nature

https://doi.org/10.1038/s41586-022-04966-w

Accelerated Article Preview

Women are Credited Less in Science than are Men

Received: 25 February 2021

Accepted: 10 June 2022

Accelerated Article Preview

Cite this article as: Ross, M. B. et al. Women are Credited Less in Science than are Men. *Nature* https://doi.org/10.1038/s41586-022-04966-w (2022)

Matthew B. Ross, Britta M. Glennon, Raviv Murciano-Goroff, Enrico G. Berkes, Bruce A. Weinberg & Julia I. Lane

Women are Credited Less in Science than are Men

Ross, M. B.[1]
Glennon, B. M. [2,3]
Murciano-Goroff, R. R.[4]
Berkes, E. G. [5]
Weinberg, B.A. [5,3]
Lane, J. I. [6*]

Affiliations: [1] Department of Economics and School of Public Policy & Urban Affairs, Northeastern University, Boston, Massachusetts, USA [2] Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania, USA[3] National Bureau of Economic Research, Cambridge, Massachusetts, USA [4] Questrom School of Business, Boston University, Boston, Massachusetts, USA [5] Department of Economics, Ohio State University, Columbus, Ohio, USA [6] Wagner School of Public Policy, New York University, New York, New York, USA.

Abstract

There is a well-documented gap in the observed number of scientific works produced by women and men in science, with clear consequences for the retention and promotion of women in science [1]. The gap might be a result of productivity differences [2-5]**, or it might be due to women's contributions not being acknowledged [6, 7]. This paper finds that at least part of this gap is due to the latter: women in research teams are significantly less likely to**

be credited with authorship than are men. The findings are consistent across three very different sources of data. Analysis of the first source - large scale administrative data on research teams, team scientific output, and attribution of credit - show that women are significantly less likely to be named on any given article or patent produced by their team relative to their peers. The gender gap in attribution is found across almost all scientific fields and career stages. The second source – an extensive survey of authors – similarly shows that women's scientific contributions are systematically less likely to be recognized. The third source – qualitative responses – suggests that the reason is that their work is often not known, not appreciated, or ignored. At least some of the observed gender gap in scientific output may not be due to differences in scientific contribution, but to differences in attribution.

Main

Gender differences in observed scientific output are well documented: women both publish and patent less than are men [1]. The causes of those differences are less understood. Analysis using individual data has suggested that women are less productive because they work in less welcoming work environments [2], have greater family responsibilities [3], have different positions in the laboratory [8] or differ in supervision [5]. Intriguing new work suggests that women are not less productive, but that their work is undervalued [9]. The analysis in this paper uses new data on research teams to suggests that women are accorded less credit than are men: simply put, they are systematically less likely to be named as authors on articles and patents.

The possibility that women receive less recognition for their scientific contributions is not hypothetical: the canonical example is that of Rosalind Franklin. Franklin's pivotal contribution to the discovery of the structure of DNA initially went unrecognized [6], and it was not until long after she died that science realized that she was wrongfully denied authorship on the original Crick and Watson paper. Indeed, her contribution was apparently only recognized because Watson's account of the discovery was so incorrect [10] and stimulated a reconstruction of events by Franklin's friends [11]. More recently, Walter Isaacson recounts Jennifer Doudna's concern that she and Emmanuelle Charpentier were being relegated to "minor players" in the history and commercial use of CRISPR-Cas9 (p210, [7]). The open questions, of course, are how many women's contributions have been missed in similar but less high-profile circumstances, and how many women have been discouraged from pursuing careers in science as a result [12].

Finding *what isn't there* from *what is there* is a fundamental problem in mathematics, and has been used to address such vastly different questions as calculating the return on investment on mutual funds (after accounting for funds that no longer exist) or the optimal placement of armor on World War II airplanes (after accounting for planes that did not return) [13]. The problem of selecting on the dependent variable is also prevalent in the social sciences, for example, in only observing the labor supply of people who participate in the labor market [14] or studying the drivers of economic development by selecting a few successful industrializing countries [15].

The first steps in answering what data were missing in these two examples are to describe the population from which the sample of observations is drawn and document the degree of missingness. Subsequent steps then characterize the sources of the missingness. The large-scale bibliometrics databases used to study scientific output consist only of named authors or inventors (not unnamed contributors), and cannot be used to find who is not named; carefully curated case studies are too small to generalize [16]. The unique data on research teams used in this paper are, by contrast, fit for purpose: they consist of information on 9,778 teams over a four-year period, the 128,859 individuals working in those teams, matched to 39,426 journal articles and 7,675 patents produced by those teams and are described in more detail in the Methods: Construction of Administrative Data section. Because the data include information about the positions held by each individual on each team, as well as their gender, it is possible to calculate for each individual whether they *did* or *did not* receive credit on a given article and to calculate differences by gender.

The evidence generated from the analysis described in this paper suggests that Rosalind Franklin is far from unique in not receiving credit for her work. If credit is defined simply as ever being named an author, women account for only 34.85% of the authors on a team, even though they make up just under half of the workforce (48.25%; Extended Data Table 2). When credit is defined as the likelihood of being credited on a given document (relative to the mean) produced by a research team, there is a 13.24% to 58.40% gap in the likelihood that women are named on any given article or patent respectively produced by their team (Extended Data Table 4, Column 5). The chances of women receiving credit on an article decrease by 4.78% relative to the baseline rate of 3.18% ($p<0.0001$; two-sided t-test; test value=-3.8; effect size=-0.0015 percentage points) for each 1 log point increase in citations (Extended Data Table 7).

The results are confirmed by appealing to another completely different source of quantitative data – a survey of 2,446 scientists about the allocation of credit (See Methods: Survey Design and Collection section and Supplementary Information Part 3). Exclusion from authorship is common and differs significantly by gender: 42.95% of women and 37.81% of men reported that they had been excluded from authorship ($p=0.0151$; two-sided t-test; test value=-2.4327; effect size=-0.0514), and significantly more women (48.97%) than men (39.13%) report that others underestimated their contribution ($p=0.0036$; two-sided t-test; test value= -2.9218; effect size=-0.0984).

Qualitative analysis – open ended narrative statements by survey respondents as well as personal interviews with consenting authors (approach detailed in the Methods sections: Survey Design and Collection and Qualitative Evidence and Supplemental Information Part 3) – were also consistent. Authors noted that the rules of credit allocation were frequently unclear, and often determined by senior investigators. A complex mix of factors, particularly field, rank, culture, and gender, was identified. However, an overarching theme was that the rules governing scientific contributions were often not codified, not understood by all members of the research team, or simply ignored. The necessary level of work required for authorship is often not clear to everyone participating on research teams, and the level of work deemed necessary to receive attribution can vary based on the idiosyncratic personal preferences and a team member's relationship with the Principal Investigator (PI). Thus, women—and other historically marginalized groups—must often put in significantly more effort in order for their scientific contributions to be recognized.

Our analyses on administrative, survey, and qualitative data suggest that even 70 years later, the same factors that led to Rosalind Franklin's denial of authorship on the pivotal work on the structure of DNA are still at work. At least some of the observed gender gap in scientific output may not be due to differences in scientific contribution, but to differences in attribution within research teams.

Attribution and Administrative Data

Unpacking the structure of research teams to understand whose work is not recognized requires identifying each individual on each research team, characterizing their position by their job title, and then determining whether or not they are named on the articles and patents produced by the research team. Administrative data can be used to provide highly granular information about who works on which research project because human resource records both document every payment that is made each pay period on each grant and provide information on each employee's job title. Currently 118 campuses from 36 participating universities provide their deidentified data to the Institute for Research on Innovation and Science at the University of Michigan which processes and standardizes the information into analytical files [17]. The earliest year that data is provided by a participating institution is 2000; the latest is 2019, and the data include information on payments of wages from individual grants to all people employed by each grant, including information on the job title for which a person is paid on a particular grant (More detail provided in the Methods Section: Construction of Administrative Data).

Teams were constructed around a central PI, their associated grants, and individuals employed on those grants from 2013-16. The scientific field of each team is identified by using the title of all associated grants and comparing the grants with a pool of text that describes each scientific field using a "Wiki-labeling" approach [17-19]. Scientific documents were linked to a team if the article or patent acknowledged one of the team's grants and/or any member of the team was listed as an author on that article or patent (More detail provided in the Methods Section: Construction of Administrative Data).

Attribution can be measured in many ways using these data. Three measures are constructed for the purposes of this paper: (i) the rate at which individuals are ever named as an author on any scientific document – the "ever-author" rate, (ii) the rate at which individuals are named as an author on a given scientific document produced by their team – the "attribution" rate and (iii) the rate at which individuals are named to any given high impact document – the "high-impact attribution" rate (See Methods: Analytical Sample).

The first, and simplest, measure is the "*ever author*" rate, which characterizes an individual as an author if he or she was ever named as an author or an inventor during the analysis period. While 16.97% of individuals are classified as authors using this measure as shown in Table 1, the probability that men are named is 21.17% compared to 12.15% of women to be so recognized. Table 1 also shows that there are two reasons for this gap: women's junior position in research teams, and underrepresentation in attribution given their position. First, women are less likely to be the senior positions that are associated with authorship. The highest authorship rate is for faculty, at 45.70%, yet only 11.30% of women (versus 19.72% of men) in the scientific workforce are faculty. Conversely, the authorship rate for research staff is 8.63%, yet 47.81% of women are research staff compared to 28.73% of men. Second, holding constant the distribution of positions (at the grand means), women are 4.82% less likely to be named as authors. In the case of graduate students, for example, 14.97% of women are ever named as an author on a document compared with 21.37% of men. The consequences of such disparities on the retention of senior women in, and the attraction of young women to, scientific careers are unlikely to be positive.

Although illustrative, the "*ever author*" rate does not fully capture differential attribution. In our motivating example, Franklin could have been named as an author on some articles or patents emanating from the research team other than the DNA paper with Crick and Watson. The second authorship measure is the "attribution rate" which represents the likelihood a woman receives credit on a given scientific document produced by her research team.

The empirical implementation of what is a relatively straightforward conceptual framework is more difficult, but the data are rich enough to allow such calculations (see Methods: Analytical Sample section for details). The denominator – the set of "*potential authorships*" – was created by associating all members of each team who were employed one year prior to the publication/application date of all associated articles/patents emanating from that team during the analysis period. Since some individuals, such as research staff, are on multiple teams, they are proportionately allocated across teams using a set of analytical weights (see Methods Analytical Sample section for details). The numerator – attribution – was defined as "*actual authorships*" on those publications and patents. Thus, the attribution rate is the ratio of actual authorships to potential authorships. The overall attribution rate for any team member on either a patent or article is 3.2%. On average across all job titles and fields, women have a 2.12% probability of being named on any scientific document while men are twice as likely to named, at 4.23% (p=.0000; two-sided t-test value: 19.5823 effect size: 2.11% Extended Data Table 2 and 3).

Women in each position are systematically less likely to be named an author on any given article or patent than are men for any given position in the organization. The data are rich enough to examine whether the observed gender differences simply reflect gender differences in organizational position rather than attribution.

Figure 1 (and SI Figure S5) makes use of information in the data about each individual's career position – faculty, postdoc, graduate student, undergraduate student or research staff – as well as the research team's field. Women do occupy more junior career positions than men. The first panel (a) of Figure 1 shows that this is indeed the case: the proportion of women in each position declines as the seniority of the position increases. At the top extreme, of potential faculty authors, 34.82 % are women. At the other extreme, 60.81% of potential research staff authors are women.

However, the first panel (a) also shows that the share of actual authorships for women is less than what would be expected given their share of potential authorships in each career position. The difference between the share of potential authorships and actual authorships for women ranges from 15.72 percentage points for research staff (p=.0000; two-sided t-test value: -15.81; effect size: 15.72 percentage points) to 7.09 percentage points for faculty (p=.0000; two-sided t-test value: -13.34; effect size: 7.09 percentage points) to 5.51 percentage points for postdocs (p=.0000; two-sided t-test value: -5.08; effect size: 5.51 percentage points; Extended Data Table 3). These gaps are visually apparent in the figure as every marker in panel (a) is positioned below the 45-degree line in panel (a) (see Extended Data Table 3 and SI Figure 5).

A similar pattern is apparent when authorship is analyzed by field. The second panel (b) of Figure 1 plots the share of actual authorships for women against the share of women among potential authors by field. For example, in biology, the share of actual authorships who are women is 15.02 percentage points lower than the share of women among potential authors (p=.0000; two-sided t-test value: -3024; effect size: 15.02 percentage points; Extended Data Table 3). In physical science, the comparable difference is 14.12 percentage points (p=.0000; two-sided t-test value: -25.44; effect size: 14.12 percentage points; Extended Data Table 3).

