Long-term Analysis of Sex Differences in Prestigious Authorships in Cardiovascular Research Supported by the National Institutes of Health

Online Supplemental Material

Carolin Lerchenmüller, MD[1,2]*; Marc J. Lerchenmueller, PhD, MPH[3]*;

Olav Sorenson, PhD[3]

[1]Cardiovascular Research Center, Massachusetts General Hospital and

Harvard Medical School, Boston, MA 02114

[2]Department of Cardiology, Angiology and Pulmonology, University Hospital Heidelberg,

69120 Heidelberg, Germany

[3]Yale University, School of Management, 165 Whitney Avenue, New Haven, CT 06511

* Co-corresponding authors. Author order alphabetical.

Short Title: Gender Gap in Cardiovascular Research

Correspondence to:

Carolin Lerchenmüller

185 Cambridge Street

Boston, MA 02114

+1 617 643 3473

clerchenmueller@mgh.harvard.edu

Marc J. Lerchenmueller

165 Whitney Avenue

New Haven, CT 06511

+1 203 606 0743

marc.lerchenmueller@yale.edu

Online Supplemental Material

Supplemental Methods

Supplemental Table 1: Cardiovascular journals ranked by five year journal impact factor

Supplemental Table 2: Sensitivity analysis of RR estimates for different samples – cardiovascular research

Supplemental Table 3: Sensitivity analysis of RR estimates for different samples – life sciences

Supplemental Figure 1: RR of prestigious authorship across the life sciences per year

Supplemental Methods

Data:

We analyzed the relationship between sex and authorship order, both among R01-supported articles published in cardiovascular journals and R01-supported articles across the life sciences that were recorded in the *PubMed* database. We used two author disambiguation algorithms that allocate authors with their forenames (for sex identification) to articles indexed in *PubMed*. We started with the Author-ity[1] algorithm, which includes all papers published before 2009 and has been shown to identify authors with greater than 99 percent accuracy.[2] We then used the Scopus algorithm for articles published from 2010 up to 2015, which has been found to identify authors across publications with accuracy similar to Author-ity.[2, 3]

To determine the probable sex of authors, we wrote a program[4] in R[5] that calls the *Genderize* algorithm. *Genderize* associates an individual's forename with the probability of being male and female.[6] For example, the algorithm designates "Chris" as male with 93% confidence, based on 8,631 verified records in the *Genderize* database. For our analyses, we considered sex determined if the assigned probability met or exceeded a 90% threshold. Comparisons of *Genderize* to other automated algorithms indicate that it provides the most accurate sex assignments.[7]

Finally, we retrieved the 5-year JIFs from the Journal Citation Report (Thomson Reuters, 2015). Note that 5-year JIFs are relatively stable over time – the 2007 scores, the first year available, had a correlation of 0.91 with the 2015 scores in our data. We assigned journals not listed in the Journal Citation Report to the lowest impact strata as the citation report sets a minimum impact threshold for inclusion in the index.

Sensitivity analysis:

Although our main sample for analysis only included articles with at least three authors, we evaluated the robustness of our results in several additional samples, for cardiovascular research (Supplemental Table 2) and across the life sciences (Supplemental Table 3). First, we dropped articles with alphabetical authorship as these articles may signal an equal allocation of credit, with the first and last authorship positions carrying no additional prestige. Second, we only considered articles for which we could designate the gender of all authors to test whether missing information on the sex of some authors might influence our results. Third, we relaxed all sampling restrictions and considered all articles in our relative risk (RR) calculations (including those authored by one or two authors). Across these different samples, our main results remained substantively unchanged.

We also calculated RRs using different confidence thresholds for designating authors' gender. These different thresholds influenced our results by a maximum of two percentage points.

As only a few journals have begun to require the publication of author contributions, we could not assess the contributions of interior authors to the publications across the 3,849 journals included in our analysis. If interior authors tend to be scientific assistants rather than individuals interested in pursuing an academic career and those assistants are disproportionally female, our approach might underestimate women's representation in prestigious authorships. In other words, the higher RR of women earning first authorships would increase, in favor of women. The gap in last authorship, however, appears so large that it seems unlikely that accounting for differential contributions of interior authors would materially impact our results on last authorship.

