is the biggest and most comprehensive openly available dataset to research publications, citations, scientists, and institutions. All other datasets, such as Web of Science and Scopus, are not open. While there are APIs to access them, they are limited.

As crucial as this dataset is, the coverage and validation of MAG are limited. A recent study showed a 60% coverage, lower than Google Scholar 88% coverage (Martín-Martín et al., 2020). Also, it is unclear if there are biases in the probability of a publication appearing in MAG. We could imagine that because women tend to publicize their work less than men (Jagsi & Silver, 2019), their work is less available on the open web, where the MAG obtains its information (Wang et al., 2019). Also, the internal workings of the MAG’s disambiguation algorithm are unclear. For example, the data for disambiguating names might be richer and better for men names than women names—men make up more than 60% of the scientific force (Huang et al., 2020).

Issues with statistical analysis

Statistical significance tests hide many subtleties in the data behind a single quantity: p-value (McShane et al., 2019). However, another critical factor is effect sizes (Greenland et al., 2016). The AMR found a significant difference in “gain” for training a woman mentee for a woman mentor. The result essentially suggests that working with women mentees is also detrimental for women mentors. But in practical terms, what is the effect size of this difference? What does this difference mean?

AMR measures the benefit of mentors working with mentees as the average citations within five years of papers co-authored together. AMR shows a significant difference of approximately 6%. Assuming a generous 100 average citations received during the collaboration period (we believe that this is *much* lower), this would mean that the difference between working with a woman mentee vs. a man mentee is only six citations. Should we be recommending such essential decisions as choosing to mentor a woman just because of *six citations*? Probably not.

Acknowledgments

The author would like to thank Jianxuan Liu for preliminary discussions about the manuscript.

References

- Achakulvisut, T., Bhagavatula, C., Acuna, D., & Kording, K. (2019). Claim extraction in biomedical publications using deep discourse model and transfer learning. arXiv preprint arXiv:1907.00962.
- Acuna, D. E., & Penner, O. (2013). The future h-index is an excellent way to predict scientists’ future impact. *Controversies in Medical Physics: a Compendium of Point/Counterpoint Debates* Volume 3, 274.
- AlShebli, B., Makovi, K., & Rahwan, T. (2020). The association between early career informal mentorship in academic collaborations and junior author performance. *Nature Communications*, 11(1), 5855. <https://doi.org/10.1038/s41467-020-19723-8>
- Azoulay, P., Fons-Rosen, C., & Graff Zivin, J. S. (2019). Does Science Advance One Funeral at a Time? *American Economic Review*, 109(8), 2889–2920. <https://doi.org/10.1257/aer.20161574>
- Bails, L. (2020). *cem: A Python implementation of Coarsened Exact Matching for causal inference* (0.1.2) [Python]. <https://github.com/lewisbails/cem>

- Bornmann, L., & Daniel, H.-D. (2009). The state of h index research. Is the h index the ideal way to measure research performance? *EMBO Reports*, 10(1), 2–6.
<https://doi.org/10.1038/embor.2008.233>
- Byars-Winston, A., & Dahlberg, M. L. (2020). *The Science of Effective Mentorship in STEM*. National Academies Press.
- Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in Academic Science: A Changing Landscape. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, 15(3), 75–141.
<https://doi.org/10.1177/1529100614541236>
- Cronin, B., & Sugimoto, C. (2014). *Beyond Bibliometrics* | *The MIT Press*. The MIT Press.
<https://mitpress.mit.edu/books/beyond-bibliometrics>
- Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.
- Curtin, N., Malley, J., & Stewart, A. J. (2016). Mentoring the Next Generation of Faculty: Supporting Academic Career Aspirations Among Doctoral Students. *Research in Higher Education*, 57(6), 714–738. <https://doi.org/10.1007/s11162-015-9403-x>
- David, S. V., & Hayden, B. Y. (2012). Neurotree: A collaborative, graphical database of the academic genealogy of neuroscience. *PLoS ONE*, 7, e46608.
<https://doi.org/10.1371/journal.pone.0046608>
- Dunning, T. (2012, September). *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge Core; Cambridge University Press.
<https://doi.org/10.1017/CBO9781139084444>
- Dutt, K., Pfaff, D. L., Bernstein, A. F., Dillard, J. S., & Block, C. J. (2016). Gender differences in recommendation letters for postdoctoral fellowships in geoscience. *Nature Geoscience*, 9(11), 805–808. <https://doi.org/10.1038/ngeo2819>
- Dworkin, J. D., Linn, K. A., Teich, E. G., Zurn, P., Shinohara, R. T., & Bassett, D. S. (2020). The extent and drivers of gender imbalance in neuroscience reference lists. *Nature Neuroscience*, 23(8), 918–926. <https://doi.org/10.1038/s41593-020-0658-y>
- Ethier, C., Acuna, D., Solla, S. A., & Miller, L. E. (2016). Adaptive neuron-to-EMG decoder training for FES neuroprostheses. *Journal of neural engineering*, 13(4), 046009.
- Fried, L. P., Francomano, C. A., MacDonald, S. M., Wagner, E. M., Stokes, E. J., Carbone, K. M., Bias, W. B., Newman, M. M., & Stobo, J. D. (1996). Career Development for Women in Academic Medicine: Multiple Interventions in a Department of Medicine.