Figure 1 HERE

It is possible, of course, that the gender differences arise from compositional differences between women and men in terms of the teams on which they work, fields, job titles, or time allocated to particular projects. In particular, women might sort into teams with different propensities to publish or on projects with different research questions. Figure 2 (and SI Figure S6) plots the estimated attribution rate for men and women on articles (left) and patents (right) as well as the differences (indicated by Δ). The estimates are generated from a series of regression models that control for these types of potential compositional differences (Extended Data Table 4). In these models, an indicator for being named is regressed on an indicator for gender as well as an increasingly expansive set of control variables. Column (1) includes no controls; Column (2) adds publication date (calendar year x month), days worked on the team, and an indicator for the individual being a PI; Column (3) adds job title indicators; Column (4) adds field controls; and Column (5) adds indicator variables for each team. Including these additional controls reduces, but does not eliminate, the disparity for women. Even in the fully specified model, which adds controls for each research team, women are 13.24% (p<0.0001; two-sided t-test; test value=-6.3788; effect size=-0.4210 percentage points) less likely to be named on articles and 58.40% (p<0.0001; two-sided t-test; test value=-10.7746; effect size=-0.7652 percentage points) less likely to be named on patents.

The estimated regression-adjusted gender differences in attribution rates across job titles and fields, controlling for a wide variety of observable factors are reported in Extended Data Tables 5 and 6. Notably, the differences hold for all job titles except undergraduates, and of 13 fields for all but four for publications and five for patents.

Figure 2 here.

The third measure reflects the fact that not all scientific documents are created equal. The omission of Franklin from the Crick and Watson paper was particularly egregious because of its high potential and ultimate scientific impact. The empirical implementation of the third measure is to attach forward citations to the articles and patents. Figure 3 shows that, when controlling for field, career position and team size, there is no significant difference between the likelihood of a woman being named relative to a man on an article with zero citations (p=0.1725; two-sided t-test; test value=1.3642; effect size=0.1392 percentage points). However, for more highly cited output, women are less likely to be named than are men. For example, on an article with 25 citations women are 19.9739% less likely to be named than are men relative to the baseline (p<0.0001; two-sided t-test; test value=-7.4982; effect size=0.6352 percentage points; Extended Data Table 7).

Figure 3 here

Attribution and Survey Data

Qualitative evidence about the reasons behind the lack of attribution can be obtained from surveys. Despite the well-known issues with selection bias, self-reporting, and low response rates, survey data can be useful for triangulating against administrative data [18]. We designed a survey of authors who appeared on at least one article in the Web of Science [19] after 2014 and who had a published and available e-mail address. We asked three core sets of questions of each individual to shed light on the findings from our analysis of administrative data (the full survey can be found in Part 3 of the Supplementary Information).

In order to get a sense of how often scientists were not appropriately credited, we asked whether respondents had ever been excluded from a paper to which they had contributed. Of the 2,660 responses, there is a clear gender difference, with 42.95% of women and 37.81% of men having been excluded as an author (p=0.0151; two-sided t-test; test value=-2.4327; difference=-0.0514). This gap is qualitatively similar to the gaps estimated using the administrative data, where men were almost twice as likely (at 21.17%) to be recognized as ever being an author/inventor as women (at 12.15%), and the attribution rate on potential authorships/inventorships for men was 4.23% relative to 2.12% for women.

We also asked why respondents thought they were not credited: Figure 4 (and SI Figure S7) summarizes the results for the 871 individuals who responded (483 men and 388 women). The most common reason was that scientific contributions were underestimated, and this was the case for far more women, at 48.97%, than men, at 39.13% (p=0.0036; two-sided t-test value=-2.9218; effect size=-0.0984). While discrimination or bias was much less likely to be cited, women were twice as likely (at 15.46%) to cite this as a reason than men (at 7.67%), and that difference was significant (p=0.0003; two-sided t-test value= -3.6623; effect size: -0.0780). Men were more likely to say their contributions did not warrant authorship (37.68% of men compared with 24.74% of women; p=0.0000; two-sided t-test value=4.1060; effect size=0.1294). Differences in responsibilities (i.e. a respondent indicated that they were not granted attribution for at least one of the following reasons: personal, non-research responsibilities, and/or left the lab) appear to account for some of the attribution gap – 17.53% of excluded women cited these reasons, compared with 12.63% of men (p=0.0432; two-sided t-test value=-2.0244; effect size=-0.0490). Taken together, these estimates suggest that a large portion of the gender gap in attribution is due to either discrimination or how contributions are perceived by collaborators, or both.

Figure 4 here

The same question - are women with the same contribution as men less likely to be credited – can be asked a different way - conditional on being credited, did women contribute more than men? Accordingly, we asked authors to indicate what they did to earn authorship on one of their most recent publications using the standardized contributions identified by Project Credit [20]. The results, reported in Figure 5 and in SI Figure S7, are consistent: women must do more to be included as an author than men on average (2,297 individuals responded: 1,371 men and 926 women). A simple unweighted count of total contributions reported shows that women report a total 6.34 contributions on average compared to 6.11 for men (p=0.0907; two-sided t-test value=-1.6925; effect size=-0.2376). Women report making statistically significantly more contributions in conceptualization (64.99% of men vs. 68.36% of women; p=0.0937; two-sided t-test value=-1.6767; effect size=-0.0337), data curation (37.42% of men vs. 44.38% of women; p=0.0008; two-sided t-test value=-3.3467; effect size of -.0697), writing the original draft (45.73% of men vs. 52.48% of women; p=0.0015; two-sided t-test value=-3.1813; effect size=-0.0675) and reviewing and editing (82.57% of men vs. 86.18% of women; p=0.0205; two-sided t-test value=-2.3178; effect size=-0.0361). The only category in which men reported a greater contribution was software (18.31% of men vs. 11.67% of women; p=0.0000; two-sided t-test value=4.3174; effect size=0.0664). There is no significant difference between men and women in either formal analysis (49.23% of men vs. 51.94% of women; p=0.2028; two-sided t-test value=-1.2740; effect size=-0.0271) or project administration (32.82% of men vs. 35.75% of women; p=0.1471; two-sided t-test value=-1.4504; effect size=-0.0292).

Figure 5 Here

Attribution and Qualitative Data

The third source of information came from the voices of scientists themselves in two ways. First, the survey permitted open-ended, written responses; 887 responses were received. 338 respondents volunteered to be interviewed; six (four women and two men) were selected for additional feedback. A number of cross-cutting themes emerged, in addition to expected differences across fields, research teams, countries, and seniority.

The first was the importance of team structure and the role of voice: researchers felt that they had to speak up for themselves to be included, and if they are unaware or too unsure of themselves to speak up, they will miss out. As one woman respondent said "I did not push to be listed as an author." Another woman respondent noted that "Being a woman [means] that quite often you contribute in one way or another to science but unless you shout or make a strong point, our contributions are often underestimated." Multiple respondents mentioned that lack of voice could disproportionately affect women, minorities, and foreign-born scientists. However, respondents also noted that speaking up could backfire as well; "Senior authors shamed me in front of group for asking for recognition (trying not to be a female-doormat stereotype backfires pretty much every time I have tried...)."

The second was a lack of clarity with respect to authorship rules, which reinforces organizational structure. Rules are often determined by senior researchers (who are disproportionately men), and are often governed by personal relationships and idiosyncratic preferences, which reportedly led to disagreements. In at least two interviews, and in many of the survey responses, the disagreements were extremely bitter; the open-ended responses included such statements as "Favoritism, narcissisms, power-play" (W); "The team backstabbed me" (W); "[…] this lack of credit from my PI to be childish and unprofessional" (M). Power imbalance were also frequently mentioned; "Publications were used as reward and punishment. The department heads were on everything…[everything] was dependent on their decision on authorship. It was difficult to get away from them as it was a way to keep people tied to them" (W).

Finally, interviewees and survey respondents were keenly aware of the importance of scientific output as a signal of quality. They felt that being left off papers had significant negative long-term consequences. Some felt that not getting credit had damaged their career; "My career would have been quite different with these two *Nature* papers" (W); "Being left off papers for which I was one of the two main leads has greatly damaged my career as a researcher and my chance to get promotion, jobs, and grant funding. I am still an academic but in a teaching role" (W); "Authorship is pivotal for career advancement, yet when trainees are excluded from authorship due to senior author decisions, there is no appeal or challenge process…Most of my fellow academics (especially women, and most especially women of color) have been harmed by faculty who decide to award authorship to other lab members who did not do the work" (W). Others were still traumatized by the experience; "It was a very tough experience and I am relieved it didn't happen earlier in my career because that would have been devastating" (W); "I'm still very angry over this disgusting behavior" (W); "[it was] one of the lowest points of my professional career" (M).

Discussion

The key finding of this work is that, regardless of the measure of scientific credit, and despite efforts to standardize credit [16], women are significantly much less likely to be credited with authorship than are men. The results are robust to variety of different checks described in detail in Supplementary Information Part 1, namely (i) differential accuracy of gender imputation for non-English and Asian names; (ii) differential match quality because of name changes and frequency; (iii) the definition of potential authors, including first and last authorship; (iv) differences by type of research output and the timing of research output relative to employment; (v) heterogeneity across fields; (vi) sample construction; (vii) definition of time working in labs; (viii) logistic model; and (ix) combinations of robustness checks. Thus, some of the well-documented "productivity" gap [1-5] may not be a gap in the contribution of women to science at all, but rather a gap in how much their contributions are recognized. The associated qualitative work suggests that the standards determining scientific attribution are not well-known or understood by all parties and are frequently disregarded: the result appears to be that women are systematically disadvantaged. While we focus here on gender, these gaps were also reported in our survey for other marginalized groups.

The evidence presented here is consistent with the notion that gender differences in science may be self-reinforcing: that the fate experienced by Rosalind Franklin and others like her discouraged numerous potentially high-impact researchers from entering science [21]. The under-representation of women in faculty positions may be the result of early discouragement among junior researchers, if women are less likely to be recognized for their contributions—especially on pivotal projects—and are consequently less likely to advance in their careers. Longitudinal work on the progress of women's careers [22], could be furthered by studying these data which could provide an empirical link between credit attribution, women's career progression, and discouragement of early stage researchers.

There are important caveats as well. Each data source has its drawbacks. The administrative data are drawn from research intensive universities. As such, the research experiences described using the administrative data may not represent the research experiences for all teams, and, to the extent that women may be under-represented in research intensive universities, may not represent the experiences of all women. Similarly, while the survey data are drawn from a broader sample, they are drawn from a sample of authors, so they do not capture the experiences of those who never authored.

Much more can be done to unpack the findings in other dimensions, such as the mechanisms whereby credit for scientific work is allocated, other dimensions of identity, and richer (e.g., non-binary and fluid) measures of gender. Although we made every effort to be aware of and to guard against confirmation bias [23, 24] by including a variety of robustness checks in the quantitative analysis, by working with survey methodologists to review the survey to ensure that the questions were not leading to a "desired" answer [25], and by developing an interview protocol that did not introduce any discussion of gender (see Supplemental Information Section 3), we encourage other researchers to work with the code and data that are available at IRIS to extend our analyses. Indeed, the unique data infrastructure highlighted in this work can be and is being expanded [17] by the addition of new universities and linkages to many different data sources. As such, it can be used by many other researchers to allow more examination of the organization of science – ranging from rich and complex data on the dynamic longitudinal interactions on what is funded (grants), who is funded (Principal Investigators), and the characteristics of the individuals and the research teams who are employed by those funds. It will also be possible in future work to examine the effect of policies instituted by the research institutions within which researchers work (at the department, campus and university level [26]) on the retention and productivity of scientists [27], student placements and career trajectories [28-30], as well as business startups [31].

In sum, and beyond the results presented here, this paper serves as the introduction to a new and rich data infrastructure that is at least as rich as the bibliometrics data infrastructure that has served as the evidence basis for the study of the science of science [32]. The infrastructure, which is currently being used by over 200 researchers can be and has been replicated in other countries [33] and provides new insights into the organization of science.

Main text references

1. Huang, J., et al., *Historical comparison of gender inequality in scientific careers across countries and disciplines.* Proceedings of the National Academy of Sciences, 2020. **117**(9): p. 4609-4616.
2. Tomaskovic-Devey, D., *Women's Work: Gender Equality vs. Hierarchy in the Life Sciences.* Administrative Science Quarterly, 2005. **50**(4): p. 661-662.
3. Whittington, K.B. and L. Smith-Doerr, *Gender and commercial science: Women's patenting in the life sciences.* The Journal of Technology Transfer, 2005. **30**(4): p. 355-370.
4. Bostwick, V.K. and B.A. Weinberg, *Nevertheless She Persisted? Gender Peer Effects in Doctoral STEM Programs*. 2018.

