US environmental science women are high impact team players¹

Mike A. Thelwall, PhD, University of Wolverhampton, UK.

In the context of continuing gender inequality in environmental science, this article investigates the evolution of the share of female first-authored journal articles in US environmental science specialisms 1996-2018 and whether there are gender differences in average citation impact. The proportion of female first authors has increased by at least 16% in all 11 Scopus environmental science categories in the USA, with none achieving a female majority by 2018. Female last authors were rarer and increased less during the same period. There is a citation (ratio) advantage for female first-authored research of 2% to 10% that could be partly accounted for by females tending to lead larger teams (up to 8% if team size is factored out). Irrespective of main author gender, extra authors produce expected citation rate increases of between 6% for an Ecology second author and 70% for at least five Waste Management and Disposal authors. The female citation advantage in one field might be substantial enough to boost their careers, although collaboration is a much stronger indicator of greater impact.

Introduction

Environmental science seems to have an average level of gender inequality for an academic discipline in the USA but with a consistent male majority overall (see below). For US first-authored research in 2017, Environmental Science was the 12th most female subject out of 26, with a male majority in all specialist fields. This varied from 0.43 females per male in Water Science & Technology to 0.52 females per male in Health, Toxicology & Mutagenesis (Thelwall, Bailey, Tobin, and Bradshaw, 2019). Environmental science seems to be rarely singled out for Science Technology, Engineering and Mathematics (STEM) initiatives to promote female participation, presumably because other subjects are in a worse situation, but there are partial exceptions (Horner-Devine, Yen, Mody-Pan, Margherio, and Forde, 2016; Thornbush, 2016). Explicit gender bias does not seem to be a major cause of female underrepresentation in science overall (Ceci and Williams, 2011) so calls to address it (Grogan, 2019) may not have large effects and subject-based factors should also be explored.

A few analyses of environmental science or related areas have mentioned gender information, but with insufficient coverage to give an overall picture (Sobratee and Slotow, 2019; Bullock, Kirchhoff, Mauro, and Boerchers, 2018; Branch and Linnell, 2016; Swihart, Sundaram, Höök, and Dewoody, 2016; Mund, Kloft, Bundschuh, Klingelhoefer, Groneberg, and Gerber, 2014; Fox, Ritchey, and Paine, 2018; Fox and Paine, 2019). The use of citations has been opposed on the basis that they disadvantage women in environmental science because they are less likely to write review articles and more likely to conduct empirical rather than theoretical research, which is slower to create (Downes and Lancaster, 2019).

This article investigates whether there are gender differences in citation impact that might help to explain the relative success of US female environmental scientists. Citations are sometimes used in appointment, promotions, tenure and funding decisions, so gender differences in citation rates might partly account for gender differences in participation rates. There are international differences in research success (e.g., Burns, and Fox, 2017) and so a focus on a single large country helps to reduce sources of variation in the results. Gender can be a factor within environmental science teamwork (Djenontin and Meadow, 2018), and team

¹ Journal of Environmental Engineering and Science, <https://doi.org/10.1680/jenes.19.00025>

size can affect research impact in general (Larivière, Gingras, Sugimoto, and Tsou, 2015) and so this was included in the analysis.

Methods

The research design was to use regression to identify whether first and last author genders influenced the citation impact of environmental science research in multiple fields, and to account for the role of team size. First and last authors tend to conduct the most tasks in a study (Larivière, Desrochers, Macaluso, Mongeon, Paul-Hus, and Sugimoto, 2016) and so the focus is on these roles. Alphabetical ordering and equal contribution lists (Broderick and Casadevall, 2019) introduce noise, when used. Data and regression details are given in the Appendix.

Results

All Environmental Science specialisms experienced substantial increases in the proportion of female first-authored articles in the US from 1996 to 2018, from a 16% to a 25% increase (Table 1). Despite these universal increases, only one specialism approached gender parity, Health, Toxicology and Mutagenesis (49.9%). The lowest proportion of females in 2018 was for Environmental Engineering (33%), perhaps because Engineering is a traditionally male topic. Female last author shares did not increase as much, between 12% and 19%, and reached lower levels in each field, with the highest share being 37% (Health, Toxicology and Mutagenesis). This presumably is due to a greater share of senior males in the fields and senior authors being more likely to be last authors.