- JAMA*, 276(11), 898–905. <https://doi.org/10.1001/jama.1996.03540110052031>
- Funk, C., & Hefferon, M. (2016). *Why do people become scientists? What scientists told us*. <https://www.pewresearch.org/fact-tank/2016/10/24/as-the-need-for-highly-trained-scientists-grows-a-look-at-why-people-choose-these-careers/>
- Ghiasi, G., Larivière, V., & Sugimoto, C. R. (2015). On the Compliance of Women Engineers with a Gendered Scientific System. *PLoS ONE*, 10(12). <https://doi.org/10.1371/journal.pone.0145931>
- Ginther, D. K., & Kahn, S. (2014). *Women's Careers in Academic Social Science: Progress, Pitfalls, and Plateaus*. Undefined. </paper/Women%27s-Careers-in-Academic-Social-Science%3A-and-Ginther-Kahn/0aa43c3a8bbc9b74f7f8a3b9716f6a23d1c3636c>
- Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, 31(4), 337–350. <https://doi.org/10.1007/s10654-016-0149-3>
- Gupta, N., Kemelgor, C., Fuchs, S., & Etzkowitz, H. (2005). Triple burden on women in science: A cross-cultural analysis. *Current Science*, 89(8), 1382–1386.
- Harvey, L. (2008). Rankings of Higher Education Institutions: A Critical Review. *Quality in Higher Education*, 14(3), 187–207. <https://doi.org/10.1080/13538320802507711>
- Hernan, M. A., & Robins, J. M. (2019). *Causal Inference*. Taylor & Francis.
- How to improve the use of metrics. (2010). *Nature*, 465(7300), 870–872. <https://doi.org/10.1038/465870a>
- Huang, J., Gates, A. J., Sinatra, R., & Barabási, A.-L. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences*, 117(9), 4609–4616. <https://doi.org/10.1073/pnas.1914221117>
- Iacus, S. M., King, G., & Porro, G. (2012). Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis*, 20(1), 1–24. <https://doi.org/10.1093/pan/mpr013>
- Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139025751>

- Jadidi, M., Karimi, F., Lietz, H., & Wagner, C. (2017). Gender disparities in science? Dropout, productivity, collaborations and success of male and female computer scientists. *Advances in Complex Systems*, 21(03n04), 1750011. <https://doi.org/10.1142/S0219525917500114>
- Jagsi, R., & Silver, J. K. (2019). Gender differences in research reporting. *BMJ*, 367. <https://doi.org/10.1136/bmj.l6692>
- Jas, M., Achakulvisut, T., Idrizović, A., Acuna, D., Antalek, M., Marques, V., ... & Ramkumar, P. (2020). Pyglmnet: Python implementation of elastic-net regularized generalized linear models. *Journal of Open Source Software*, 5(47).
- Kaatz, A., Lee, Y.-G., Potvien, A., Magua, W., Filut, A., Bhattacharya, A., Leatherberry, R., Zhu, X., & Carnes, M. (2016). Analysis of NIH R01 Application Critiques, Impact and Criteria Scores: Does the Sex of the Principal Investigator Make a Difference? *Academic Medicine : Journal of the Association of American Medical Colleges*, 91(8), 1080–1088. <https://doi.org/10.1097/ACM.0000000000001272>
- King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J., & West, J. D. (2017). Men Set Their Own Cites High: Gender and Self-citation across Fields and over Time. *Socius*, 3, 2378023117738903. <https://doi.org/10.1177/2378023117738903>
- Lee, T. G., Acuna, D. E., Kording, K. P., & Grafton, S. T. (2019). Limiting motor skill knowledge via incidental training protects against choking under pressure. *Psychonomic bulletin & review*, 26(1), 279-290.
- Levine, R. B., Mechaber, H. F., Reddy, S. T., Cayea, D., & Harrison, R. A. (2013). “A Good Career Choice for Women”: Female Medical Students’ Mentoring Experiences A Multi-Institutional Qualitative Study. *Academic Medicine*, 88(4), 527–534. <https://doi.org/10.1097/ACM.0b013e31828578bb>
- Lindner, M., Torralba, K., & Khan, N. (2018). Scientific productivity: An exploratory study of metrics and incentives. *PLOS ONE*.
- Ma, Y., Mukherjee, S., & Uzzi, B. (2020). Mentorship and protégé success in STEM fields. *Proceedings of the National Academy of Sciences*, 117(25), 14077–14083. <https://doi.org/10.1073/pnas.1915516117>
- Makarova, E., Aeschlimann, B., & Herzog, W. (2016). Why is the pipeline leaking? Experiences of young women in STEM vocational education and training and their adjustment strategies. *Empirical Research in Vocational Education and Training*, 8(1), 2. <https://doi.org/10.1186/s40461-016-0027-y>

