

Is Science Built on the Shoulders of Women? A Study of Gender Differences in Contributorship

Benoit Macaluso, MS, Vincent Larivière, PhD, Thomas Sugimoto, MPA,
and Cassidy R. Sugimoto, PhD

Abstract

Purpose

Women remain underrepresented in the production of scientific literature, and relatively little is known regarding the labor roles played by women in the production of knowledge. This study examined labor roles by gender using contributorship data from science and medical journals published by the Public Library of Science (PLOS), which require each author to indicate their contribution to one or more of the following tasks: (1) analyzed the data, (2) conceived and designed the experiments, (3) contributed reagents/materials/analysis tools, (4) performed the experiments, and (5) wrote the paper.

Method

The authors analyzed contribution data from more than 85,000 articles published between 2008 and 2013 in PLOS journals with respect to gender using both descriptive and regression analyses.

Results

Gender was a significant variable in determining the likelihood of performing a certain task associated with authorship. Women were significantly more likely to be associated with performing experiments, and men were more likely to be associated with all other authorship roles. This holds true controlling for academic age: Although experimentation

was associated with academically younger scholars, the gap between male and female contribution to this task remained constant across academic age. Inequalities were observed in the distribution of scientific labor roles.

Conclusions

These disparities have implications for the production of scholarly knowledge, the evaluation of scholars, and the ethical conduct of science. Adopting the practice of identifying contributorship rather than authorship in scientific journals will allow for greater transparency, accountability, and equitable allocation of resources.

What constitutes women's work in science? The rhetoric of the "unique" and "special" talents of women relegated them to roles as amateurs and technical assistants in the production of knowledge in the early 20th century.¹ This secondary role was sustained by the relative underproduction and subsequent attrition of women in the labor market and the hierarchical nature of scientific inquiry promoted in the Big Science era, the latter part of the 20th century in which science shifted from small teams and individual work to larger teams requiring larger infrastructure and instrumentation for research. In contemporary science and medicine, women are matriculating at a greater rate than men, but still remain underrepresented in scientific output.²

Furthermore, very little is known about the labor roles played by women and whether sex segregation in the production of science persists, despite claims for universalism in science.³

The inability to conduct large-scale analyses on this question has been largely a function of the idiosyncratic authorship practices of the 20th century, which provided authors on a byline (ordered in various ways, by disciplinary practice), without an acknowledgment of actual contributions.⁴ These practices have led to several concerns—most notably the lack of public accountability of authors when issues of fraudulent research arise and the prevalence of ghost and honorific authorship. To counter this, medical journal editors have been proactive in arguing for greater specificity around authorship roles: For example, the International Committee of Medical Journal Editors recommendations include explicit criteria for what constitutes authorship on an article.⁵

However, practices of ghost and honorific authorship continue^{6,7} as these standards fail to capture the precise roles played by each author on the article—an issue that is further complicated by rising

rates of hyperauthorship (the increase in number of authors on a given article).⁸ The notion of replacing authorship with contributorship was advanced in the late 1990s as a "radical change" in scholarly publishing⁹ whereby authors would be listed by their roles, rather than in a ranked order. Fifteen years later, a few journals have begun to include contributor information, without abandoning the byline. Some have done this systematically (e.g., *JAMA*), but most collect data idiosyncratically—either in the form of acknowledgments or by providing an open field for specifying contributions (e.g., the *BMC Medical Education* journal). The lack of systematic data collection limits large-scale data mining.

The Public Library of Science (PLOS) is one early adopter of the practice of identifying author contributions. PLOS, a nonprofit open access publisher of seven high-impact journals, provides criteria for authorship and requests that the contribution of each author on the byline be identified in the following ways: analyzed the data; conceived and designed the experiments; contributed reagents/materials/analysis tools; performed the experiments; and wrote

Please see the end of this article for information about the authors.

Correspondence should be addressed to Cassidy R. Sugimoto, School of Informatics and Computing, Indiana University Bloomington, 919 E. 10th St., Bloomington, IN 47408; telephone: (812) 856-2323; e-mail: sugimoto@indiana.edu.

Acad Med. 2016;91:1136–1142.

First published online June 7, 2016
doi: 10.1097/ACM.0000000000001261

Supplemental digital content for this article is available at <http://links.lww.com/ACADMED/A370>.

the paper. Each author can be assigned to one or more of these roles (other roles have been present historically; however, those enumerated represent the five main categories).