5. Mairesse, J. and M. Pezzoni, *Does Gender Affect Scientific Productivity?* Revue économique, 2015. **66**(1): p. 65-113.
6. Klug, A., *Rosalind Franklin and the discovery of the structure of DNA.* Nature, 1968. **219**(5156): p. 808-810.
7. Isaacson, W., *The Code Breaker: Jennifer Doudna, Gene Editing, and the Future of the Human Race*. 2021: Simon and Schuster.
8. Bostwick, V.K. and B.A. Weinberg, *Nevertheless she persisted? Gender peer effects in doctoral STEM programs.* Journal of Labor Economics, 2022. **40**(2): p. 397-436.
9. Hofstra, B., et al., *The diversity–innovation paradox in science.* Proceedings of the National Academy of Sciences, 2020. **117**(17): p. 9284-9291.
10. Watson, J.D., *The double helix: a personal account of the discovery of the structure of DNA.* 1968.
11. Sayre, A., *Rosalind Franklin and dna*. 2000: WW Norton & Company.
12. Council, N.R., *Gender differences at critical transitions in the careers of science, engineering, and mathematics faculty*. 2010: National Academies Press.
13. Ellenberg, J., *How not to be wrong: The power of mathematical thinking*. 2015: Penguin.
14. Heckman, J.J., *The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models*, in *Annals of economic and social measurement, volume 5, number 4*. 1976, NBER. p. 475-492.
15. Geddes, B., *How the cases you choose affect the answers you get: Selection bias in comparative politics.* Political analysis, 1990. **2**: p. 131-150.
16. Conti, A. and C.C. Liu, *Bringing the lab back in: Personnel composition and scientific output at the MIT Department of Biology.* Research Policy, 2015.
17. Chang, W.-Y., et al., *A Linked Data Mosaic for Policy Relevant Research on Science.* Harvard Data Science Review, 2022. **4:2**.
18. Meyer, B.D., W.K. Mok, and J.X. Sullivan, *Household surveys in crisis.* Journal of Economic Perspectives, 2015. **29**(4): p. 199-226.
19. Aghaei Chadegani, A., et al., *A comparison between two main academic literature collections: Web of Science and Scopus databases.* Asian social science, 2013. **9**(5): p. 18-26.
20. Allen, L., et al., *Publishing: Credit where credit is due.* Nature, 2014. **508**(7496): p. 312-313.
21. Kemelgor, C. and H. Etzkowitz, *Overcoming isolation: Women's dilemmas in American academic science.* Minerva, 2001. **39**(2): p. 153-174.
22. Leahey, E., J.L. Crockett, and L.A. Hunter, *Gendered academic careers: Specializing for success?* Social forces, 2008. **86**(3): p. 1273-1309.
23. Klayman, J. and Y.-W. Ha, *Confirmation, disconfirmation, and information in hypothesis testing.* Psychological review, 1987. **94**(2): p. 211.
24. Lehner, P.E., et al., *Confirmation bias in complex analyses.* IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2008. **38**(3): p. 584-592.
25. Kreuter, F., S. Presser, and R. Tourangeau, *Social desirability bias in cati, ivr, and web surveysthe effects of mode and question sensitivity.* Public opinion quarterly, 2008. **72**(5): p. 847-865.
26. Owen-Smith, J., *Research Universities and the Public Good: Discovery for an Uncertain Future*. 2018: Stanford University Press.
27. Funk, R.J. and J. Owen-Smith, *A dynamic network measure of technological change.* Management Science, 2017. **63**(3): p. 791-817.
28. Zolas, N., et al., *Wrapping it up in a person: Examining employment and earnings outcomes for Ph.D. recipients.* Science, 2015. **350**(6266): p. 1367-1371.
29. Zolas, N., et al., *Wrapping it up in a person: Tracing flows from funded research into the economy.* Science, 2015.
30. Buffington, C., et al., *STEM Training and Early Career Outcomes of Female and Male Graduate Students: Evidence from UMETRICS Data linked to the 2010 Census.* American Economic

Review, 2016.
31. Goldschlag, N., et al., *The Link Between University R&D, Human Capital and Business Startups*, C. Corrado, et al., Editors. 2018, University of Chicago Press.
32. Lane, J., et al., *New linked data on science investments, the scientific workforce and the economic and scientific results of science*. 2014.
33. Chevalier, G., et al., *A new approach for estimating research impact: An application to French cancer research.* Quantitative Science Studies, 2020. **1**(4): p. 1586-1600.

Tables

Table 1: Gender differences in position and "ever authorship"

Job Title	Frequency of job titles in the full sample			Likelihood of ever receiving attribution		
	Total	Women	Men	Total	Women	Men
Faculty	14.85%	11.30%	19.72%	45.70%	41.25%	48.86%
Postdocs	8.63%	6.00%	9.08%	25.17%	22.35%	27.31%
Graduate Students	24.15%	17.42%	25.06%	18.69%	14.97%	21.37%
Research Staff	35.41%	47.81%	28.73%	8.63%	6.59%	11.01%
Undergraduates	16.96%	17.48%	17.42%	2.61%	2.22%	3.10%
Total / Average	100%	100%	100%	16.97%	12.15%	21.17%

Notes: This table provides descriptive statistics that show the percentage of employees who worked in university teams between 2013 and 2016 published papers between 2014 and 2016, and who appeared on at least one of those publications as an author. The percentages are computed over the128,859 unique employees in the dataset. The totals include men, women, and those whose gender was not imputed. See Extended Data Table 1 and the section on the Construction of Administrative Data in the Methods Section.

Figure legends

Figure 1: Women are less likely to be named authors in all career stages and in all fields
Note: Each panel plots the probability that a potential author on a scientific document (article or patent) is a woman (x-axis) against the probability actual authors are women (y-axis). A potential authorship is defined as an employee in a lab between 2013 and 2016 and a publication or patent produced by that lab between 2014 and 2016. There are 17,929,271 potential article authorships and 3,203,831 potential patent inventorships in our sample. The markers in each panel are sized by the total number of actual authorships in the category. The 45-degree line represents parity in the gender composition of potential and actual authorships. Individual data on potential and actual authorships are visualized in SI fig 5. Panel (a) reports the disparity across job titles; panel (b) reports the disparity across fields. The observations are weighted by the inverse number of teams per employee times the inverse number of potential articles or patents per employee.

Figure 2: Women are still less likely to be named even when controls are included
Note: The figure shows the probability that an individual in a team is an author on any given article (Panel a) or patent (Panel b) from that team. Panel A is estimated using 17,929,271

observations (potential article authorships). Panel B is estimated using 3,203,831 observations (potential patent inventorships). All probabilities are obtained from ordinary least squares regressions of being named on gender and the indicated controls (reported in Table 4). For the purpose of plotting probabilities and gender differences holding all else fixed (Δ), we hold all of the controls at their respective means. Because on average men have higher values than women on the controls that increase the probability of attribution, as more controls are included, the predicted probabilities for men decline and those for women increase. Specification 1 includes no controls while specification 2 controls for whether a potential author is the PI of the team, the days worked on the team, and indicators for the publication date (calendar year x month). Specification 3 also includes controls for the job title of the potential author. Specification 4 adds field controls for the team, and specification 5 adds individual indicator variables for each team (these team indicators subsume the fields indicator). The observations are weighted by the inverse number of teams per employee times the inverse number of potential articles or patents per employee. The data are represented as the arithmetic mean of the indicator variable representing if the individual was ever an author (Panel a) or ever an inventor (Panel b) on any given article or patent. Individual data on the probability of women or men being named to articles or patents are visualized in SI fig 6. The error bars extend from that mean to create a 95% confidence interval based on 1.96 times the standard error of the mean. Standard errors are clustered by team and employee.

Figure 3: Women are much less likely to be named on high impact articles
Note: The figure shows the probability that an individual in a team is an author on any given article (Panel a) or patent (Panel b) in relation to the number of citations to the document receives. These estimates were obtained from an ordinary least squares regression of being named on an indicator for gender interacted with the log of total forward citations plus one (reported in Table 6). The regression used in Panel A is estimated on 17,929,271 potential article authorships. The regression used in Panel B is estimated on 3,203,831 potential patent inventorships. The observations are weighted by the inverse number of teams per employee times the inverse number of potential articles or patents per employee. Estimates include controls for publication date (calendar year x month), PI status, days worked on the team, job title, and research team fixed effects. Each data point represents the estimated marginal effect of the number of citations on the predicted probability of being named an author (Panel a) or inventor (Panel b). The error bars extend from the point estimate of the estimate marginal effect by +/- 1.96 times the standard error and display the 95% confidence interval of the marginal effect. Standard errors are clustered by team and employee.

Figure 4: Women are more likely to report that their contributions were underestimated or that there was discrimination
Note: The figure shows the percentage of men and women respectively who responded to the question (Q2b) "What is the most likely reason that you were not listed as an author on that paper?" by selecting a given category. The survey was sent to 28,000 scientists who had published in an academic journal listed in the Web of Science and who listed themselves with a public profile on the ORCID database. The figure summarizes the responses of the 871 scientists who responded to our survey and completed this question. Respondents were able to select multiple options, therefore the responses sum to more than 100%. The probability is computed as the arithmetic mean of the binary responses. The data are represented for each category as the

arithmetic mean of the indicator variable corresponding to a respondent selecting that category. Individual data on the reason an individual is not named are visualized in SI fig 7. The error bars are centered on that mean and extend on each side to create a 95% confidence interval based on 1.96 times the standard error of the mean. The difference in the probability of indicating "Contribution did not justify authorship" between men and women who respond is 0.1294 (p=0.0000; two-sided t-test value=4.1060). The difference in the probability of indicating "Others underestimated my contributions" between men and women who respond is -0.0984 (p=0.0036; two-sided t-test value=-2.9218). The difference in the probability of indicating "Discrimination/stereotyping/bias" between men and women is -0.0780 (p=0.0003; two-sided t-test value= -3.6623). Additional t-tests of the differences in the probability of indicating a reason across men and women who respond can be found in the main text.

Figure 5: Women report making more contributions than men on authored papers

Note: We sent a survey to 28,000 scientists who had published in an academic journal listed in the *Web of Science* and who had a public profile in the ORCID database. Of those sent the survey, 2,297 scientists responded and completed the question, (Q1a) "How did you contribute to the paper? Check all that apply." The above figure shows the percentage of these respondents who selected each category. The probability is computed as the arithmetic mean of binary indicators representing if the respondent selected each category. Each respondent was asked about a paper associated with them on *Web of Science*. Respondents were able to select multiple options, therefore the responses sum to more than 100%. The data are represented for each category as the arithmetic mean of the indicator variable corresponding to a respondent selecting that category. Individual data on the contribution by gender are visualized in SI fig 8. The error bars are centered on that mean and extend on each side to create a 95% confidence interval based on 1.96 times the standard error of the mean.

Methods

This section has four parts. The first describes the data construction and variable operationalization used in the analysis of administrative data; the second describes the analysis of the administration data; the third, the construction of the survey data; and the fourth, the qualitative responses and interviews.

Construction of Administrative Data

The analytical linked dataset, which consists of observations on 128,859 individuals employed on 9,778 research teams from 2013-16 linked to 39,426 subsequent articles and 7,675 patents, is constructed from three sources: internal Finance and Human Resources administrative data from 20 universities and 57 colleges and campuses [34], representing over 40% of total academic R&D spending in the United States, journal articles from the Web of Science and patent data derived from the universe of patents from the US Patent and Trademark Office (USPTO).

Finance and Human Resources data

The first source is derived from Financial and Human resources (FHR) data, called UMETRICS, on all personnel paid on sponsored research projects for 106 college and campuses from 33 universities from 2001 and 2022 (the exact years covered vary by institution) [34]. (A full list of participating institutions, which are primarily research-intensive, can be found at https://iris.isr.umich.edu/.)
Each pay period, the FHR system at each university records the details of charges to each sponsored project, including for each person paid on each grant and reports the information to the Institute for Research on Innovation and Science [35]. These administrative data are different from the level-of-effort data that are submitted by PIs as part of their annual and final report to an agency in at least three ways. First, they represent actual payroll data, drawn from the FHR system every pay-period, rather than the estimate provided by the Principal Investigator (PI) or the team administrator once a year. An intensive hand curated effort that compared the results from an early effort found that the FHR reports are more granular and comprehensive than the PI or team administrator reports [36, 37]. For example, all personnel names (including co-PIs) are recorded in the FHR reports, but many names are not recorded in the former. Second, the UMETRICS data capture all sources of funding, and are much more comprehensive than data from a single agency. The UMETRICS data include federal funding sources as well as funding from philanthropic foundations, state and local governments, industry, and international organizations. Third, the data reflect actual expenditures in every accounting time period, not just funds that are obligated at the beginning of a grant. So, if, as often happens, there is a no-cost extension, or more funds are spent earlier in the project, that spending and the work of the relevant team members, is captured in the data. There are limitations. If personnel do not charge time to the grant, their effort is not captured in the data – we are unaware of any source that would capture unpaid work. If there are gender differences in unpaid research work, the analysis will not be able to capture such differences.