Table 1. Gender identification and share statistics for the environmental science Scopus narrow fields 1996-2018.

Narrow field	Articles with US only authors	Gendered first & last authors %	Female first in 1996 %	Female first in 2018 %	Change 1996-2018 %	Female last in 1996 %	Female last in 2018 %	Change 1996-2018 %
Ecological Modeling	7427	58.7	9.9	30.4	20.5	12.8	22.9	10.2
Ecology	93931	62.8	20.5	39.9	19.4	16.6	27.2	10.6
Environmental Chemistry	48014	55.5	18.8	36.8	17.9	14.3	27.2	13.0
Environmental Engineering	23556	51.6	15.1	33.2	18.2	13.3	26.1	12.8
Global and Planetary Change	11223	66.2	15.3	40.2	25.0	12.1	29.4	17.3
Health, Toxicology & Mutagenesis	30491	59.6	26.1	49.9	23.8	16.1	37.2	21.2
Manag., Monitoring, Policy & Law	37193	68.1	16.1	36.3	20.2	15.3	30.2	14.9
Nature & Landscape Conservation	22289	72.8	19.4	41.1	21.7	16.6	28.7	12.0
Pollution	28832	53.8	18.6	36.2	17.6	18.6	26.6	8.0
Waste Management & Disposal	19329	53.7	19.4	35.1	15.7	18.6	28.6	10.1
Water Science and Technology	53577	50.7	14.5	30.6	16.1	13.1	23.5	10.4

Female authors tend to lead larger teams as first or last authors in two fields: Environmental Engineering, and Health, Toxicology and Mutagenesis (Table 2). The average team size for female first authors is larger in all fields except one, but only statistically significant in two.

Table 2. Average (geometric mean) team sizes 2018 for Scopus narrow Environmental Science fields (US only) by first and last author gender.

Narrow field	Articles	Female first team size	Male first team size	Female last team size	Male last team size
Ecological Modeling	240	3.45	3.12	3.47	3.15
Ecology	3115	3.27	3.15	3.29	3.16
Environmental Chemistry	1311	3.96	3.77	3.82	3.84
Environmental Engineering	716	3.79***	2.98	3.78***	3.05
Global and Planetary Change	579	2.80	2.92	2.84	2.88
Health, Toxicology and Mutagenesis	1133	4.50***	3.74	4.38*	3.95
Management, Monitoring, Policy and Law	1698	2.55	2.40	2.46	2.45
Nature and Landscape Conservation	1141	3.09	3.04	3.14	3.03
Pollution	985	3.79	3.53	3.71	3.59
Waste Management and Disposal	618	3.68	3.42	3.49	3.51
Water Science and Technology	1469	3.17	2.98	3.18	3.00

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ Wilcoxon male vs. female tests; larger values are starred

For the extended regression model including team size factors, there is a statistically significant female first author advantage in 5/11 fields and a statistically significant female last author advantage in 1/11 fields. The largest advantage is 7.6%, for Health, Toxicology and Mutagenesis female first authors. There is a statistically significant female last author disadvantage in three fields, with a maximum of 4.2%, in Water Science and Technology (Table 3).

Additional authors statistically significantly boost expected citation rates in all fields, and this boost is almost always substantial and greater than any gender advantage. The main exception is that a second author provides only a 5.9% increase in the expected citation rate (ratio of logged citations to the world average logged citations for the publishing field and year) for Ecology. The largest increase is for Waste Management and Disposal, with a second author associating with a 53.9% higher expected citation ratio and at least five authors associating with a 70.1% higher expected citation ratio (Table 3).

Table 3. Regression coefficients for the full regression model for Scopus narrow Environmental Science fields (US only, first and last author gendered). Publication year regression coefficients are not shown.