Finally, we also considered an alternative definition of RR, including last authors in the denominator for the rates of both first and last authorship. This definition does, however, inflate the RR of first authorship in favor of women due to their lower number of last authorships relative to men. Since it would appear unlikely that many senior authors would defer their last authorship for the first author position, our more conservative RR estimate presented in the main manuscript represents the more appropriate calculation, conceptually and empirically.

Confidence intervals and statistical significance:

To assess statistical significance, we calculated 95% confidence intervals for the relative risk of prestigious authorships for women versus men. Specifically, we calculated binomial ratio confidence limits as follows: Given two binomial proportions, p1 and p2, emerging from samples with respective sizes n1 and n2, the formula for the confidence interval for the ratio p1 / p2 is:

$$\frac{p_1}{p_2} e^{\pm \Phi^{-1}(\alpha/2)\sqrt{(1-p_1)/(n_1 p_1)+(1-p_2)/(n_2 p_2)}}$$

with Φ^{-1} denoting the normal percent point function. To address the (unlikely) event that either proportion p1 or p2 are close to zero in our samples (leading to a division by zero in the above formula), we calculated the Bayes estimators of p1 and p2 with F and M denoting, for example, the number of female and male first authorships as follows:

$$p_1 = \frac{F + 0.5}{n_1 + 1} \qquad\qquad p_2 = \frac{M + 0.5}{n_2 + 1}$$

Supplemental Table 1: Cardiovascular journals ranked by journal impact factor

Item	Journal Title	Five year impact factor (JCR 2015)
1	Circulation	16.2
2	Journal Of The American College Of Cardiology	15.7
3	European Heart Journal	15.7
4	Circulation Research	11.2
5	Nature Reviews Cardiology	10.5
6	Jacc-Cardiovascular Imaging	7.3
7	Jacc-Cardiovascular Interventions	7.2
8	Jacc-Heart Failure	7.2
9	Circulation-Heart Failure	7.1
10	Circulation-Cardiovascular Interventions	6.5
11	Circulation-Cardiovascular Imaging	6.3
12	Circulation-Cardiovascular Quality And Outcomes	5.8
13	Cardiovascular Research	5.8
14	European Journal Of Heart Failure	5.7
15	Journal Of Heart And Lung Transplantation	5.6
16	Journal Of Cardiovascular Magnetic Resonance	5.4
17	Basic Research In Cardiology	5.4
18	Journal Of The American Heart Association	5.2
19	Circulation-Arrhythmia And Electrophysiology	5.1
20	Heart	4.9
21	Journal Of The American Society Of Echocardiography	4.9
22	Journal Of Molecular And Cellular Cardiology	4.8
23	Circulation-Cardiovascular Genetics	4.6
24	International Journal Of Cardiology	4.4
25	American Heart Journal	4.3
26	Heart Rhythm	4.2
27	European Heart Journal-Cardiovascular Imaging	4.1
28	Cardiovascular Diabetology	3.9
29	Progress In Cardiovascular Diseases	3.8
30	Journal Of Cardiac Failure	3.6
31	American Journal Of Physiology-Heart And Circulatory Ph	3.5
32	Clinical Research In Cardiology	3.5
33	European Journal Of Preventive Cardiology	3.5
34	Heart Failure Reviews	3.5
35	Journal Of Thoracic And Cardiovascular Surgery	3.5
36	Nutrition Metabolism And Cardiovascular Diseases	3.4
37	Circulation Journal	3.3
38	American Journal Of Cardiology	3.3
39	Eurointervention	3.2
40	Europace	3.2
41	Respiratory Medicine	3.2
42	Revista Espanola De Cardiologia	3.1
43	Journal Of Cardiovascular Translational Research	3.0
44	Canadian Journal Of Cardiology	3.0
45	Cardiovascular Drugs And Therapy	2.9
46	Journal Of Cardiovascular Electrophysiology	2.8
47	Journal Of Cardiovascular Computed Tomography	2.8
48	Trends In Cardiovascular Medicine	2.7
49	Current Problems In Cardiology	2.7
50	European Journal Of Cardio-Thoracic Surgery	2.7