The analysis focuses on a subset of 57 college campuses from 20 universities which consistently provided data for the period covering 2013-16 (see pages 10-11 and 23 of the UMETRICS summary documentation [34]; SI Part 2). This restriction ensures that employment spells are long enough to reasonably identify PIs and teams as well as to observe the scientific documents produced by those teams from 2014-16. The full data include administrative level information from 392,125 unique federal and non-federal awards, including 23,307,254 wage payments to 643,463 deidentified individuals [26].

Research teams

The construction of research teams was informed by the work of Stephan[28] who operationalized the concept of a research team to be a collection of scientists working jointly on projects with common funding and resources. The UMETRICS data are ideally suited to create measures of teams at scale using this definition, because the administrative data provide detailed information of all people charging time to each grant in each payroll period [29, 30].
The composition of each team is constructed as follows. The PI is at the center of each team. The Pis in the data are identified by selecting faculty members who have been continuously paid on at least one research grant per year from 2013-16 and whose associated wage payments always list faculty as their job title. The PI-associated grants are identified if at least one wage payment was made to the PI during the sample period and shared evenly if they involve multiple Pis. Research center grants, which are characterized as grants with 12 or more faculty – the 99th percentile of the grants, were excluded. Based on the grants associated with the PI, we identify the set of graduate students, post-docs, research staff, undergraduates, and non-PI faculty who are paid on those grants. The set of scientists paid on the grants associated with the PI collectively make up the research team. This procedure yields a total of 9,778 teams, with 128,859 employees, between 2013 and 2016.
The number of teams and potential authorships varies considerably across people in the sample. To ensure that our estimates are not dominated by people who are on many teams or on teams with many articles, we weight our data so that each person receives equal weight and for each person, each team receives equal weight. If $N_{Teams,i}$ denotes the number of teams that person i is on and $N_{PA,t}$denotes the number of potential authorships (i.e., articles and/or patents) on team t, then the weight applied to person i's potential authorships on team t is $\frac{1}{N_{Teams,i} \cdot N_{PA,t}}$. Thus, each person is weighted by the inverse of the number of teams on which he or she appears times the inverse number of potential authorships for that team. Each unique employee therefore has an overall weight of one in the sample. Our results, however, are robust to various alternative weightings.

Gender

Gender is algorithmically assigned using a combination of Ethnea [30] and Python's Gender Guesser algorithms. Ethnea is first used to assign gender based on the first names and ethnicity (algorithmically assigned from the family name) of each employee. When the first name gives ambiguous results, the middle name is used. If gender is still ambiguous, Python's Gender Guesser is applied to the individual's first name, but not the middle name. Gender can be identified for 107,239 (83.2% of sample), of whom 51,738 are women and 55,502 are men.
The accuracy of the imputation was tested against two sources of ground truth. The first source is self-reported, administrative data on gender for 12,867 faculty from one institution participating

in UMETRICS. The algorithm correctly predicts the self-reported gender in 93% of the cases: the precision is 93.35% for men and 92.51% for women. The second source is derived from a match of the UMETRICS data with the Survey of Earned Doctorates (SED) [38]. The Survey of Earned Doctorates, which is an annual survey (with a 93% response rate) of all doctorate graduates from U.S. universities, directly asks respondent to report their gender. The precision of the algorithm was 97.29% for men and 94.06% for women. Robustness checks are reported in the Supplementary materials. We note that a limitation is that our gender construct does not allow for non-binary or fluid gender identities. Addressing non-binary and/or fluid gender identities is an important direction for future research.

Job Titles

Job titles for each employee, which are also referred to in the text as positions, or roles, are constructed from the FHR records [39]. Some employees may hold different job titles on the same or different teams; in those instances, the title is equally weighted based on the number of days they were paid in each title within each team.

Scientific fields

The scientific field of each team is identified by using the title of all associated grants and comparing the grants with a pool of text that describes each scientific field using a "Wiki-labeling" approach [15, 16]. This approach is used to assign a likelihood score that a given grant award title belongs to a given research field category, as categorized by the NCSES Survey of Graduate Students and Postdoctorates in Science and Engineering. Each team's field is estimated by taking the field of each grant and weighting the grant's relative importance to the team's portfolio by the direct expenditure of each grant over the analysis period.

Publications

Publications are drawn from the *Web of Science (WoS)* database produced and maintained by Clarivate Analytics, which contains publication and citation information on approximately 69.3 million total articles from 1900 to 2018. The analysis focuses on articles published from 2014-16 and linked to individuals observed in UMETRICS from 2013-16, although we include some additional robustness checks on other year ranges and other publication types in the supplementary online materials.

Patents

Patents are drawn from the *PatentsView (PV)* visualization and analysis platform, which contains 6.8 million total patents from 1976 to 2018 [40]. The analysis focuses on a subset of patents that has an application date from 2014-16 and is linked to individuals observed in UMETRICS from 2013-16. Additional robustness checks on other year ranges are include in the supplementary online materials.

Linked Administrative Records

The links between UMETRICS and authorship on articles and patents were generated by combining information on the individual and grants listed explicitly on the scientific documents as well as the implicit network structure of co-authorships and grant collaborations. In UMETRICS, the data includes the individual's name (including partial name in the case of hyphenated names), the institution and the grant number but, crucially, also other people on each grant. The same is the case in the publication and patent data. We identify all patents/articles associated with a given inventor/author by leveraging PatentsView's algorithmically assigned Inventor ID and the union of the Web-of-Science's Researcher ID and the ORCID when they are available. Key to our approach, these identity clusters allow us to link a given inventor or author's full patent and publication history to an individual's employee ID in UMETRICS such that we not only see those documents associated with a specific set of grants or a particular time period, but their entire patenting and publishing history over their career. The multi-step procedure, which uses data post 2000, is detailed in Ross et al. [41]. There are five steps. The first relies on an exact match of UMETRICS award numbers to either the award numbers cited in the government interest field in the patents or the award numbers cited in the acknowledgement section of the publication. The second step relies on grant matches. It links inventors in Patentsview and authors in Web-of-Science to people paid on UMETRICS grants using a sequential process of exact and fuzzy matching, with matched names removed from the pool for subsequent rounds. Candidate matches are disqualified for mismatches on institutional affiliation and dissimilarity of text between awards and publications and patents. The third step relies on network matches. It uses exact and fuzzy name matching to find coinventors (in Patentsview), coauthors in Web-of-Science) and collaborators (in UMETRICS). Candidate matches are disqualified for mismatches on institutional affiliation and dissimilarity of text between awards and publications and patents. The fourth step links people by blocked affiliations. Affiliation names are matched by blocking on the UMETRICS university affiliation to the affiliations in PV and WoS (using a hand curated, disambiguated list of university names), and using the stepwise matching and validation processes described in the second step. As before, candidate matches are disqualified for mismatches on institutional affiliation and dissimilarity of text between awards and publications and patents. The fifth and final step relies on an approximate match of unmatched grants. It uses the pool of articles/patents associated with the identity clusters linked in steps 2-4 (namely, employees in UMETRICS linked with their associated inventor IDs and research ID or ORCID). The restriction that grant numbers on these documents be deterministically matched is loosened, and a fuzzy match is allowed between grants in UMETRICS and those unmatched in Step 1 but associated with linked individuals.

<u>Analytical Sample</u>

All publications and patents that acknowledge one of the team's grants and/or has an author/inventor from the team are linked to the team. This results in a total of 47,101 scientific documents (39,426 articles and 7,675 patents) published between 2014-16 which were linked to employees and teams observed in UMETRICS at any point in the prior year, i.e., from 2013-16. Summary information about the individuals and the teams is provided in Extended Data Table 1). Additional information about the differences between authors and non-authors in the sample as well as some basic descriptive information surrounding grant funding sources is provided in Part 2 of the Supplementary Information.

The resulting linkages permit the calculation of the overall "ever author" rate, which is 16.97% overall, but only 12.15% for women and 21.17% for men (Extended Data Table 2). The "attribution rate" is constructed by generating a pool of "potential authorships" as follows. All individuals with a faculty job title are considered eligible to be potential authors on all articles or patents produced by a team during the analysis period. All individuals with a non-faculty title had to have been employed by the team in the year prior to the article of an article or application for a patent. We relax this time constraint for non-faculty job titles in the supplement which generally increases the size of the gender gap reported in the main estimates.

The resulting analytical dataset consists of 21,133,102 potential authorship observations (17,929,271 on articles and 3,203,831 on patents) of which 367,231 were actual authorships. 48.2% of potential authorships were by women, while 31.8% of actual authorships were by women. If these numbers are converted to rates, the attribution rate on scientific documents was 3.17%. The attribution rate for articles alone is 3.2% while it is 1.3% for patents (Note to Extended Data Table 2). Although both of these attribution rates are relatively low, this is largely due to the inclusion of undergraduate students and research staff in our sample as well as those observed working for short time periods. These employees are rarely observed in the actual authorships and result in a lower the overall attribution rate. The regression analyses reported in the subsequent sections control for both position and the number of days worked in the team; Part 1 of the Supplementary Information provides results excluding undergraduates and research staff. The results are robust in each specification.

The third attribution measure - the impact of scientific articles and patents – is constructed by attaching forward citations (as of 2018) reported in the WoS and PV datasets to the potential authorship sample. Because earlier documents in the sample (e.g., those from 2014) have more time to receive citations than later documents (e.g., those from 2016), we include publication date (calendar year x month) controls, as in our other models.

Effect sizes are calculated as the percentage point differences between the contrasted groups unless otherwise noted in the text.

Empirical Strategy

The empirical approach was to estimate linear regressions using a model of the form

$$P[named_{i,t,e,l}|\ldots = \beta_0 + \beta_1\, woman_{i,e} + X_{i,e} + M_{i,t} + O_{i,e} + Team_{i,l} + \mu_{i,t,e,l} \qquad (1)$$

where i potential authorship observations are characterized by an employee e working on team l in the year prior to a document with a publication or application date t (calendar year x months). The primary variable of interest, $woman_{i,e}$, is an indicator of whether a potential authorship was attributable to an employee who was a woman. Equation 1 is estimated on the sample of 17,929,271 potential authorships on journal articles while the patent results are estimated on the sample of 3,203,831 potential authorships.

A series of regressions were estimated. The first set (Extended Data Table 3) included controls, $X_{i,e}$, which sequentially include indicator variables for the publication/application month associated with a potential authorship/inventorship, the team's Principal Investigator, the number of days worked in the team, and an indicator of whether the individual's gender was unknown. Idiosyncratic trends in the data are accounted for by including a series of $M_{i,t}$ calendar year x months and year fixed effects based on the date when article i was published or patent i applied for; an individual's position in the team is accounted for through a series of $O_{i,e}$ position variables which capture the days that an individual worked in a particular position as a share of the total days worked on the research team. Differences across research teams are accounted by including a series of $Team_{i,l}$ team fixed effects and we denote the disturbances in the data using $\mu_{i,t,e,l}$. The second set (Extended Data Table 4) re-estimated Equation 1 with the same controls but by job title; the third set (Extended Data Table 5) re-estimated the same equation with the same controls by field. The final set (Extended Data Table 6) restricted the sample to high impact publications and patents.

Survey Design and Collection

The survey was sent to individuals who had previously published in academic research journals identified through their public profiles on ORCID, a platform in which academic researchers post their educational credentials, work history, and publication records. Information on the survey instrument, email recruitment, and interview protocols is available in Part 3 of the Supplementary Information.

The main database was the ORCID 2017 database, which includes the publicly viewable information from profiles shown on the ORCID website as they appeared in 2017: 897,264 profiles listed a complete name as well as educational credentials, work history information, or both.

Email addresses associated with the researchers of these profile were then derived from those emails listed on published and publicly available research articles which were available from the *Web of Science*. *Web of Science* also provides the associated email addresses for 128,602 of the 897,264 ORCID profiles. Because the focus was on asking academic researchers about their experience with being named or not being named as coauthors on publications, the ORCID profiles were restricted to those that could be linked with a published academic paper in the *Web of Science* database between 2014 and 2018: 98,134 profiles fulfilled those criteria.

Finally, some individuals create multiple ORCID profiles and some email addresses are recycled for multiple people over time. In order to avoid emailing the same individual multiple times, each email had only one associated ORCID profile. After resolving duplicates, there were 98,022 unique ORCID profiles that matched our sample criteria.