In this study, we analyze the contribution data found in 85,260 articles published between 2008 and 2013 in PLOS journals with respect to gender (controlling for variables such as discipline, authorship status, academic age, discipline, team composition, and country of author). Specifically, we seek to address whether there are gendered differences in scientific labor roles. This allows us to reveal for the first time, to our knowledge, a large-scale analysis of the differing roles played by the sexes in contemporary knowledge production.

Method

Data

We used two sources of data: Thomson Reuters's Web of Science (WoS) and all articles published by PLOS, available on the PLOS Web site in XML format. As of 2014, WoS had published more than 50 million articles in almost 20,000 journals. PLOS has published 8 peer-reviewed scientific journals, largely in the biomedical area: *PLOS Biology* (founded in 2003), *PLOS Medicine* (2004), *PLOS Genetics* (2005), *PLOS Computational Biology* (2005), *PLOS Pathogens* (2005), *PLOS ONE* (2006), and *PLOS Neglected Tropical Diseases* (2007). *PLOS Clinical Trials* was published in 2006 and 2007 only. Nearly 95% of the articles used in this study were classed as biomedical (41.8%), clinical medicine (44.9%), or biology (7.9%).

The digital object identifiers (DOIs) for PLOS articles are provided in article-level metrics, which we used to match each PLOS article with the corresponding record in WoS. (For a full description of this matching process, see Supplemental Digital Appendix 1 at <http://links.lww.com/ACADMED/A370>.)

The dataset of PLOS articles, including authors' contributions as well as all WoS metadata, served as the sampling frame for the study. From this, we excluded any document types that were not standard articles and review articles. Furthermore, we included articles published only between 2008 and 2013 because WoS only

provided full first names from this time period. We also excluded the journal *PLOS Clinical Trials* from the analysis because it did not publish any articles during the period covered. Also excluded were articles lacking contributorship data ($n = 962$) as well as those for which a match could not be established between PLOS and WoS ($n = 369$). Four duplicates were also excluded. The final dataset included 85,260 articles published in seven PLOS journals between 2008 and 2013, for which we managed to establish a link between the PLOS and the WoS record.

Processing

The data underwent several rounds of processing before analysis. First, we parsed the contributions field to extract contribution type. There are several formats in which this could appear, so specialized code had to be developed for this. One of the unique features is that authors are listed by initials, rather than names, so the initial had to be matched back to author names (see Supplemental Digital Appendix 1 at <http://links.lww.com/ACADMED/A370>). We assigned academic age and gender to the authors, based on full name data. We matched the gender of authors according to the gender assignment tables developed by Larivière and colleagues.² This list uses given names and country combinations to assign gender to authors of articles. On the whole, this conversion list managed to assign a gender to 88.1% of authorships (i.e., author–article combinations), of which 32.5% were female and 55.7% were male. Initials and unisex names accounted for 0.2% and 2.7%, respectively, while the unknown rate was 8.9% (similar to what was found in Larivière et al.² for all WoS articles—i.e., 8.4%).

We estimated academic age of authors using their year of first publication, as recorded in WoS. To obtain such age, authors found in the WoS were disambiguated automatically by the Center for Science and Technology Studies (Leiden University) using the algorithm developed by Caron and van Eck.¹⁰ Given that the majority of authors had less than 30 years of publication experience, our analyses focus on those who have between 0 (i.e., the first year in which a contributor was listed on a publication) and 30 years. It should be noted that, throughout this article, when the term “academic age” is used, it refers to years since first publication.

Analysis

We performed descriptive analyses on the entire dataset. An analytic sample was constructed for the regression using those observations that contained all necessary variables (i.e., author position [first and last], academic age, number of authors, percent of female authors, country, and discipline): 270,103 observations were used of 589,906 possible observations. Differences between the regression and descriptive samples can be found in the Supplemental Digital Appendix 1 at <http://links.lww.com/ACADMED/A370>. The most significant difference is the near-total exclusion of 2013 data from the regression, given that academic age data were unavailable for this year. Given this, regressions should be interpreted as describing 2008–2012, whereas the descriptive figures (excluding those with age data) describe the 2008–2013 data. We conducted the regression analyses using SAS statistical software version 9.4 (SAS Inc., Cary, North Carolina).