Supplemental Table 1 (continued)

Item	Journal Title	Five year impact factor (JCR 2015)
51	Journal Of Nuclear Cardiology	2.5
52	Cardiology In Review	2.4
53	Cardiovascular Therapeutics	2.4
54	Archives Of Cardiovascular Diseases	2.4
55	Journal Of Cardiovascular Pharmacology And Therapeutic	2.3
56	European Journal Of Cardiovascular Nursing	2.3
57	Journal Of Cardiovascular Pharmacology	2.3
58	Clinical Cardiology	2.2
59	Cardiovascular Pathology	2.2
60	Cardiovascular Toxicology	2.1
61	Current Opinion In Cardiology	2.1
62	Journal Of Cardiovascular Nursing	2.1
63	Cardiovascular And Interventional Radiology	2.1
64	Journal Of Cardiology	2.0
65	Catheterization And Cardiovascular Interventions	2.0
66	American Journal Of Cardiovascular Drugs	2.0
67	Bmc Cardiovascular Disorders	2.0
68	Journal Of Cardiopulmonary Rehabilitation And Preventic	1.9
69	Heart And Vessels	1.8
70	International Journal Of Cardiovascular Imaging	1.7
71	Cardiovascular Ultrasound	1.6
72	Journal Of Cardiothoracic And Vascular Anesthesia	1.5
73	Journal Of Interventional Cardiac Electrophysiology	1.5
74	Heart & Lung	1.5
75	International Heart Journal	1.5
76	Coronary Artery Disease	1.4
77	Annals Of Noninvasive Electrocardiology	1.4
78	Pediatric Cardiology	1.3
79	Heart Lung And Circulation	1.3
80	Journal Of Cardiovascular Medicine	1.3
81	Pace-Pacing And Clinical Electrophysiology	1.3
82	Cardiology Clinics	1.3
83	Echocardiography-A Journal Of Cardiovascular Ultrasoun	1.3
84	Cardiology Journal	1.2
85	Journal Of Interventional Cardiology	1.2
86	Journal Of Cardiovascular Surgery	1.2
87	Perfusion-UK	1.2
88	Congenital Heart Disease	1.2
89	Journal Of Electrocardiology	1.2
90	Arquivos Brasileiros De Cardiologia	1.2
91	Interactive Cardiovascular And Thoracic Surgery	1.2
92	Journal Of Cardiothoracic Surgery	1.1
93	Scandinavian Cardiovascular Journal	1.1
94	Journal Of Invasive Cardiology	1.1
95	Journal Of Cardiac Surgery	0.9
96	Hellenic Journal Of Cardiology	0.9
97	Thoracic And Cardiovascular Surgeon	0.8
98	Cardiology In The Young	0.8
99	Journal Of Heart Valve Disease	0.8
100	Reviews In Cardiovascular Medicine	0.8
101	Texas Heart Institute Journal	0.7
102	Herz	0.7
103	Revista Portuguesa De Cardiologia	0.6
104	Minerva Cardioangiologica	0.5
105	Heart Surgery Forum	0.4
106	Current Cardiology Reports	0.0
107	Heart Failure Clinics	0.0

Supplemental Table 2: Sensitivity analysis of RR estimates for different samples – cardiovascular research