Three studies were piloted before the main study. After imputing the gender of the individuals represented by the ORCID profiles using first names and the Ethnea database, 10,000 (imputed) ORCID profiles belonging to men and 10,000 (imputed) ORCID profiles belonging to women were randomly selected to receive the survey in addition to 6,500 profiles that had gender ambiguous names.

Qualitative Evidence

In addition to the open-ended text in which researchers could record their experiences, the last question of the survey solicited researchers “to interview over Zoom regarding their experiences with the allocation of credit in research teams.” Respondents were told that, if they were interested in talking about their experiences with the allocation of scientific credit on teams, they could enter their email addresses to be contacted for a follow-up interview. A team of two authors (of both genders for three interviews, and of one gender for three interviews) of this paper interviewed six individuals for half an hour each. Four were women, two were men. Gender was never raised as an issue by the team but was raised by the interviewees. The detailed interview protocol is available in Supplemental Information Part 3.

Methods References

34. Insitute for Research on Innovation and Science, *Summary Documentation for the IRIS UMETRICS 2019 Data Release*. 2019, Summary Documentation for the IRIS UMETRICS 2019 Data Release.
35. IRIS, *Summary Documentation for the IRIS UMETRICS 2020 Data Release*, T.I.f.R.o.I.S. (IRIS), Editor. 2021.
36. Lane, J. and L. Schwarz, *Creating new administrative data to describe the scientific workforce: The STAR METRICS program*, in *Social Science Research Network*. 2012: https://ssrn.com/abstract=2085187
37. Nelson, L. and S.W. Sedwick, *STAR METRICS: A Participant's Perspective.* NCURA Mag, 2011. **43**: p. 24-25.
38. National Science Foundation, *Survey of Earned Doctorates*, National Center of Science and Engineering Statistics, Editor. 2013-2016.
39. Ikudo, A., et al., *Occupational Classifications: A Machine Learning Approach.* Journal of Economic and Social Measurement, 2020.
40. Institute For Research On, I. and R. Science, *UMETRICS-Patent Crosswalk and Patent Data Files*. 2018, Institute for Research on Innovation and Science (IRIS).
41. Ross, M., et al., *"A Network Approach to Bibliometric Record Linkage with Semantic Validation"*. 2020, Ohio State University: Columbus, Ohio.

End Notes

Acknowledgments: Funding: NSF grants 2100234, 1932689, 1761008, 1760544; NIA, OBSSR, and NSF SciSIP through P01 AG039347; R01 GM140281, UL1 TR002733, Alfred P. Sloan Foundation, Ewing Marion Kauffman Foundation, Schmidt Futures, Patrick J. McGovern Foundation; Weinberg was supported on P01 AG039347 by the NBER directly and on a subaward from NBER to Ohio State. Research reported in this publication was developed using the Collaborative Archive & Data Research Environment (CADRE) project (http://doi.org/10.26313/rdy8-4w58). Thus, the authors received some support through CADRE's funders. CADRE was developed with support from a National Leadership Grant from the Institute of Museum and Library Services (IMLS; grant number LG-70-18-0202-18), including cost-share from the Big Ten Academic Alliance Library Initiatives (BTAA), Microsoft Research, the Web of Science Group, and academic university libraries (Ohio State University, Michigan State University, Purdue University, University of Michigan, University of Minnesota, Penn State University, University of Iowa and Rutgers University). Assistance: Frauke Kreuter, Ekaterina Levitskaya, Jason Owen Smith, Natsuko Nicholls, Jiusi "Josie" Xiao.

Author Contributions: J.L., B.G. and B.W. conceived the framework, and together with M.R., R.M-G., and E.B. co-wrote the paper. M.R., R. M-G, and E.B. designed and carried out the UMETRICS data analysis. B.G. and R.M-G designed and carried out the survey data collection and analysis.

Competing Interests: The authors have no competing interests.

Additional Information

Supplementary information is available for this paper.

Correspondence and requests for materials should be addressed to Julia Lane (julia.lane@nyu.edu).

Reprints and permissions information is available at www.nature.om/reprints.

Institutional Review Board Approval: University of Pennsylvania Institutional Review Board (IRB Protocol # 850522) approved the survey. University of Pennsylvania Institutional Review Board (IRB Protocol # 850522), Boston University Institutional Review Board (IRB Protocol #6412X) the New York University Institutional Review Board (IRB Protocol #IRB-FY2022-6243) and the Ohio State University Institutional Review Board (IRB Protocol 2022E0133) approved the followup interviews.

Data availability

The datasets generated during and/or analysed during the current study are available, as well as the associated code, at the Virtual Data Enclave repository at the Institute for Research on Innovation and Science at the University of Michigan. Access information is provided here https://iris.isr.umich.edu/research-data/access/. Patent data were obtained from Patentsview (https://patentsview.org/), which is publicly available. Web of Science data were obtained from CADRE at Indiana University (https://iuni.iu.edu/resources/datasets/cadre).

The survey data are not available as per the University of Pennsylvania IRB protocols. Aggregate statistics from the survey data can be made available to researchers (upon request) for replication.

Code availability

All the Stata code (Version 17) and Python code (Version 3.7.6) used is available in the Virtual Data Enclave at the University of Michigan. Access information is provided here https://iris.isr.umich.edu/research-data/access/

Extended Data Table Legends

Extended Data Table 1: Team and Individual descriptive statistics

The table reports arithmetic means of the teams and the individuals who worked at least one day in any of the four years from 2013-16. Fields do not sum to 1 because fields are unassigned for about 11.5% of teams. Some employees hold different job titles over the timeframe; in those instances, they are first divided equally across titles within each team and then equally across teams. Shares of men and women do not sum to one because the gender of about 17% of the employees could not be algorithmically assigned.

Extended Data Table 2: Attribution Rates by Job Title and Field

With the exception of two rows, “Total Count of Authors/Authorship” and “Total in Workforce”, the first three columns (a) summarize the share of "ever authors". The numerator is individuals in each category who are ever named as an author on a publication or a patent. The denominator is the total number of individuals in that category. The second three columns (b) summarize the share of authorships: The denominator – the set of "potential authorships" – was created by associating all members of each team who were employed one year prior to the publication/application date of all associated articles/patents emanating from that team during the analysis period. Since some individuals, such as research staff, are on multiple teams, they are proportionately allocated across teams and papers using a set of analytical weights (see Methods Analytical Sample section for details). The numerator - attribution - was defined as "actual authorships" on those publications and patents (see Methods: Analytical Sample for details). The "Total Count of Authors/Authorship" row summarizes the total counts of "ever authors" in the first three columns (a) and the weighted total counts of actual authorships in the second three columns (b). The "Total in Workforce" row summarizes the total counts of individuals in our sample. Note that the "Total" column includes those in our sample who we could not identify as either men or women.

Extended Data Table 3: Gender Difference in Attribution Rate by Job Title and Field

The table summarizes the share of actual and potential authorships that are women. The first column shows the percentage of actual authorships who are women. The second column shows the share of potential authorships who are women. The third column provides the effect size, defined as the difference in percentage points between the share of actual authorships and the share of potential authorships who are women. The fourth column displays the estimated standard error for those differences based on the bootstrapping procedure described in the Methods: Analytical Sample section. The last column provides the two-sided t-test statistic for the effect size being equal to zero using the bootstrap estimated standard error (see Methods: Analytical Sample for details).

Extended Data Table 4: Gender differences in attribution

The sample consists of 17,929,271 potential article authorships and 3,203,831 potential patent inventorships. The top panel is estimated on the sample of potential article authorships and the bottom panel is estimated on the sample of potential patent inventorships. The dependent means are 3.18% and 1.31%, respectively. Specification (1) includes none of the control variables discussed above and estimates the gender gap to be 1.97 and 1.50 percentage points for articles and patents. Specifications (2-5) gradually introduce controls for days worked, PI status, publication month, job title, field, and team (which subsumes field). The observations are weighted by the inverse number of teams per employee times the inverse number of potential articles or patents per employee. Each coefficient is tested against the null hypothesis of being equal to 0 using a two-sided t-test. We do not adjust for multiple hypothesis testing. Standard errors are clustered by team and employee and are in parentheses. Statistical significance indicated by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Extended Data Table 5: Gender differences in attribution by job title

Estimates based on a sample of 17,929,271 potential article authorships and 3,203,831 potential patent inventorships. The observations are weighted by the inverse number of teams per employee times the inverse number of potential articles or patents per employee. All estimates include controls for article/patent date (calendar year x month), PI status, days worked on the team, job title, and team fixed effects. Each coefficient is tested against the null hypothesis of being equal to 0 using a two-sided t-test. We do not adjust for multiple hypothesis testing. Standard errors are clustered by team and employee and are in parentheses. Statistical significance indicated by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

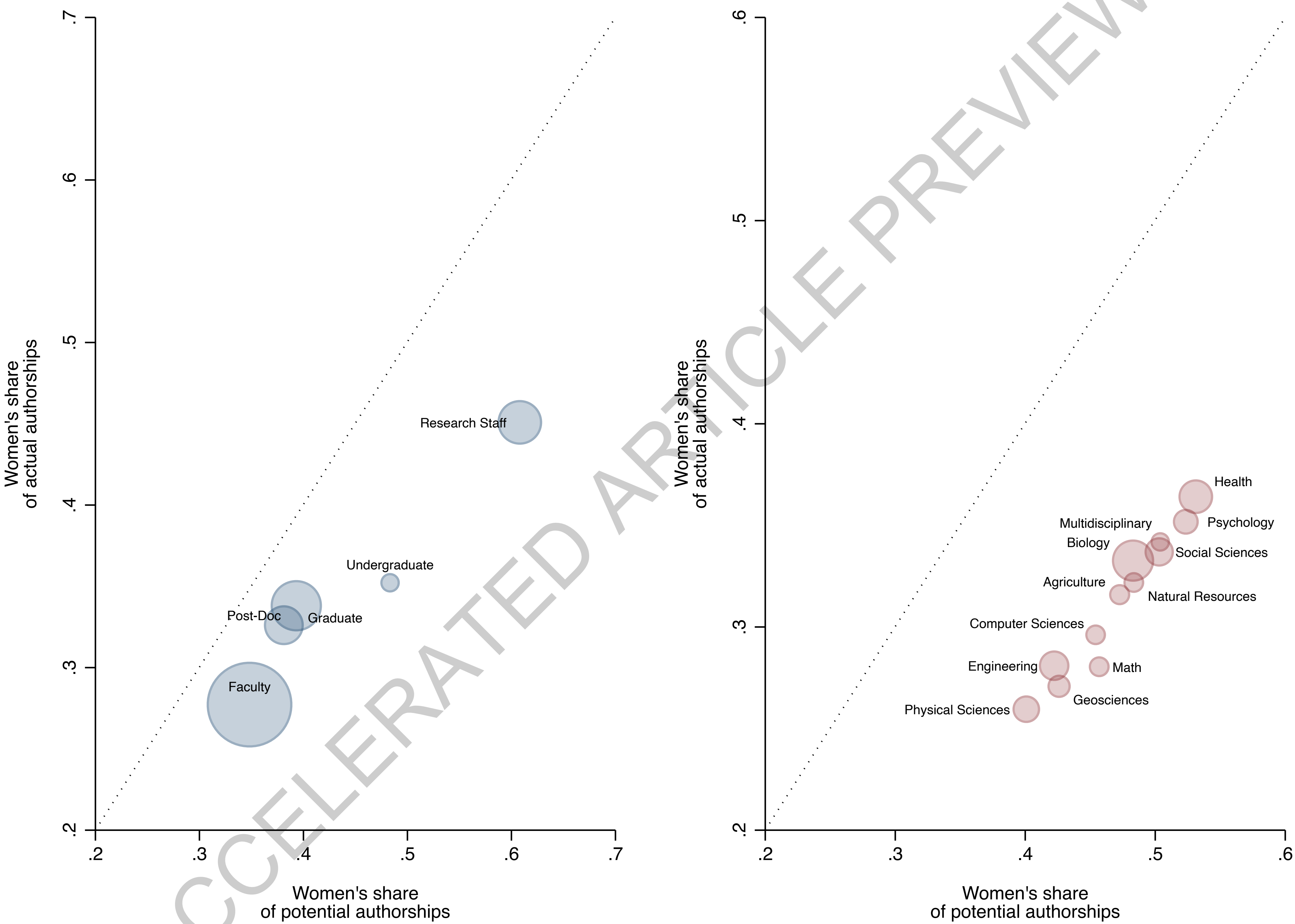
Extended Data Table 6: Gender differences in attribution by field

Estimates based on a sample of 17,929,271 potential article authorships and 3,203,831 potential patent inventorships. The observations are weighted by the inverse number of teams per employee times the inverse number of potential articles or patents per employee. All estimates include controls for article/patent date (calendar year x month), PI status, days worked in the team, job title, and team fixed effects. Each coefficient is tested against the null hypothesis of being equal to 0 using a two-sided t-test. We do not adjust for multiple hypothesis testing. Standard errors are clustered by team and employee and are in parentheses. Statistical significance indicated by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