- Makarova, E., Aeschlimann, B., & Herzog, W. (2019). The Gender Gap in STEM Fields: The Impact of the Gender Stereotype of Math and Science on Secondary Students' Career Aspirations. *Frontiers in Education*, 4. <https://doi.org/10.3389/feduc.2019.00060>
- Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & Delgado López-Cózar, E. (2020). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations. *Scientometrics*. <https://doi.org/10.1007/s11192-020-03690-4>
- McKiernan, E. C., Schimanski, L. A., Muñoz Nieves, C., Matthias, L., Niles, M. T., & Alperin, J. P. (2019). Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations. *ELife*, 8, e47338. <https://doi.org/10.7554/eLife.47338>
- McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon Statistical Significance. *The American Statistician*, 73(sup1), 235–245. <https://doi.org/10.1080/00031305.2018.1527253>
- Moed, H. F., & Halevi, G. (2015). Multidimensional assessment of scholarly research impact. *Journal of the Association for Information Science and Technology*, 66(10), 1988–2002. <https://doi.org/10.1002/asi.23314>
- Mullen, C. A. (2010). Naturally Occurring Student-Faculty Mentoring Relationships: A Literature Review. In *The Blackwell Handbook of Mentoring: A Multiple Perspectives Approach*. John Wiley & Sons.
- Myers, K. R., Tham, W. Y., Yin, Y., Cohodes, N., Thursby, J. G., Thursby, M. C., Schiffer, P., Walsh, J. T., Lakhani, K. R., & Wang, D. (2020). Unequal effects of the COVID-19 pandemic on scientists. *Nature Human Behaviour*, 4(9), 880–883. <https://doi.org/10.1038/s41562-020-0921-y>
- Noy, S., & Ray, R. (2012). Graduate Students' Perceptions of Their Advisors: Is There Systematic Disadvantage in Mentorship? *The Journal of Higher Education*, 83(6), 876–914. <https://doi.org/10.1080/00221546.2012.11777273>
- Ockenfels, A., & Weimann, J. (1999). Types and patterns: An experimental East-West-German comparison of cooperation and solidarity. *Journal of Public Economics*, 71(2), 275–287. [https://doi.org/10.1016/S0047-2727\(98\)00072-3](https://doi.org/10.1016/S0047-2727(98)00072-3)
- Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological Methods*, 19(4), 459–481. <https://doi.org/10.1037/a0036434>
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. John Wiley & Sons.

- Rice, D. B., Raffoul, H., Ioannidis, J. P. A., & Moher, D. (2020). Academic criteria for promotion and tenure in biomedical sciences faculties: Cross sectional analysis of international sample of universities. *BMJ*, 369. <https://doi.org/10.1136/bmj.m2081>
- Rose, G. L., Rukstalis, M. R., & Schuckit, M. A. (2005). Informal Mentoring Between Faculty and Medical Students. *Academic Medicine*, 80(4), 344–348.
- Seagram, B. C., Gould, J., & Pyke, S. W. (1998). AN INVESTIGATION OF GENDER AND OTHER VARIABLES ON TIME TO COMPLETION OF DOCTORAL DEGREES. *Research in Higher Education*, 39(3), 319–335. <https://doi.org/10.1023/A:1018781118312>
- Smith, M. J., Weinberger, C., Bruna, E. M., & Allesina, S. (2014). The Scientific Impact of Nations: Journal Placement and Citation Performance. *PLOS ONE*, 9(10), e109195. <https://doi.org/10.1371/journal.pone.0109195>
- Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, Prediction, and Search*. Springer-Verlag. <https://doi.org/10.1007/978-1-4612-2748-9>
- Squazzoni, F., Bravo, G., Grimaldo, F., Garcia-Costa, D., Farjam, M., & Mehmani, B. (2020). *No Tickets for Women in the COVID-19 Race? A Study on Manuscript Submissions and Reviews in 2347 Elsevier Journals during the Pandemic* (SSRN Scholarly Paper ID 3712813). Social Science Research Network. <https://doi.org/10.2139/ssrn.3712813>
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. *PLOS ONE*, 8(5), e64841. <https://doi.org/10.1371/journal.pone.0064841>
- Varkey, P., Jatoi, A., Williams, A., Mayer, A. P., Ko, M., Files, J., Blair, J., & Hayes, S. (2012). The positive impact of a facilitated peer mentoring program on academic skills of women faculty. *BMC Medical Education*. <https://doi.org/10.1186/1472-6920-12-14>
- Wang, K., Shen, Z., Huang, C., Wu, C.-H., Eide, D., Dong, Y., Qian, J., Kanakia, A., Chen, A., & Rogahn, R. (2019). A Review of Microsoft Academic Services for Science of Science Studies. *Frontiers in Big Data*, 2. <https://doi.org/10.3389/fdata.2019.00045>
- Wolfinger, N. H., Mason, M. A., & Goulden, M. (2008). Problems in the Pipeline: Gender, Marriage, and Fertility in the Ivory Tower. *The Journal of Higher Education*, 79(4), 388–405. <https://doi.org/10.1080/00221546.2008.11772108>