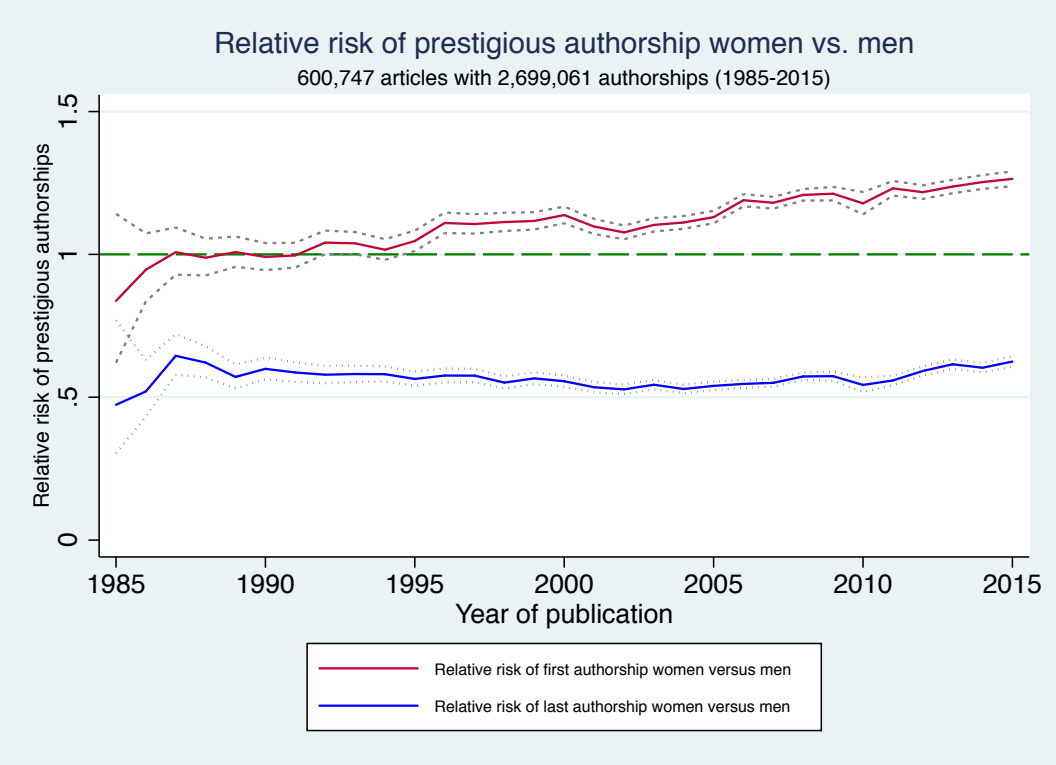
			95% CI			95% CI	
	Nr. of articles	RR first authorship	Lower bound	Upper bound	RR last authorship	Lower bound	Upper bound
Results from main text (for comparison)							
1985–1994	1,066	0.79	0.66	0.96	0.41	0.32	0.55
1995–2004	4,024	0.96	0.90	1.03	0.53	0.48	0.58
2005–2015	6,928	1.05	1.01	1.11	0.52	0.48	0.55
Excl. articles with alphabetical authorship							
1985–1994	1,003	0.84	0.70	1.02	0.41	0.32	0.55
1995–2004	3,876	0.96	0.89	1.03	0.52	0.48	0.58
2005–2015	6,754	1.06	1.01	1.11	0.51	0.48	0.55
Excl. articles with missing gender of authors							
1985–2015*	2,675	1.03	0.96	1.09	0.48	0.44	0.53
Including articles with single and dual authorship(s)							
1985–1994	1,320	0.75	0.63	0.89	0.42	0.34	0.54
1995–2004	4,671	0.94	0.88	1.01	0.55	0.51	0.60
2005–2015	8,089	1.01	0.96	1.05	0.53	0.50	0.56
Sex designation threshold of 85% confidence							
1985–1994	1,071	0.78	0.65	0.95	0.42	0.33	0.55
1995–2004	4,026	0.97	0.90	1.04	0.53	0.49	0.58
2005–2015	6,937	1.05	1.01	1.10	0.52	0.49	0.56
Sex designation threshold of 95% confidence							
1985–1994	1,066	0.78	0.64	0.95	0.40	0.31	0.54
1995–2004	4,021	0.96	0.89	1.03	0.53	0.48	0.59
2005–2015	6,917	1.05	1.00	1.10	0.53	0.49	0.56
Alternative definition of relative risk							
1985–1994	1,066	0.95	0.78	1.16	0.41	0.32	0.55
1995–2004	4,024	1.08	1.00	1.16	0.53	0.48	0.58
2005–2015	6,928	1.15	1.10	1.20	0.52	0.48	0.55