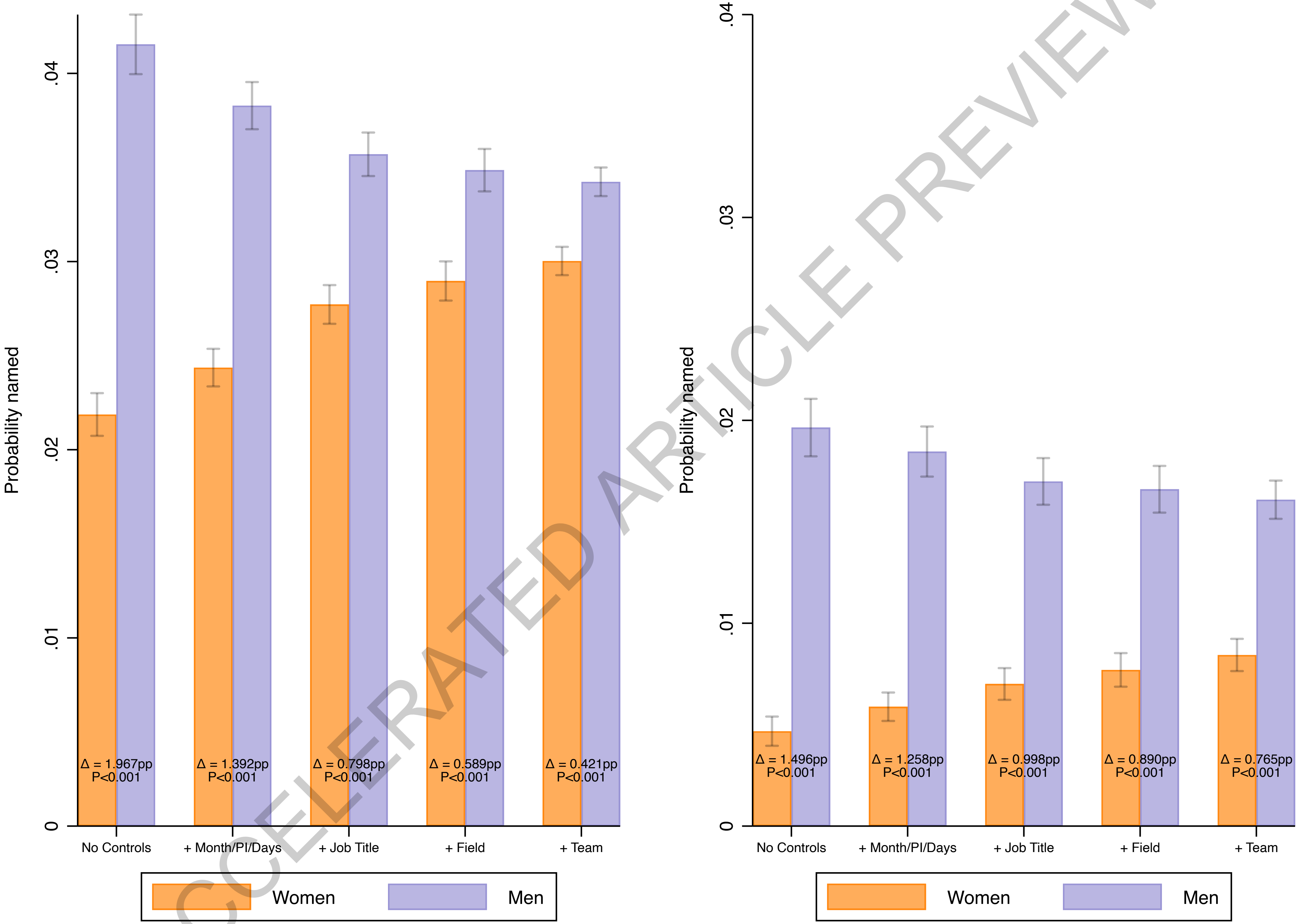
Extended Data Table 7: Gender differences in high impact attribution
Estimates based on a sample of 17,929,271 potential article authorships and 3,203,831 potential patent inventorships. The observations are weighted by the inverse number of teams per employee times the inverse number of potential articles or patents per employee. All estimates include controls for article/patent date (calendar year x month), PI status, days worked in the team, job title, and team fixed effects. Each coefficient is tested against the null hypothesis of being equal to 0 using a two-sided t-test. We do not adjust for multiple hypothesis testing. Standard errors are clustered by team and employee and are in parentheses. Statistical significance indicated by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Extended Data Table 8: Survey Response Rates
This table describes the details of the three pilots and main study. The first column details the date of the first email, while the second column details the date of the reminder email, typically two weeks after the first email. The main study had two reminder emails because on March 1st, an error in the Qualtrics survey system caused an abnormally high number of emails to bounce (~11,000). On March 9th, after the error had been addressed, we re-sent reminder emails to those respondents that had been missed due to the error. Column three describes the sampling strategy; random sampling was used for the pilots, but once we learned that there were far fewer women in the population than there were men, we adjusted to a gender-stratified sampling strategy in order to gain enough power for two-sided t-tests comparing responses from men and women. Specifically, 10,000 (imputed) ORCID profiles belonging to men and 10,000 (imputed) ORCID profiles belonging to women were randomly selected to receive the survey in addition to 6,500 profiles that had gender ambiguous names. Column 4 indicates the number of emails sent, while columns 5-7 document the response rate. A large fraction of emails either bounced or received an automated response.

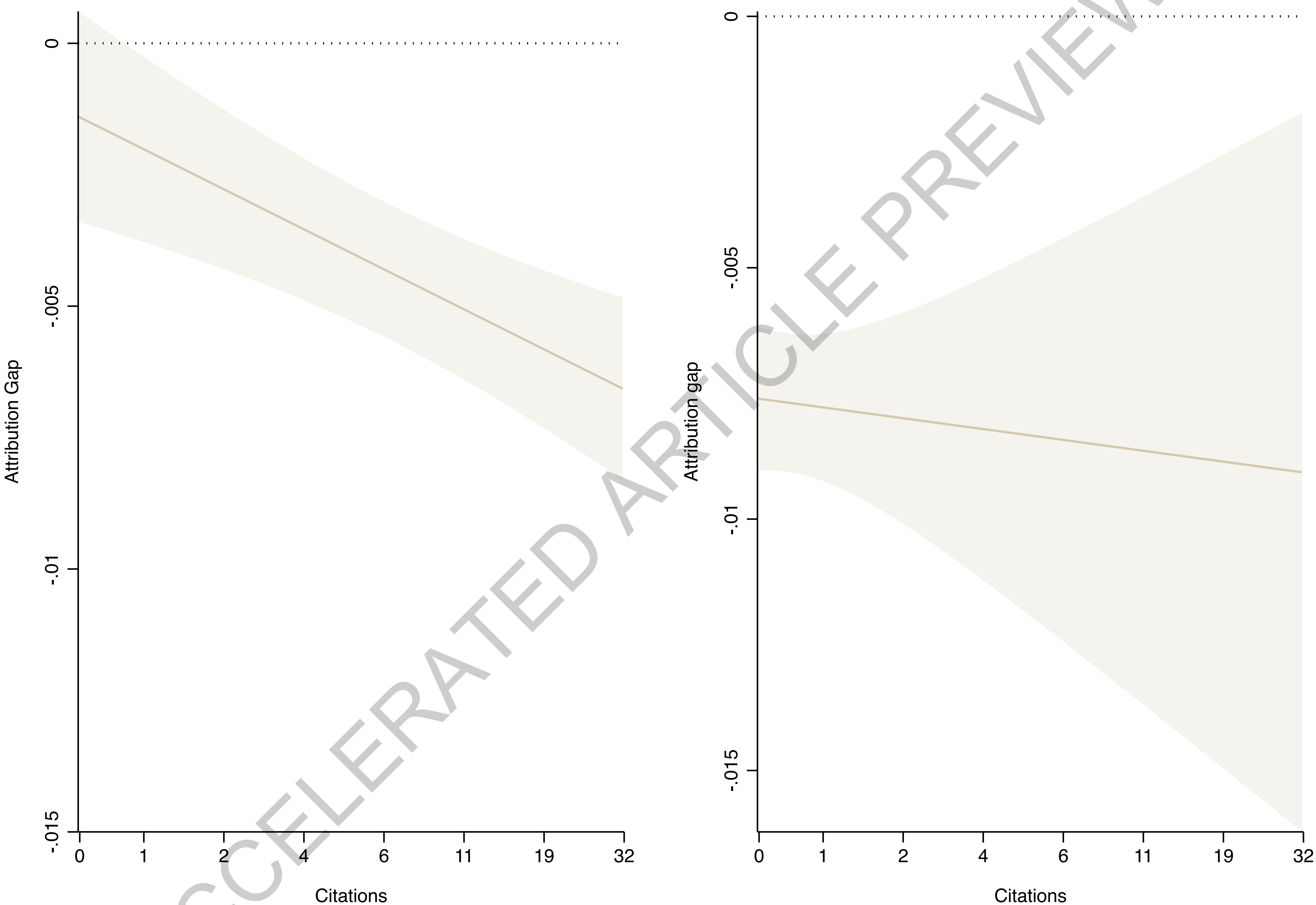
Extended Data Table 9: Characteristics of survey respondents
The table reports arithmetic means of the demographic characteristics of survey respondents.



Women's share of actual authorships
Women's share of potential authorships
.2
.3
.4
.5
.6
.7
Research Staff
Undergraduate
Post-Doc
Graduate
Faculty
Health
Psychology
Multidisciplinary
Biology
Social Sciences
Agriculture
Natural Resources
Computer Sciences
Engineering
Math
Geosciences
Physical Sciences

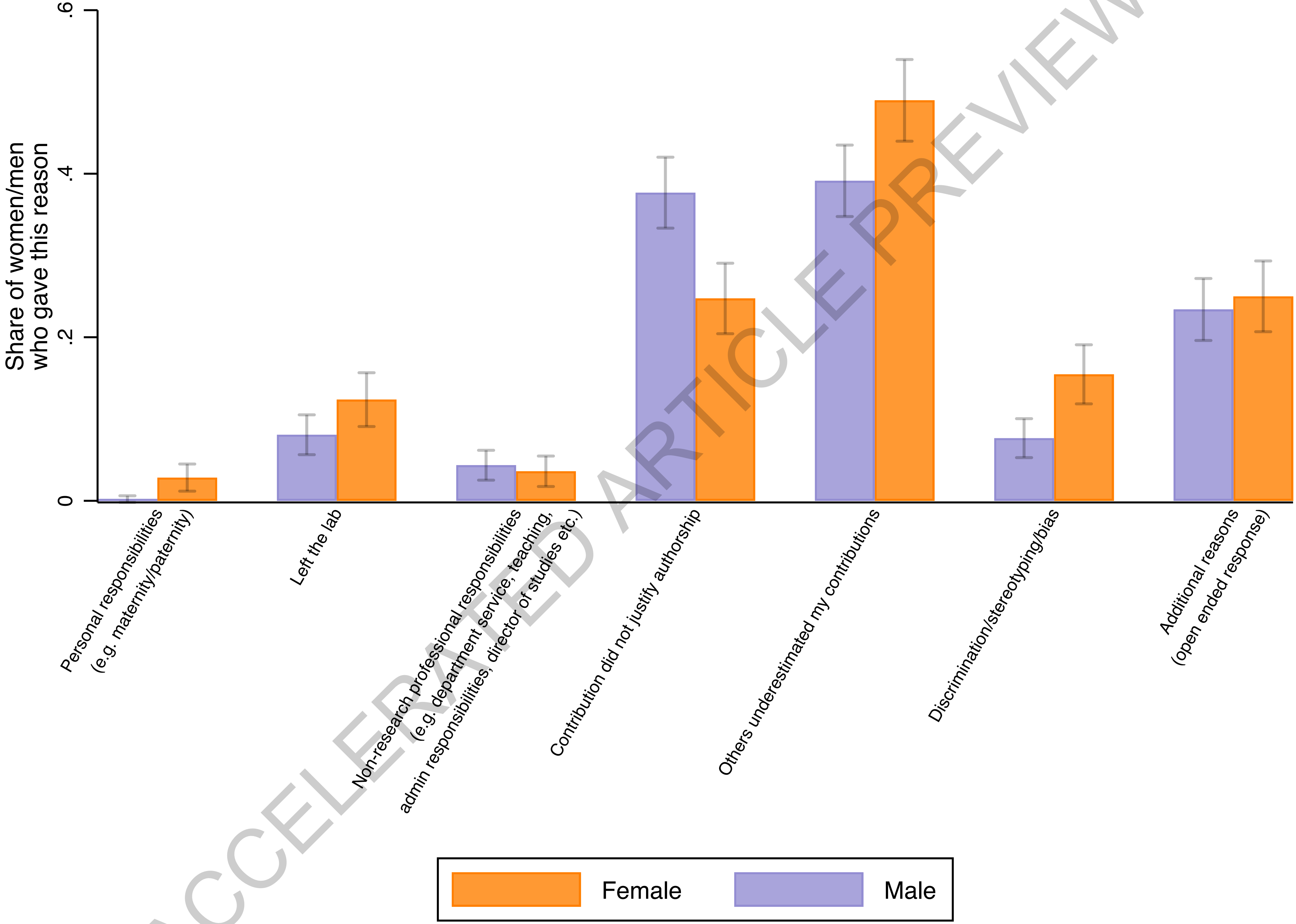


Probability named
0
.01
.02
.03
.04
Δ = 1.967pp
P<0.001
Δ = 1.392pp
P<0.001
Δ = 0.798pp
P<0.001
Δ = 0.589pp
P<0.001
Δ = 0.421pp
P<0.001
No Controls
+ Month/PI/Days
+ Job Title
+ Field
+ Team
Women
Men
Probability named
0
.01
.02
.03
.04
Δ = 1.496pp
P<0.001
Δ = 1.258pp
P<0.001
Δ = 0.998pp
P<0.001
Δ = 0.890pp
P<0.001
Δ = 0.765pp
P<0.001
No Controls
+ Month/PI/Days
+ Job Title
+ Field
+ Team
Women
Men