Results

Hierarchy of science

Although not explicit, ranked authorship carries implicit assumptions of roles in certain disciplines. For example, in the biomedical sciences, the corresponding author is often the project investigator and senior researcher, while the first author is the one who took a lead role in conducting the research. We analyzed contributorship by the five previously defined roles (data analysis, experiment design/conception, material contribution, experimentation, writing) and author position. Figure 1 presents the proportion of men and women associated with certain labor roles by the gender of the dominant author. When there was a female first or corresponding author, women were more likely to be associated with all tasks except contributing materials. In the case of male corresponding or first author, men were more likely to be associated with all tasks except experimentation. The largest gaps can be seen in terms of experimentation with female dominant authors—women were significantly more likely to be associated with experimentation in the case of dominant female authors.

Academic age roles

Women had, on average, younger academic age than men in our sample, and academic

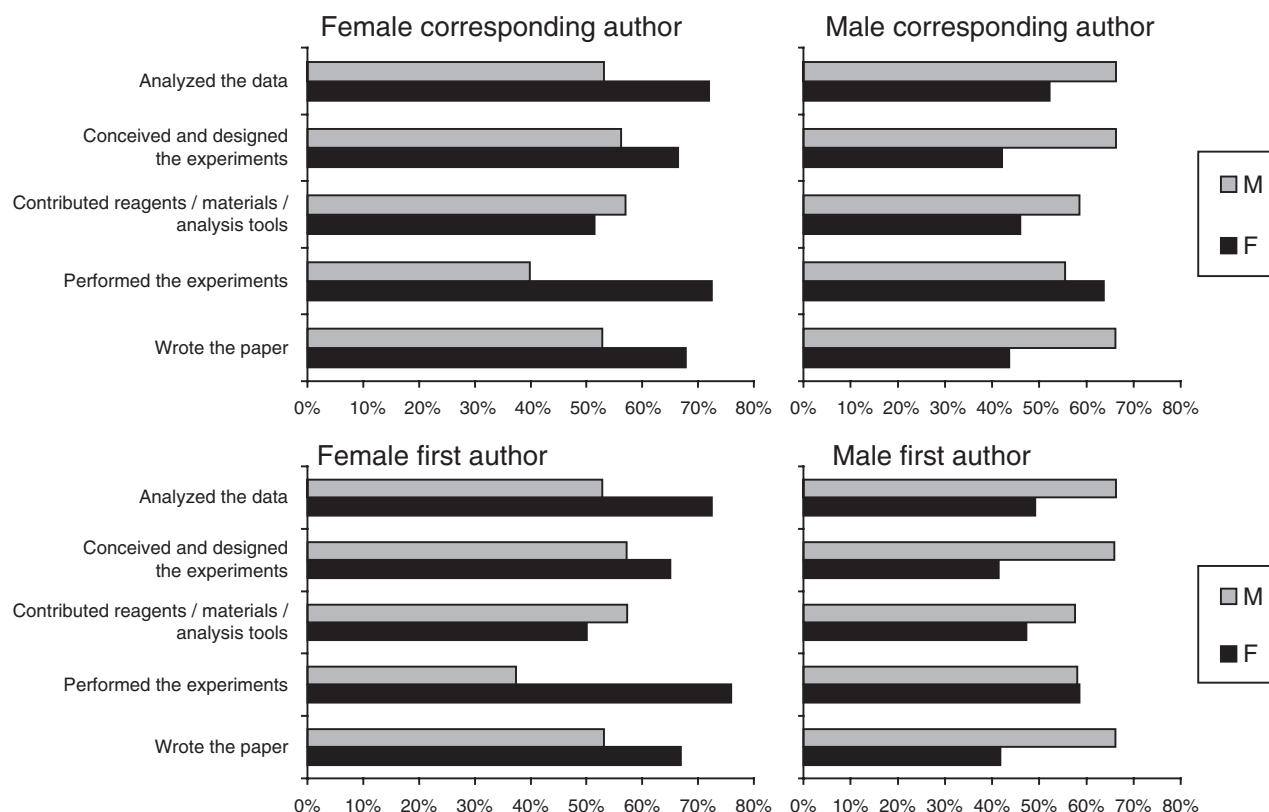


Figure 1 Proportion of male and female authors contributing to each labor role, by gender of corresponding author (top panel) and gender of first author (bottom panel), from a study of author contributorship roles and gender, 2008–2013.

age had an effect on labor roles. Figure 2 presents the proportion of all authors of a given academic age and gender associated with a particular role. For example, nearly 80% of women and 60% of men in their first year of publishing were associated with performing experiments, a proportion that decreased over time (i.e., more experienced researchers are less likely to perform experiments). However, a clear gap remained in the contributions of experimentation, regardless of academic

age, with women consistently doing proportionally more of this task.