*Insufficient number of data points to analyze by decades

Supplemental Table 3: Sensitivity analysis of RR estimates for different samples – life sciences

			95% CI			95% CI	
	Nr. of articles	RR first authorship	Lower bound	Upper bound	RR last authorship	Lower bound	Upper bound
Results from main text (for comparison)							
1985–1994	77,066	1.01	0.99	1.02	0.59	0.57	0.60
1995–2004	204,414	1.09	1.08	1.10	0.54	0.53	0.55
2005–2015	319,267	1.20	1.19	1.21	0.57	0.57	0.58
Excl. articles with alphabetical authorship							
1985–1994	70,921	1.01	0.99	1.02	0.58	0.57	0.59
1995–2004	192,787	1.09	1.08	1.10	0.54	0.53	0.54
2005–2015	306,434	1.21	1.20	1.22	0.57	0.57	0.58
Excl. articles with missing gender of authors							
1985–1994	9,112	1.06	1.03	1.10	0.57	0.54	0.61
1995–2004	47,443	1.08	1.07	1.10	0.56	0.55	0.57
2005–2015	89,586	1.14	1.13	1.15	0.59	0.58	0.60
Including articles with single and dual authorship(s)							
1985–1994	104,909	0.98	0.97	1.00	0.66	0.65	0.67
1995–2004	255,258	1.05	1.04	1.06	0.58	0.57	0.59
2005–2015	381,978	1.13	1.13	1.14	0.60	0.59	0.60
Sex designation threshold of 85% confidence							
1985–1994	77,336	1.01	0.99	1.02	0.59	0.58	0.60
1995–2004	204,841	1.09	1.08	1.10	0.54	0.54	0.55
2005–2015	320,020	1.20	1.19	1.20	0.58	0.57	0.58
Sex designation threshold of 95% confidence							
1985–1994	76,810	1.01	0.99	1.02	0.58	0.57	0.59
1995–2004	204,065	1.09	1.08	1.10	0.54	0.53	0.54
2005–2015	318,538	1.21	1.20	1.21	0.57	0.56	0.57
Alternative definition of relative risk							
1985–1994	77,066	1.18	1.16	1.20	0.59	0.57	0.60
1995–2004	204,414	1.25	1.24	1.26	0.54	0.53	0.55
2005–2015	319,267	1.32	1.31	1.33	0.57	0.57	0.58

Supplemental Figure 1: RR of prestigious authorship across the life sciences per year



Relative risk (RR) of prestigious authorship from 1985 to 2015. The green dashed horizontal line marks the point of equal gender probabilities of prestigious authorship. The red and blue lines represent the RR of first and last authorships, respectively. The grey dashed and dotted lines indicate 95% binomial ratio confidence intervals for the RR of first and last authorship, respectively.

References

1. Torvik VI and Smalheiser NR. Author name disambiguation in MEDLINE. *ACM Trans Knowl Discov Data*. 2009;3:1-29.

2. Lerchenmueller MJ and Sorenson O. Author disambiguation in PubMed: Evidence on the precision and recall of Author-ity among NIH-funded scientists. *PLoS ONE*. 2016;11:e0158731.

3. Kawashima H and Tomizawa H. Accuracy evaluation of Scopus Author ID based on the largest funding database in Japan. *Scientometrics*. 2015;103:1061-1071.

4. Lerchenmueller MJ. Gender designation via Genderize API with unlimited request *doi:106084/m9figshare4563814*. 2016.

5. R: A language and environment for statistical computing. R Foundation for Statistical Computing [computer program]. Vienna, Austria; 2015.

6. genderizeR: gender prediction based on first names. R package version 1.2.0. 2015. https://github.com/kalimu/genderizeR (accessed March 28th 2016).

7. CodingNews. Gender detection. 2015;http://codingnews.info/post/gender-detection.html (accessed 11/15/17).