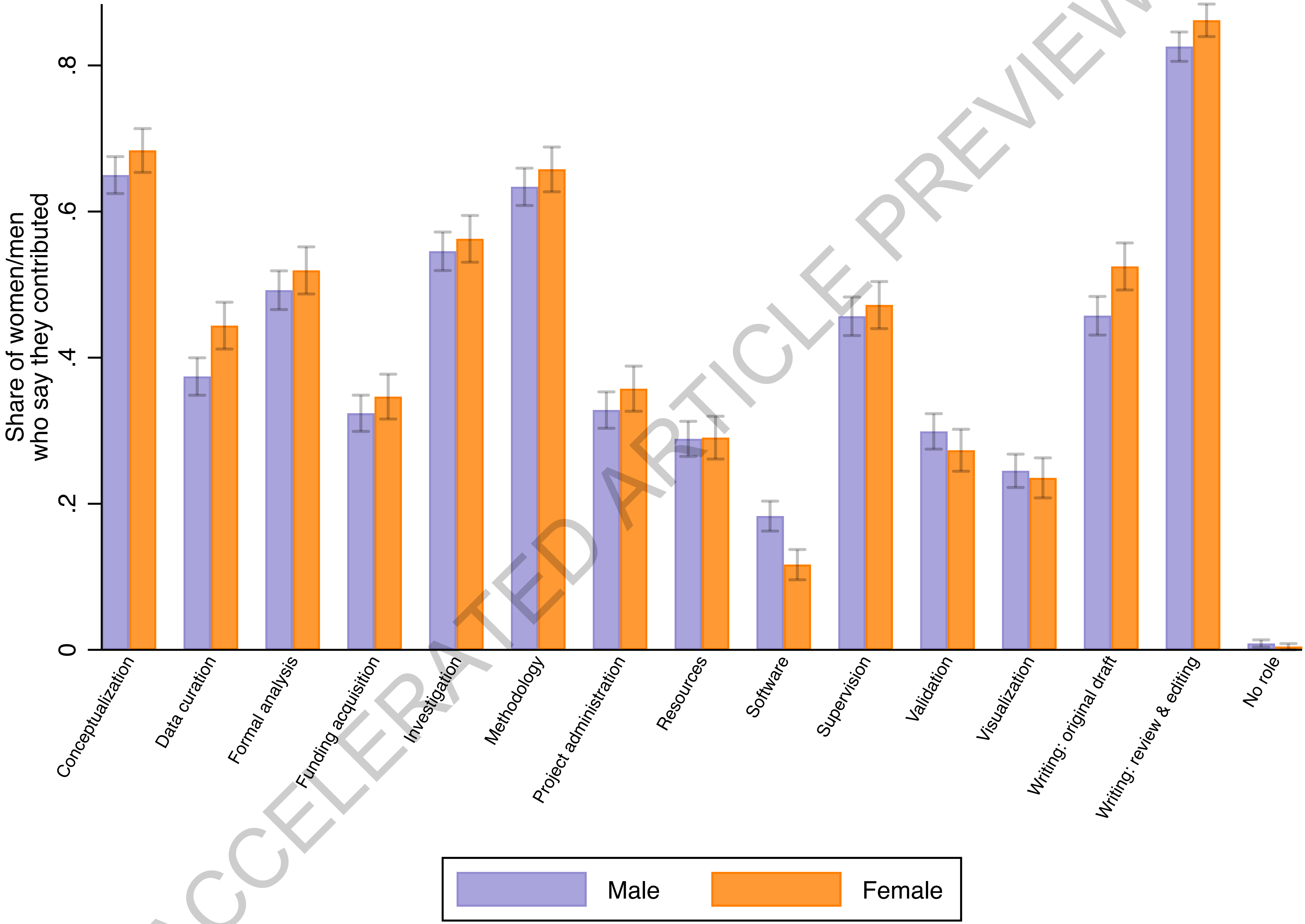


Citations: Mean = 11.24 , Std. SD = 33.90

Citations: Mean = 0.23, SD = 0.95



Share of women/men
who gave this reason
0
.2
.4
.6
Personal responsibilities
(e.g. maternity/paternity)
Left the lab
Non-research professional responsibilities
(e.g. department service, teaching,
admin responsibilities, director of studies etc.)
Contribution did not justify authorship
Others underestimated my contributions
Discrimination/stereotyping/bias
Additional reasons
(open ended response)
Female
Male



Share of women/men who say they contributed
0
.2
.4
.6
.8
Conceptualization
Data curation
Formal analysis
Funding acquisition
Investigation
Methodology
Project administration
Resources
Software
Supervision
Validation
Visualization
Writing: original draft
Writing: review & editing
No role
Male
Female

Extended Data Table 1

Team	Mean	Individual	Mean
Employees	46.91	Women	40.15%
Articles	26.60	Men	43.07%
Patents	4.66	Total Days Worked	257.58
Agriculture	4.16%	Potential Articles	139.13
Biology	18.44%	Potential Patents	24.86
Computer Sciences	3.80%	Unique Authored Articles	0.49
Engineering	8.76%	Unique Authored Patents	0.07
Geosciences	4.56%	Faculty	14.85%
Health	14.79%	Post-Doc	8.63%
Math	3.70%	Graduate Students	24.15%
Multidisciplinary	3.77%	Undergraduate Students	16.96%
Natural Resources	4.03%	Research Staff	35.41%
Physical Sciences	6.24%		
Psychology	7.15%		
Social Sciences	9.16%		
Total Teams	9,778	Total Employees	128,859

Extended Data Table 2

	(a) Rate of "Ever Authors"			(b) Rate of Authorship for a Given Document		
	Total	Women	Men	Total	Women	Men
Total Rate	16.97%	12.15%	21.17%	3.17%	2.12%	4.23%
Total Count of Authors/Authorship	18,034	6,284	11,750	3,444	1,097	2,347
Total in Workforce	107,240	51,738	55,502			
	Job Title					
	Total	Women	Men	Total	Women	Men
Graduates	18.69%	14.97%	21.37%	2.45%	2.19%	2.77%
Post-Docs	25.17%	22.35%	27.31%	4.04%	3.39%	4.32%
Faculty	45.70%	41.25%	48.86%	11.54%	9.34%	13.00%
Res. Staff	8.63%	6.59%	11.01%	1.25%	0.88%	1.67%
Undergraduates	2.61%	2.22%	3.10%	0.43%	0.33%	0.57%
	Field					
	Total	Women	Men	Total	Women	Men
Agriculture	17.00%	12.31%	20.81%	3.00%	2.02%	3.93%
Biology	19.59%	14.65%	24.26%	3.60%	2.52%	4.71%
Computer Sciences	16.98%	11.99%	20.88%	3.10%	2.06%	4.08%
Engineering	19.09%	13.35%	22.78%	3.22%	2.18%	4.09%
Geosciences	18.43%	12.86%	22.32%	3.43%	2.23%	4.45%
Health	17.47%	13.02%	22.55%	3.06%	2.12%	4.19%
Math	17.52%	12.37%	21.73%	3.58%	2.24%	4.84%
Multidisciplinary	15.64%	11.21%	19.80%	2.70%	1.86%	3.63%
Natural Resources	16.54%	11.99%	20.54%	2.99%	2.02%	3.98%
Physical Sciences	21.15%	14.98%	25.03%	3.88%	2.58%	4.93%
Psychology	16.31%	11.89%	21.08%	2.96%	2.02%	4.10%
Social Sciences	16.28%	11.80%	20.51%	2.89%	1.96%	3.90%

Extended Data Table 3

	Actual (real)	Potential (real)	Effect Size (real)	SE (bootstrap)	t-Test
Grad Students	33.81%	39.32%	-5.51%	0.0086	-6.38
Postdocs	32.61%	38.11%	-5.51%	0.0108	-5.08
Faculty	27.73%	34.82%	-7.09%	0.0053	-13.34
Research Staff	45.09%	60.81%	-15.72%	0.0099	-15.81
Undergrad Student	35.22%	48.33%	-13.11%	0.0244	-5.37
Agriculture	31.59%	47.26%	-15.67%	0.0059	-26.77
Biology	33.27%	48.29%	-15.02%	0.0050	-30.24
Computer Sciences	29.62%	45.41%	-15.79%	0.0056	-28.45
Engineering	28.10%	42.22%	-14.13%	0.0049	-29.07
Geosciences	27.08%	42.60%	-15.52%	0.0060	-25.89
Health	36.41%	53.11%	-16.70%	0.0051	-32.61
Math	28.04%	45.69%	-17.64%	0.0065	-27.12
Multidisciplinary	34.19%	50.38%	-16.19%	0.0063	-25.74
Natural Resources	32.19%	48.35%	-16.16%	0.0056	-28.94
Physical Sciences	25.95%	40.08%	-14.12%	0.0056	-25.44
Psychology	35.18%	52.35%	-17.17%	0.0053	-32.25
Social Sciences	33.70%	50.30%	-16.60%	0.0053	-31.35

Extended Data Table 4

	(1)	(2)	(3)	(4)	(5)
			Articles		
Woman	-0.01967***	-0.01392***	-0.00798***	-0.00589***	-0.00421***
	(0.00083)	(0.00071)	(0.00069)	(0.00069)	(0.00066)
Effect size relative to mean	-0.6186	-0.4377	-0.2509	-0.1852	-0.1324
			Patents		
Woman	-0.01496***	-0.01258***	-0.00998***	-0.00890***	-0.00765***
	(0.00078)	(0.00071)	(0.00069)	(0.00070)	(0.00071)
Effect size relative to mean	-1.1420	-0.9603	-0.7618	-0.6794	-0.5840
			Controls		
Month		X	X	X	X
PI Flag		X	X	X	X
Days		X	X	X	X
Job Title			X	X	X
Field				X	
Team					X

Extended Data Table 5

	Articles	Patents
Woman * Faculty	-0.01310***	-0.03432***
	(0.00280)	(0.00296)
Woman * Postdoc	-0.00662***	-0.00172
	(0.00237)	(0.00200)
Woman * Grad. Student	-0.00035	-0.00273***
	(0.00132)	(0.00099)
Woman * Res. Staff	-0.00443***	-0.00436***
	(0.00070)	(0.00084)
Woman * Undergraduate	0.00013	0.00202***
	(0.00082)	(0.00070)

Extended Data Table 6

	Articles	Patents
Woman * Agriculture	-0.00311	0.01076
	(0.01425)	(0.01666)
Woman * Biology	-0.01379**	-0.02820***
	(0.00558)	(0.00578)
Woman * Computer Science	-0.00899	-0.00019
	(0.02301)	(0.01679)
Woman * Engineering	-0.00086	0.00990
	(0.01180)	(0.01067)
Woman * Geosciences	-0.04529**	-0.00838
	(0.01838)	(0.02062)
Woman * Health	0.01811**	0.00379
	(0.00728)	(0.00737)
Woman * Math	0.02704	-0.01591
	(0.02569)	(0.02276)
Woman * Multidisciplinary	0.03135*	-0.00271
	(0.01686)	(0.01598)
Woman * Natural Resources	-0.02702	-0.01450
	(0.01783)	(0.02085)
Woman * Physical Science	-0.01337	-0.04382***
	(0.01518)	(0.01415)
Woman * Psychology	-0.02263*	0.00125
	(0.01311)	(0.01375)
Woman * Social Science	0.00148	0.00650
	(0.01027)	(0.00956)
Woman * Other	-0.00121	-0.00307
	(0.00303)	(0.00233)