The bottom of Figure 2 demonstrates the difference between female and male contributorships by academic age. The gap between male and female contributions in conception, contributing materials, and writing the paper equalized as the contributors age academically. Analyzing data, however, shifted from a male-dominated activity in early years

of an academic's publication record to a female-dominated activity in later years.

Team size and structure

The adage “many hands make light work” might be expected to apply in scientific collaborations: As the number of authors increases, the proportion participating in any role should decrease. We saw such a trend in most categories—with proportional contributions decreasing at nearly equal rates for men and women

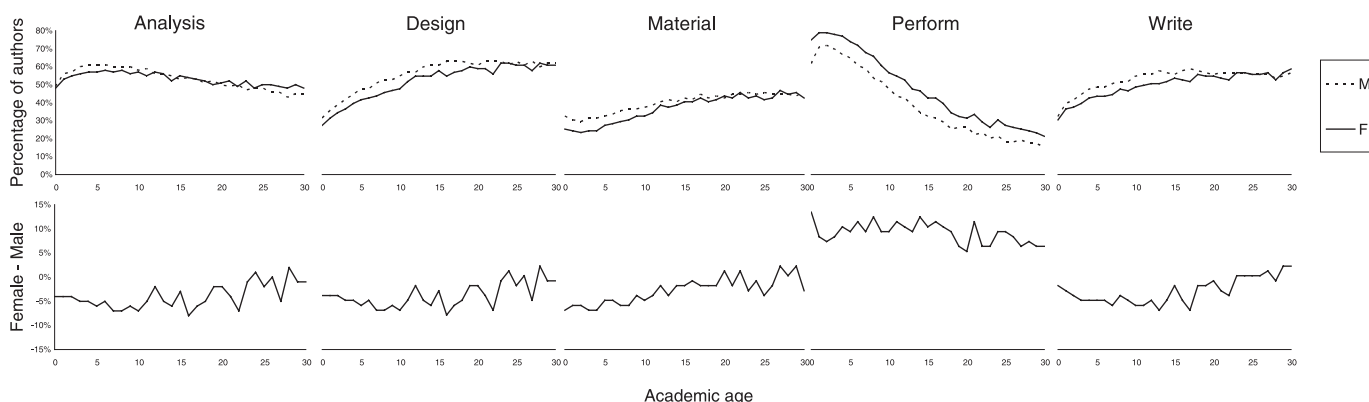


Figure 2 Proportion of male and female authorships associated with contributorship by academic age (defined as number of years since first publication) (top panel) and difference between female and male contributions by academic age (bottom panel), from a study of author contributorship roles and gender, 2008–2013.

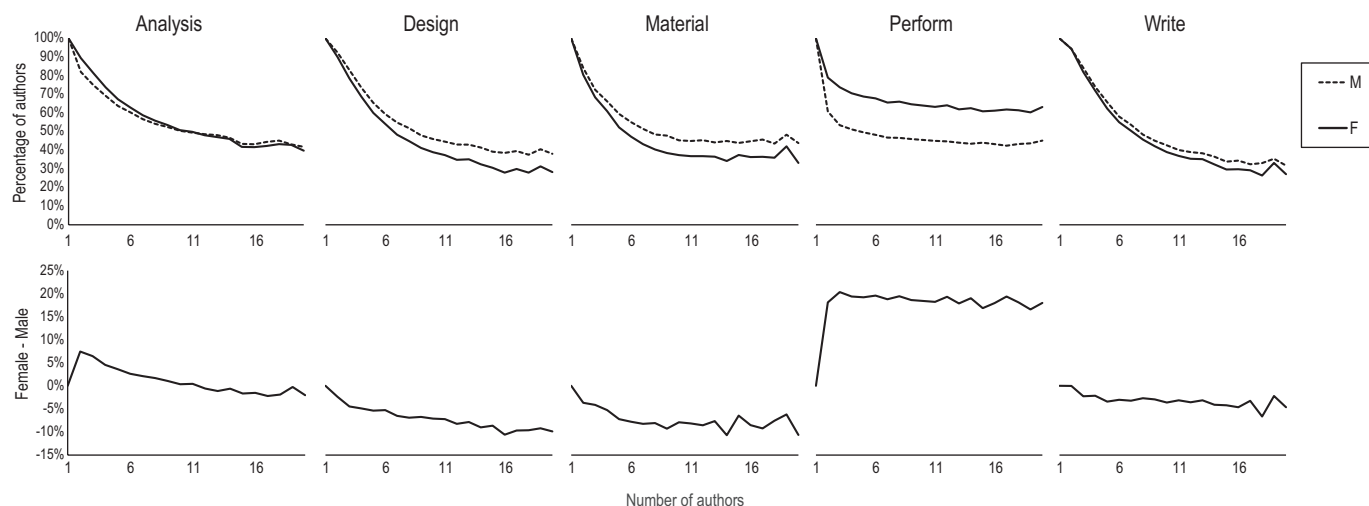


Figure 3 Proportional representation of female and male authors by number of authors on the article (top panel), and difference between female and male contributions by number of authors on the article (bottom panel), from a study of author contributorship roles and gender, 2008–2013.

(Figure 3). For performing experiments, however, a stable gap persisted between men and women with very little change in proportional representation as the team size increases. A similar pattern was observed in contributing material, with male dominance in this area. This demonstrates that an increase in team size does not lead to a more gender-balanced distribution of labor.

The relationship between the gender of the corresponding author and the gender of the contributors moderates these findings. For example, when there was a male corresponding author, the proportion of women contributing to experimentation remained stable as the team size increased (Figure 4). However, for male authors on a team with a male

corresponding author, the proportion decreased as team size increased. Men were proportionally less likely to conduct experiments when there was a female corresponding author.

Irrespective of team size, women were more likely than men to be associated with one or two types of contribution

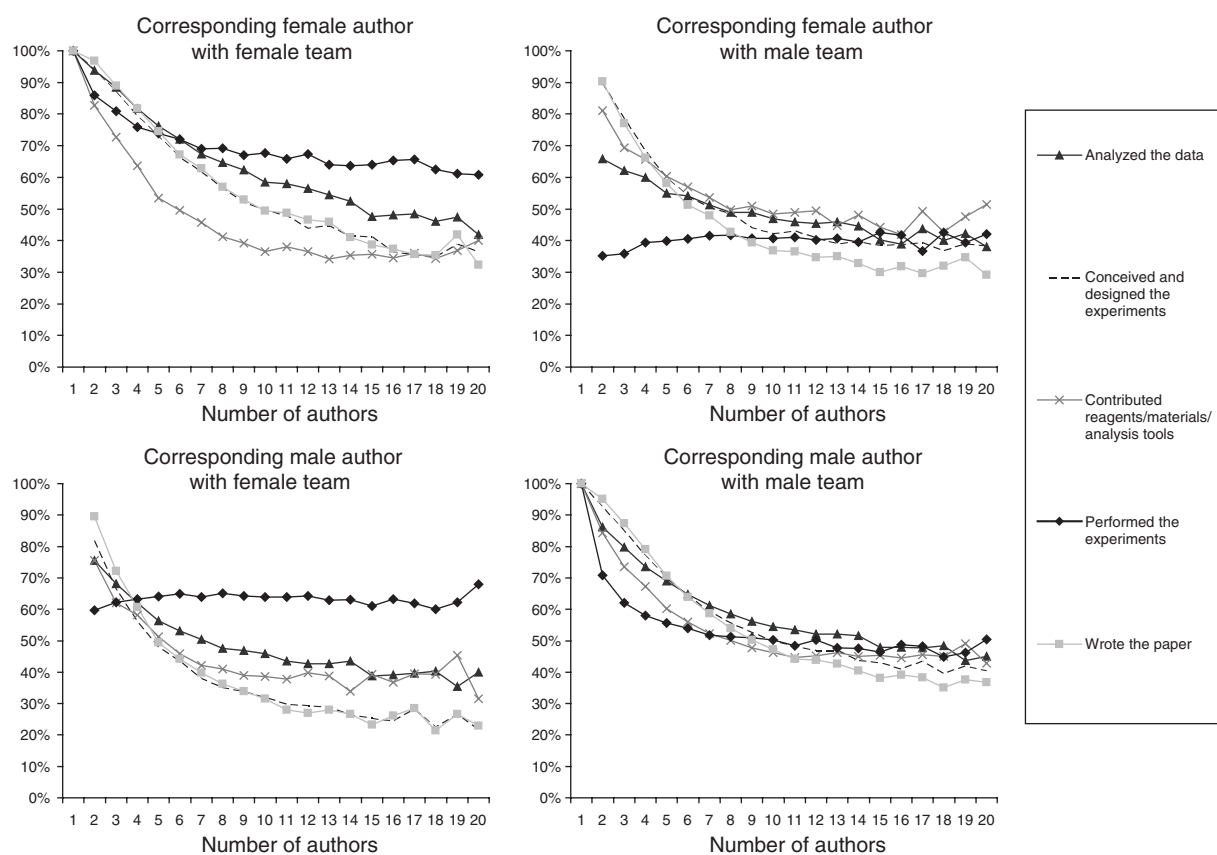


Figure 4 Proportion of male and female contribution to authorship as a function of the total number of authors, by gender of corresponding author, from a study of author contributorship roles and gender, 2008–2013.

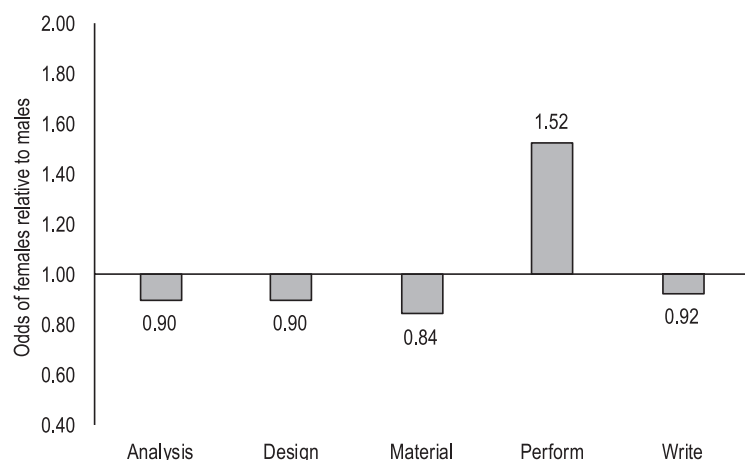


Figure 5 Odds of females identified as contributor for each type of contribution relative to males, from a study of author contributorship roles and gender, 2008–2013.

per publication (and only slightly more likely to be associated with all five contribution types). Men were more likely than women to be associated with three or four contribution types per publication.

Gender roles

Given the differences observed in academic age, dominant author, and team size, we conducted a series of logistic regressions to control for these variables. The results of the logistic regressions demonstrate a significant ($P < .0001$) relationship between gender and contributorship type when controlling for all other variables. The odds of a female author being listed as performing the experiment were 1.52 times the odds of male author (Figure 5). The odds for all other tasks were more likely to be a male author.

Regressions were conducted using the various types of contributorship (e.g., analysis, design, performing experiments) as the dependent variable and controlling for gender, authorship status, corresponding authorship status, academic age, number of authors, country, field, and publication year (Appendix 1). In each case, gender remained a significant variable in determining likelihood of performing a specific role.

Discussion

There exist in the production of science inequalities in the distribution of scientific labor, with women more likely to be associated with the “physical” labor (i.e., performing experiments), whereas men are more likely to be associated with resource contributions and “conceptual” labor (i.e.,

conceiving and designing experiments and writing the article). Gendered labor roles remained significant after controlling for academic age, discipline, country, authorship position, and proportion of male and female authors. These differences in labor roles may explain some of the disparities in the rates of scientific publication between men and women,² particularly in prestigious first and last author positions.¹¹

Prior research has established that women are underrepresented among senior author positions generally¹¹ and in prominent medical journals specifically.¹² Our findings extend this research by demonstrating that the gender of the lead author is related to the roles played by other scientists on the team. Furthermore, this study has shown that the relationship between team size and proportional contribution to various tasks differs by gender of the corresponding author. The data suggest that, in the case of male corresponding authors, the proportion of authors contributing to a single task decreases as the team size increases. The same trend was not observed for female corresponding authors. The normative value of either model is arguable. On the one hand, there is evidence that specialization in scientific roles leads to higher-quality science.^{13,14} However, one could also argue that the growing rates of fraud in authorship (and the disproportionate representation of men in fraud cases)¹⁵ suggests that the growing distance of authors from substantial aspects of authorship may have negative consequences.

Research on contributorship can reveal critical information on the mechanisms of science including, but not limited to, the role

of gender and other variables in the effective functioning of the scientific workforce. However, this is dependent on the collection of high-quality contributorship data, a practice not widely employed by journals or made available in machine-readable ways. In the field of medical education, there are very few journals that do this. The *BMC Medical Education* journal provides a designated section for author contributions. Other journals, such as *Medical Education*, *Medical Teacher*, and *Advances in Health Sciences Education*, provide no indication regarding contributorship.

Our research has, to our knowledge, provided the first large-scale view of the gendered nature of contemporary science production. Our findings are limited, however, by the small range of the study time frame and the singular publisher from which we gathered data. The use of the first publication date as a proxy for academic age also introduces some imprecision into the age measurement (further complicated by the lack of 2013 data). Furthermore, given the newness of contribution statements, we require additional studies verifying the degree to which the stated contributions accurately reflect the work in the lab.

Future research depends on the availability of high-quality data on contributorship. Such a change will take a concerted effort on the part of journal editors, science policy makers, and scholars to advocate for and implement this reporting into the scholarly communication system. This is also an imperative for training—educating the next generation of scholars on equitable systems for distribution of labor and allocation of credit in scientific work. Replacing authorship with contributorship not only illuminates potential disparities in the scientific workforce but may also mitigate scientific malpractice as scientists would be required to be more explicit about their roles, potentially lessening the opportunities for misappropriation of credit in scientific work. This is an important intervention point for those educating people in academic medicine. As authorship has changed, so too must our training, metrics, and documentation standards.

Funding/Support: The authors acknowledge funding from the Social Science and Humanities Research Council of Canada as well as from the Canada Research Chairs program.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.

B. Macaluso is researcher, Observatoire des sciences et des technologies, Université du Québec à Montréal, Montréal, Québec, Canada.

V. Larivière is Canada Research Chair on the Transformations of Scholarly Communication, Université de Montréal, scientific director, Érudit journal platform, and associate scientific director, Observatoire des sciences et des technologies, Université du Québec à Montréal, Montréal, Québec, Canada.

T. Sugimoto is researcher, Center for Education and Evaluation Policy, Indiana University Bloomington, Bloomington, Indiana.

C.R. Sugimoto is associate professor, School of Informatics and Computing, Indiana University Bloomington, Bloomington, Indiana.

References

- Rossiter MW. Women Scientists in America: Struggles and Strategies to 1940. Baltimore, Md: Johns Hopkins University Press; 1982.
- Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto CR. Bibliometrics: Global gender disparities in science. *Nature*. 2013;504:211–213.
- Merton RK. The Sociology of Science: Theoretical and Empirical Investigations. Chicago, Ill: University of Chicago Press; 1973.
- Biagoli M, Galison P. Scientific Authorship: Credit and Intellectual Property in Science. New York, NY: Routledge; 2002.
- International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. *JAMA*. 1993;269:2282–2286.
- Mowatt G, Shirran L, Grimshaw JM, et al. Prevalence of honorary and ghost authorship in Cochrane reviews. *JAMA*. 2002;287:2769–2771.
- Bates T, Anić A, Marusić M, Marusić A. Authorship criteria and disclosure of contributions: Comparison of 3 general medical journals with different author contribution forms. *JAMA*. 2004;292:86–88.
- Cronin B. Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *J Am Soc Inf Sci Technol*. 2001;52:558–569.
- Rennie D, Yank V, Emanuel L. When authorship fails. A proposal to make contributors accountable. *JAMA*. 1997;278:579–585.
- Caron E, van Eck NJ. Large scale author name disambiguation using rule-based scoring and clustering. In: Proceedings of the 19th International Conference on Science and Technology Indicators: Context Counts: Pathways to Master Big Data and Little Data. Leiden, the Netherlands: Centre for Science and Technology Studies-Leiden University; 2014:79–86.
- West JD, Jacquet J, King MM, Correll SJ, Bergstrom CT. The role of gender in scholarly authorship. *PLoS One*. 2013;8:e66212.
- Jagsi R, Guancial EA, Worobey CC, et al. The “gender gap” in authorship of academic medical literature—a 35-year perspective. *N Engl J Med*. 2006; 355:281–287.
- Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. *Science*. 2007;316:1036–1039.
- Jones B. The burden of knowledge and the “death of the Renaissance man”: Is innovation getting harder? *Rev Econ Stud*. 2009;76:283–317.
- Fang FC, Bennett JW, Casadevall A. Males are overrepresented among life science researchers committing scientific misconduct. *MBio*. 2013;4:e00640–e00612.

Appendix 1

Regression Analysis Results, From a Study of Author Contributorship Roles and Gender, 2008–2013

Variable	B	SE	P value	Odds
Dependent variable: Analysis^a				
Intercept	0.4069	0.6275	.5167	
Female	−0.1072	0.0104	< .0001	0.898346
Unknown gender	−0.0924	0.0143	< .0001	0.91174
First author	2.3825	0.0555	< .0001	10.83195
Last author	0.7962	0.0125	< .0001	2.2171
Corresponding author	0.3042	0.3042	< .0001	1.35554
Academic age	−0.0121	0.000452	< .0001	0.987973
Number of authors	−0.0302	0.000694	< .0001	0.970251
Percent authors female	0.1262	0.0222	< .0001	1.134509
Country			< .0001	
Field			< .0001	
Publication year			< .0001	
Dependent variable: Design^b				
Intercept	2.0272	0.8175	.0131	
Female	−0.1729	0.0113	< .0001	0.841222
Unknown	−0.1737	0.0154	< .0001	0.840549
First author	1.683	0.0535	< .0001	5.381677
Last author	2.3855	0.0171	< .0001	10.86449
Corresponding author	0.7458	0.7458	< .0001	2.108127
Academic age	0.0431	0.000489	< .0001	1.044042
Number of authors	−0.0393	0.000811	< .0001	0.961462
Percent authors female	0.0699	0.0241	.0037	1.072401
Country			< .0001	
Field			< .0001	
Publication year			< .0001	

(Appendix continues)

Appendix 1

(Continued)

Variable	B	SE	P value	Odds
Dependent variable: Materials^c				
Intercept	-01.4909	0.7699	.0528	
Female	-0.1728	0.0102	< .0001	0.841306
Unknown	-0.0895	0.0138	< .0001	0.914388
First author	-0.1049	0.0451	< .0001	0.900415
Last author	-0.1649	0.0124	< .0001	0.847978
Corresponding author	0.0858	0.0858	.0551	1.089588
Academic age	0.0291	0.000445	< .0001	1.029528
Number of authors	-0.00174	0.000558	.0018	0.998262
Percent authors female	-0.1381	0.0214	.0037	0.871012
Country			< .0001	
Field			< .0001	
Publication year			< .0001	
Dependent variable: Perform experiment^d				
Intercept	-0.7563	0.6072	.2129	
Female	0.4205	0.0108	< .0001	1.522723
Unknown	0.212	0.0148	< .0001	1.236148
First author	2.0301	0.0605	< .0001	7.614848
Last author	-0.6188	0.0143	< .0001	0.53859
Corresponding author	-0.3874	0.0598	< .0001	0.67882
Academic age	-0.0825	0.000525	< .0001	0.920811
Number of authors	-0.0127	0.000629	< .0001	0.98738
Percent authors female	0.0826	0.0234	.0004	1.086107
Country			< .0001	
Field			< .0001	
Publication year			< .0001	
Dependent variable: Write^e				
Intercept	2.8428	1.0728	.0081	
Female	-0.078	0.0117	< .0001	0.924964
Unknown	-0.1794	0.0161	< .0001	0.835772
First author	2.5613	0.0593	< .0001	12.95264
Last author	2.3059	0.016	< .0001	10.0332
Corresponding author	0.706	0.0594	< .0001	2.025872
Academic age	0.0321	0.000496	< .0001	1.032621
Number of authors	-0.044	0.00087	< .0001	0.956954
Percent authors female	-0.2813	0.0249	< .0001	0.754802
Country			< .0001	
Field			< .0001	
Publication year			< .0001	

^aCox and Snell pseudo- R^2 : 0.1640.^bCox and Snell pseudo- R^2 : 0.2688.^cCox and Snell pseudo- R^2 : 0.0269.^dCox and Snell pseudo- R^2 : 0.2359.^eCox and Snell pseudo- R^2 : 0.3081.