	Model 1					
Narrow field	Female first	Female last	2 auth	3 auth	4 auth	5+ auth
Ecological Modeling	-0.007	0.009	0.203***	0.120***	0.252***	0.303***
Ecology	0.026***	0.009	0.059***	0.144***	0.145***	0.224***
Environmental Chemistry	0.018**	0.008	0.157***	0.226***	0.214***	0.275***
Environmental Engineering	0.030*	-0.018	0.418***	0.540***	0.495***	0.542***
Global and Planetary Change	0.019	0.020	0.199***	0.298***	0.294***	0.383***
Health, Toxicology and Mutagenesis	0.076***	0.028**	0.194***	0.182***	0.236***	0.364***
Management, Monitoring, Policy & Law	-0.011	-0.002	0.209***	0.216***	0.236***	0.284***
Nature and Landscape Conservation	-0.006	-0.008	0.139***	0.153***	0.164***	0.208***
Pollution	0.030**	-0.031**	0.447***	0.519***	0.519***	0.597***
Waste Management and Disposal	0.017	-0.035*	0.549***	0.634***	0.637***	0.701***
Water Science and Technology	-0.001	-0.042***	0.336***	0.471***	0.422***	0.475***

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Removing author numbers from the regression equations equates to attributing team recruitment to the actions of the first and/or last author, rather than treating it as an independent variable. For reduced model 2, there is a statistically significant female first author advantage in 8/11 fields and a statistically significant female last author advantage in 1/11 fields. The largest advantage is 10.0%, for Health, Toxicology and Mutagenesis female first authors. There is a statistically significant female last author disadvantage in three fields, with a maximum of 4.8%, in Water Science and Technology (Table 4). Treating first and last authors separately (Models 3 and 4) makes little difference. Thus, if first and/or last authors are credited with team sizes as their achievement, the female first author citation advantage tends to be stronger and more likely to be statistically significant.

Table 4. Regression coefficients for three reduced regression models for Scopus narrow Environmental Science fields (US only, first and last author gendered), ignoring team size. Publication year regression coefficients are not shown.

	Model 2		Model 3	Model 4
Narrow field	Female 1st	Female last	Female 1st	Female last
Ecological Modeling	0.009	0.019	0.011	0.020
Ecology	0.030***	0.007	0.031***	0.013*
Environmental Chemistry	0.028***	0.009	0.029***	0.013
Environmental Engineering	0.063***	-0.010	0.061***	0.002
Global and Planetary Change	0.033*	0.017	0.037*	0.026
Health, Toxicology and Mutagenesis	0.100***	0.029**	0.104***	0.044***
Management, Monitoring, Policy & Law	-0.005	-0.016	-0.010	-0.017
Nature and Landscape Conservation	-0.004	-0.012	-0.006	-0.013
Pollution	0.060***	-0.027*	0.056***	-0.017
Waste Management and Disposal	0.043**	-0.048**	0.035*	-0.039*
Water Science and Technology	0.027**	-0.033**	0.022*	-0.028**

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Conclusions

Several factors in the data and analysis steps may have affected the outcomes. The gender identification algorithm may have led to the exclusion of minority cultures in the USA and those with gender neutral names, as well as people that uses gender neutral versions of their name in their publications (e.g., Pat). Articles may be misclassified if they are in a multidisciplinary journal or a journal that does not reflect their core contribution. The Scopus categories may not fully reflect the nuances of academic research fields. Second order effects, such as seniority, may affect gender differences. Team size effects may work differently from modelled here.

The results point to a small degree of sexual dimorphism in average citation impact in environmental science, favouring females, despite the male majority in all fields. This provides a broader picture than previous bibliometric analyses of small environmental science specialisms and finer grained results than previous science-wide studies. The female citation advantage is probably too small to have a noticeable impact on academic careers, however, with the possible exception of Health, Toxicology and Mutagenesis (up to 10%).

Appendix

Data

The eleven specialist environmental science categories in Scopus were downloaded, covering records from 1996 to 2018, during November 2018 to January 2019. Scopus has wider coverage than the logical alternative, the Web of Science (Mongeon & Paul-Hus, 2016). Records were removed if any author affiliation country recorded was not the US or if the document type was not a journal article (also excluding reviews) to get a standard baseline. The number of authors was recorded by counting the number of distinct author IDs for each article. The first name of the first and last authors was also extracted, along with a list of the fields in which the article was classified (normally based on its publishing journal). The gender of the first and last author was guessed by matching their first name, when reported, to a list of first names commonly used by males or females in the USA, as reported in census data and by genderapi.com. Testing on human checked academic authors from 2017 from another paper (Thelwall, Bailey, Tobin, & Bradshaw, 2019) shows that this method has an accuracy of 98.5% but leaves some authors unclassified and tends to overestimate females by 0.5%. Articles were excluded if the gender of the first or last author could not be detected.

For the citation analysis, papers published after 2014 were discarded to give at least three full years for each article to be cited, which is usually adequate (Abramo, Cicero, & D'Angelo, 2011). Citation counts cannot be directly compared between years because older articles have an unfair advantage. They should not be compared between fields because high citation specialisms have an advantage. They also should not be averaged because individual highly cited articles can dominate the results. To solve these problems, citation counts were replaced by field and year normalised citation ratios, NLCS (Normalised Log-transformed Citation Scores), as follows (Thelwall, 2017). First, the citation count c for each article was replaced by a log-transformed version $l = \ln(1 + c)$ to reduce the influence of individual highly cited articles. Second, the log-transformed citation score l was replaced by a field and year normalised version, $n = l/f$, where f is the average of all l for articles in the same field and year as c . If the article is within multiple categories, then these field/year averages are themselves averaged and the field and year normalised formula is $n = l/\bar{f}$ where the denominator is the average of the field averages of all fields containing the article. In either case, this is the NLCS.

Analysis

Gender associations with team size were assessed using Wilcoxon tests, which are appropriate for author number data even though it is skewed. For the same reason, average author numbers are reported as geometric means rather than arithmetic means.

Overall gender and team size trends were assessed using regression rather than averaging across all the data because there were substantial differences over time in the average impact of US research. The field and year normalised citation count NLCS for each article was used as the dependent variable. The main independent variables were first author gender F and last author gender L , both binary. Average NLCS vary over time with changes in the coverage of Scopus and the relative impact of US research relative to the world average and so additional binary variables P_y were added for each publication year (using 1996 as the default) to prevent changes over time from interacting with gender. Whilst this could be addressed through multilevel modelling with additional assumptions, there was enough data to accommodate the necessary reduced degrees of freedom without much loss of fitting power and this is therefore preferable to the additional assumptions necessary for multilevel modelling. The regression equation was therefore as follows, fitted separately to each narrow subject area.

$$NLCS = \alpha + \beta_1 F + \beta_2 L + \beta_3 P_{97} + \dots + \beta_{18} P_{12} \quad (1)$$

The above equation was also reduced by removing the first or last authors in order to assess the influence of these authorship positions separately. An expanded regression equation was also made by including team size binary variables T_2, T_3, T_4, T_{5+} for articles with 2, 3, 4 and 5+ authors, respectively. These four alternative equations allow model triangulation. Other factors are known to systematically influence citation counts, such as the choice of an appropriate title (Fox & Burns, 2015), but these tend to be under the control of the authors and so are not relevant as additional independent variables.

The regression was fitted using a weighted least squares algorithm that can cope with the heteroscedasticity caused by variations in the spread of normalised citation counts over time (Rousseeuw, Croux, Todorov, Ruckstuhl, Salibian-Barrera, Verbeke, & Maechler, 2019; Yohai, 1987). No outliers were found.

References

- Abramo G, Cicero T, and D'Angelo CA, (2011) Assessing the varying level of impact measurement accuracy as a function of the citation window length. *Journal of Informetrics* **5(4)**: 659-667.
- Bendels MH, Müller R, Brueggmann D and Groneberg DA (2018) Gender disparities in high-quality research revealed by Nature Index journals. *PloS ONE* **13(1)**: e0189136.
- Branch TA and Linnell AE (2016) What makes some fisheries references highly cited? *Fish and Fisheries* **17(4)**: 1094-1133.
- Broderick NA and Casadevall A (2019) Gender inequalities among authors who contributed equally. *eLife* **8**: e36399.
- Bullock R Kirchhoff D Mauro I and Boerchers M (2018) Indigenous capacity for collaboration in Canada's energy forestry and mining sectors: research metrics and trends. *Environment Development and Sustainability* **20(2)**: 883-895.
- Burns CS and Fox CW (2017) Language and socioeconomics predict geographic variation in peer review outcomes at an ecology journal. *Scientometrics* **113(2)**: 1113-1127.
- Ceci SJ and Williams WM (2011) Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences* **108(8)**: 3157-3162.

- Djenontin INS and Meadow AM (2018) The art of co-production of knowledge in environmental sciences and management: Lessons from international practice. *Environmental Management* **61(6)**: 885-903.
- Downes BJ and Lancaster J (2019) Celebrating women conducting research in freshwater ecology... and how the citation game is damaging them. *Marine and Freshwater Research* MF18436 <https://doi.org/10.1071/MF18436>
- Fox CW and Burns CS (2015) The relationship between manuscript title structure and success: editorial decisions and citation performance for an ecological journal. *Ecology and Evolution* **5(10)**: 1970-1980.
- Fox CW and Paine CT (2019) Gender differences in peer review outcomes and manuscript impact at six journals of ecology and evolution *Ecology and Evolution* **9**: 3599-3619
- Fox CW, Ritchey JP and Paine CT (2018) Patterns of authorship in ecology and evolution: First last and corresponding authorship vary with gender and geography. *Ecology and Evolution* **8(23)**: 11492-11507.
- Grogan KE (2019) How the entire scientific community can confront gender bias in the workplace. *Nature Ecology and Evolution* **3(1)**: 3.
- Horner-Devine MC, Yen JW, Mody-Pan PN, Margherio C and Forde S (2016) Beyond traditional scientific training: The importance of community and empowerment for women in ecology and evolutionary biology. *Frontiers in Ecology and Evolution* **4**: 119 <https://doi.org/10.3389/fevo.2016.00119>
- Larivière V, Desrochers N, Macaluso B, Mongeon P, Paul-Hus A and Sugimoto CR (2016) Contributorship and division of labor in knowledge production. *Social Studies of Science* **46(3)**: 417-435.
- Larivière V, Gingras Y, Sugimoto CR and Tsou A (2015) Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology* **66(7)**: 1323-1332.
- Mongeon P and Paul-Hus A (2016) The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics* 106(1) 213-228.
- Mund M, Kloft B, Bundschuh M, Klingelhofer D, Groneberg D and Gerber A (2014) Global research on smoking and pregnancy - A scientometric and gender analysis. *International Journal of Environmental Research and Public Health* **11(6)**: 5792-5806.
- Rousseeuw P, Croux C, Todorov V, Ruckstuhl A, Salibián-Barrera M, Verbeke T and Maechler M (2019) robustbase: Basic Robust Statistics R package version 092-8 .
- Sobratee N and Slotow R (2019) A critical review of lion research in South Africa: The impact of researcher perspective research mode and power structures on outcome bias and implementation gaps. *Frontiers in Ecological Evolution* **12(4)**: 81 <https://doi.org/10.3389/fevo.2019.00081>
- Swihart RK, Sundaram M, Höök TO and Dewoody JA (2016) Factors affecting scholarly performance by wildlife and fisheries faculty. *The Journal of Wildlife Management* **80(3)**: 563-572.
- Thelwall M, Bailey C, Tobin C and Bradshaw N (2019) Gender differences in research areas methods and topics: Can people and thing orientations explain the results? *Journal of Informetrics* **13(1)**: 149-169.
- Thelwall M (2017) Three practical field normalised alternative indicator formulae for research evaluation. *Journal of Informetrics* **11(10)**: 128–151.
- Thornbush M (2016) Introduction to the special issue on gender and geoethics in the geosciences. *International Journal of Environmental Research and Public Health* **13**: 39. doi:10.3390/ijerph13040398

Yohai VJ (1987) High breakdown-point and high efficiency robust estimates for regression.
The Annals of Statistics **15(2)**: 642-656.