Extended Data Table 7

	Articles	Patents
Woman	-0.00139	-0.00760***
	(0.00102)	(0.00073)
log(citation + 1)	0.00083***	-0.00011
	(0.00024)	(0.00057)
Woman * log(citation + 1)	-0.00152***	-0.00043
	(0.00040)	(0.00111)

Extended Data Table 8

	Date of first email	Date of reminder email	Sampling structure	Number of emails sent	Number of emails that bounced or received an automated response	Number of respondents who opened the survey	Number of respondents who completed the survey
Pilot 1	Jan 20, 2022	Feb 8, 2022	Random	500	79	85	74
Pilot 2	Jan 25, 2022	Feb 8, 2022	Random	500	73	77	75
Pilot 3	Feb 3, 2022	Feb 17, 2022	Random	500	69	68	65
Main study	Feb 15, 2022	March 1 and March 9, 2022a	Stratified on gender	26,500	4,116	2,705	2,446

Extended Data Table 9

Respondent Characteristics	Arithmetic Mean
Women	39.97%
Non-binary / Fluid / Prefer not to answer	0.82%
Age	49.72
Received BA in the US	14.08%
Hispanic / Latin / Spanish Origin	24.07%
White	83.24%
Black or African American	1.52%
Asian	14.55%
American Indian or Alaska Native / Native Hawaiian or Other Pacific Islander	0.69%
Faculty	52.24%
Post-Doc	13.92%
Research Staff	21.69%
Graduate Student	8.23%
Undergrad/Other/ Prefer not to answer	3.93%
Arts / Humanities / Other	1.72%
Computer Science	3.22%
Engineering	11.99%
Environmental Sciences	10.23%
Life Sciences	33.19%
Mathematical Sciences	2.84%
Other Sciences	4.47%
Physical Sciences	18.49%
Psychology	4.04%
Social Sciences	9.80%

nature portfolio

Corresponding author(s): Julia Lane

Last updated by author(s): 5/28/2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The UMETRICS data are hosted in the the Virtual Data Enclave at the University of Michigan. The data collection process from each university is described here https://iris.isr.umich.edu/wp-content/uploads/2022/01/new-member-handbook-2022.pdf Because the data are drawn directly from university HR and Finance systems, and each university can have different systems, the handbook notes "The task of compiling and transmitting administrative data from your HR, procurement, and research systems may feel daunting. Some institutions have systems operating on very different platforms and are challenged at the thought of integrating disparate data sets, while others express concern about having to commit significant resources to compiling data. At IRIS, we have worked with institutions that are quite diverse in how they manage data and we will walk through all of these issues with your data point of contact. Our technical director, Kevin Bjorne (kbjorne@umich.edu), has an outstanding record of helping institutions manage this process effectively. Kevin estimates the initial data transmission may take about 40 hours of institutional effort, and considerably less time for subsequent data transmissions. For institutions that participated in the federal STARMETRICS program this time can be much reduced by adapting existing scripts, as IRIS data are based on STARMETRICS data formats. Please contact us at IRIS-info@umich.edu to schedule an individual phone call or conference call to review the process if you have not done so already"
Data analysis	All the Stata code (Version 17) and Python 3.7.6 code used is available in the Virtual Data Enclave at the University of Michigan

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The datasets generated during and/or analysed during the current study are available, as well as the associated code, at the Virtual Data Enclave repository at the Institute for Research on Innovation and Science at the University of Michigan. Access information is provided here https://iris.isr.umich.edu/research-data/access/ . Patent data were obtained from Patentsview (https://patentsview.org/), which is publicly available. Web of Science data were obtained from CADRE at Indiana University (https://iuni.iu.edu/resources/datasets/cadre).
The survey data are not available as per the University of Pennsylvania IRB protocols. Aggregate statistics from the survey data can be made available to researchers (upon request) for replication

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Sex and gender are the core of the analysis. Both males and females were studied, as well as those for whom no gender could be identified
Population characteristics	UMETRICS data: 128,859 individuals from 72 college and campuses who were paid on a grant in the period 2013-2016 from a participating institution. 51, 737 were female. 55,500 were male. Gender could not be determined for 21,622. Survey: 2,446 Individuals who: (1) had a public profile on ORCID, (2) had an associated email address, and (3) published at least one academic paper in the Web of Science database between 2014 and 2018. 978 were female, 20 were fluid/ undefined gender, 1143 were male. The mean age was 49.72 years. 344 identified as Hispanic, 2,036 were White, 37 were Black, 356 were Asian, 17 were American Indian or Alaska Native / Native Hawaiian or Other Pacific Islander Full details are presented in ED Tables 1, 2, 9, and 10.
Recruitment	For UMETRICS inclusion: All individuals who were paid on a research grant at participating institutions and whose data were provided by the institution were included in the study. For survey inclusion: We began by identifying individuals who had a public profile on ORCID, had an associated email address, and published at least one academic paper in the Web of Science database between 2014 and 2018. After adjusting for duplicates, there were 98,022 unique ORCID profiles that matched our sample criteria. We ran three pilots that took samples of 500 individuals each that matched this criteria. We then stratified on gender for the main study, sampling 10,000 male, 10,000 female, and 6,500 gender-ambiguous names (based on the Ethnea database). We emailed each of the individuals described above through the Qualtrics platform with a recruitment script and personalized email link (which incorporated information about the published article they were linked to through Web of Science). The full email text and survey information can be found in Section C of the Supplementary Online Materials. Because our sample is based on those individuals who choose to respond to the survey, self-selection bias may exist. In particular, perhaps those who are most concerned about issues around attribution would be those most likely to choose to complete the survey. This could result in an inflated rate of respondents stating they have been left off of papers. However, since gender is not mentioned in the recruitment script, we do not expect this bias to differ across gender. For interview inclusion: 338 individuals indicated on the survey that they would be open to an interview, and provided an email address. We selected six individuals among those 338 to interview. For UMETRICS inclusion: Inclusion in the UMETRICS database did not involve active recruitment. [INSERT MORE DETAIL]
Ethics oversight	University of Pennsylvania Institutional Review Board (IRB Protocol # 850522) approved the survey. University of Pennsylvania Institutional Review Board (IRB Protocol # 850522), Boston University Institutional Review Board (IRB Protocol #6412X) and the New York University Institutional Review Board (IRB Protocol #IRB-FY2022-6243) and the Ohio State University Institutional Review Board (IRB Protocol 2022E0133) approved the followup interviews.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

Both quantitative and qualitative. The quantitative component primarily relies on UMETRICS administrative data, constructing a potential attribution rate to a realized attribution rate within university administrative data, and how these rates differ by gender. The survey and interviews focused on the allocation of credit more broadly, with both quantitative components such as the roles on published papers, while the qualitative component had open-ended responses on the reasons behind not receiving credit.

Research sample

The UMETRICS dataset is constructed from three sources: internal Finance and Human Resources administrative data from 72 colleges and campuses, representing over 40% of total academic R&D spending in the United States, journal articles from the Web of Science and patent data derived from the universe of patents from the US Patent and Trademark Office (USPTO). The analysis focuses on a subset of 52 college campuses which consistently provided data for the period covering 2013-16. This restriction ensures that employment spells are long enough to reasonably identify PIs and teams as well as to observe the scientific documents produced by those teams from 2014-16. The full data include administrative level information from approximately 440,000 unique federal and non-federal awards, including approximately 23 million wage payments to about 650,000 deidentified individuals. The sample represents over 40% of federal funding made to academic institutions. The population of funding from non-federal (philanthropic, state, industry, and local) funding is unknown, so it is not possible to determine the representativeness of the non-federal portion of the funding. Similarly, the population of research teams is unknown, as is the population of individuals supported on research grants, so it is difficult to determine the representativeness of the UMETRICS sample. We are not aware of another large-scale dataset other than UMETRICS that could be used for this analysis.

The survey data was drawn from a sample of authors with ORCID IDs who recently published an academic article in Web of Science and had an associated email address. The sample was selected because of the personalized nature of the survey; each respondent received a personalized survey link to their email address, and the personalized survey included questions about a specific paper that they had published. The respondents to the survey overrepresented faculty members, women, and academics who received their Bachelor's degree outside of the US.

Sampling strategy

The UMETRICS data represent the universe of all transactions data for the campuses at participating institutions for the years in which they submitted data. Universities are recruited through consistent partnership with the Association of American Universities (www.aau.edu) and the Association of Public LandGrant Universities (www.aplu.org), as well as with the United Negro College Fund (https://Uncf.org) and Excelencia (https://edexcelencia.org). Details of membership are provided here https://iris.isr.umich.edu/membership/.

We subset the data to those campuses that consistently provided data for the period covering 2013-2016. Full details are available here https://iris.isr.umich.edu/research-data/2019datarelease/

For the survey, below is our calculation of the estimated sample size needed for two-sample comparison of proportions

Test Ho: p1 = p2, where p1 is the proportion in population 1
and p2 is the proportion in population 2
Assumptions:

alpha = 0.0500 (two-sided)
power = 0.8000
p1 = 0.2500
p2 = 0.3000
n2/n1 = 1.00

Estimated required sample sizes:

n1 = 1291
n2 = 1291

Based on the pilot samples (which drew a random sample), women composed a small proportion of the respondents to our survey. This is likely due to underrepresentation in the scientific academic community more generally. As a result, we stratified by gender for the main study: 10,000 women, 10,000 men, and 6,500 gender-ambiguous profiles were randomly selected to survey.

Data collection

The UMETRICS data are transactions data produced by each university. The information about how the data are produced, processed and standardized are here https://iris.isr.umich.edu/membership/for-current-members/

The survey data were collected through an online web-based (Qualtrics) survey. The full information is below
1. Target Population and Accrual:
The target population was researchers with scholarly publications. We accessed the target population through a sample of Web of Science published authors.

2. Key Inclusion Criteria:
All subjects must have published an article in a scholarly journal or have worked on an article that was eventually published.

3. Key Exclusion Criteria:
Not applicable

4 Subject Recruitment and Screening:
We emailed a sample constructed for our survey from public ORICD records and the Web of Science, as detailed below. The ORCID database contains CV-style information of millions of academic researchers. We use publicly available information that researchers have chosen to make public. Each ORCID record is associated with an ORCID ID, which is a unique identifier for the academic researcher.

We focused on the 897,264 ORCID records that listed a complete name in addition to at least one employment spell or at least one educational degree. We filtered these ORCID records to only those for which we have an associated email address (128,602). Because the ORCID database does not contain email addresses, we link to the Web of Science database, which contains e-mail addresses and the bibliometric information on a wide range of academic publications. Because the focus was on asking academic researchers about their experience with being named or not being named as coauthors on publications, the ORCID profiles were restricted to those that could be linked with a published academic paper in the Web of Science database between 2014 and 2018: 98,022 profiles fulfilled those criteria and were not duplicates.

5. Early Withdrawal of Subjects:
Participation was completely voluntary; all respondents could simply not complete the full survey and were informed that they can stop participating at any time.

6. Vulnerable Populations:
Not applicable

7. Populations vulnerable to undue influence or coercion:
Not applicable.

STUDY DESIGN:
We launched the survey after three pilots, which were doing using random samples of 500 names each. We sent out the survey with one follow-up reminder after one week. The survey was designed to gain a deeper understanding as to how credit is distributed, and whether that credit distribution varies for men and women. The survey was emailed out to respondents and was hosted on the Qualtrics platform. The survey was designed to take fewer than five minutes.

We followed up with one-on-one interviews if respondents indicated that they'd like to be contacted after the survey (in response to the final question on each survey: "We are seeking individuals to interview regarding their experiences with the allocation of credit in research teams. If you would be interested in talking with us about your experiences, please enter your email below. Your responses will be kept confidential."). The interviews occurred over Zoom for 30 minutes, and were recorded and transcribed in the instances when the respondent gave permission.

The data were analyzed at Britta Glennon's office at Wharton, and a de-identified and aggregated version of the data was shared with her co-authors at their institutions.

Julia Lane (at NYU) and Raviv Murciano-Goroff (at Boston University) also obtained IRB approval to conduct interviews with Britta Glennon.

Timing	The UMETRICS data is 2013-2016; the publication and patent data (which are publicly available) go through 2019. The Survey data collection began in January 2022 and concluded in April 2022.
Data exclusions	Not applicable.
Non-participation	Participation was completely voluntary; all respondents could simply not complete the full survey and were informed that they can stop participating at any time.
Randomization	NA

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging