Supplemental Methods:

Cleaning Author Names:

The task of linking together all publications by a given author presents two potential risks: 1) the potential to link two different authors with the same name together and 2) the potential to fail to link two different presentations of an author's name together due to differences in initials, capitalization, punctuation, etc. These problems are known as name homography and name variability, respectively. There is ongoing work using name disambiguation using machine learning to train models to disambiguate authors, and these tools rely on the inclusion of other variables such as email, organization, and co-authors. While beneficial, these models remain exploratory, and emails and organizations are not reliably available in PubMed nor are they consistent within an author across time. Therefore, we chose to rely specifically on processing and cleaning author names as they were presented in the author list without further attempts at disambiguation.

To address issues of name variability across journal styles, all names (first and last) were cleaned in the following manner. Names were processed to clean Unicode text and replace special characters such as periods, dashes, and quotation marks with spaces. After this processing, authors with blanks for first or last name were removed. Then, all text was converted to lowercase strings in order to control for any differences in capitalization (e.g. Mcnally vs McNally). Finally, after trimming leading and lagging white space from first names and last names, each author's name was defined as the concatenated string of their first and last name separated by a space.

Gini Coefficient Calculation:

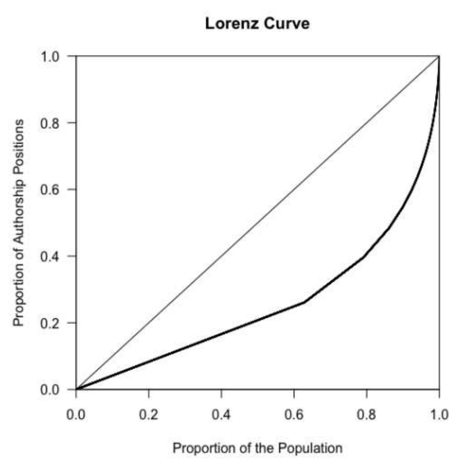
Author information was extracted from PubMed for each author listed on an article in the format outlined below. Each line of the database represents an author position, and includes metadata for the article (article number, journal, year published) as well as the author specific information (first name (forename), last name, and order within the author block). To match authors across articles, a concatenated string of the author's first and last name was created (indicated by the column "composite"). The concatenated string was cleaned using the procedures outlined above.

article_number	ForeName	LastName	order	composite	journal	year
18163451	Julie M	Barkmeier-Kraemer	2	Julie M Barkmeier-Kraemer	Mov. Disord.	2008
18163451	Michelle R	Ciucci	1	Michelle R Ciucci	Mov. Disord.	2008
18163451	Scott J	Sherman	3	Scott J Sherman	Mov. Disord.	2008
18163453	Aileen K	Ho	1	Aileen K Ho	Mov. Disord.	2008
18163453	John L	Bradshaw	2	John L Bradshaw	Mov. Disord.	2008
18163453	Robert	Iansek	3	Robert Iansek	Mov. Disord.	2008

Therefore, the number of rows of the data frame indicates the total number of author positions available. Then the composite name can be used as a factor level variable to calculate a frequency table, which contains the number of author positions occupied by each author.

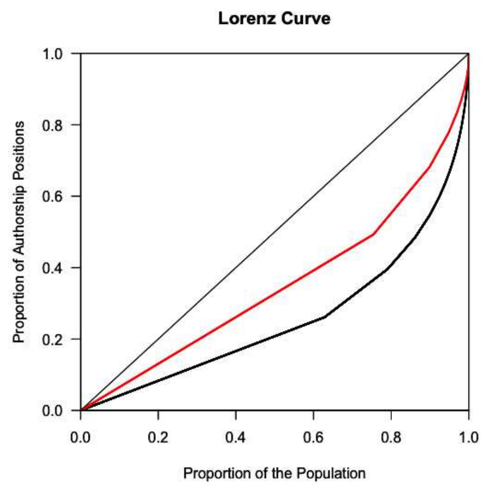
The Gini() function within the DescTools package in R (0.99.30) uses this frequency table to plot a Lorenz Curve and determine the Gini Coefficient (GC). The Lorenz Curve first sorts the population in increasing order based on their share of resources, in this case, number of author positions held. The x-axis presents the cumulative proportion of the population (authors) while the y axis shows the cumulative proportion of resources (author positions) held by that section of the population (moving from left to right).

This Lorenz Curve summarizes the distribution of authorship across the entire author block. If all authors held the same number of author positions, the curve would follow the diagonal line in the center of the image. However, because authors hold different numbers of author positions within our database, the Lorenz Curve deviates from the diagonal line, indicating inequality in resources held. In the example above, we can see that the bottom 60% of authors only hold about 25% of author positions. Alternatively, the



bottom 80% of authors only hold approximately 40% of the author positions. This indicates inequality in the number of author positions held across different authors.

The GC summarizes the Lorenz Curve by calculating the area between the diagonal line of equality and the curve. This space is the inequity gap. The further the curve is from the diagonal line, the greater the inequality, and therefore the greater the GC.



The GC ranges from 0 to 1, with 0 representing complete equality of resources (in this case the Lorenz Curve would follow the diagonal line) and 1 representing complete inequality in resources (or one individual holding all the resources). The figure above shows the Lorenz Curve for the overall authorship in black (GC = 0.49) and the Lorenz Curve for first authorship in red (GC = 0.30). As you can see, the curve for first authorship is closer to the diagonal line than the curve for overall authorship, which is also reflected in the lower GC for first authorship. This indicates that there is more equity in authorship distributions in the first author position than the last author position. It is important to note that the GC represents an exact quantification of the Lorenz Curve, and does not include confidence intervals.

Sensitivity Analysis on the Inclusion of Initials in the ORCID Data Frame:

Since the initials of one author may not be used consistently across publications, we also cleaned the ORCID database to remove initials, and recalculated the number of ORCID records with matching names. Initials were defined as follows:

After replacing special characters such as periods, dashes, and quotation marks with spaces, strings were separated at spaces to make an array of individual name parts (e.g. “Jane L” becomes (“Jane”, “L”)). Individual name parts of only one letter long (e.g. “L”) were removed from the name string. Recognizing that authors around the world may use individual characters for their first name, this rule was only applied to name parts including Latin alphabet characters (including characters with accents). The pre-processing above to replace periods and dashes with spaces will identify initials in the formats of “J.L.” or “J-L” but not “JL”. To capture these initials, first names that were entirely capitalized and only contained three or fewer Latin alphabet characters, were categorized as initials and removed. This cleaning was applied in addition to the cleaning methods outlined for the main analysis above.

Since the removal of initials in the ORCID cohort only led to a very small increase the number of authors with name homography, we chose not to clean initials in the main analysis.

Supplemental Material:

Figure S1:

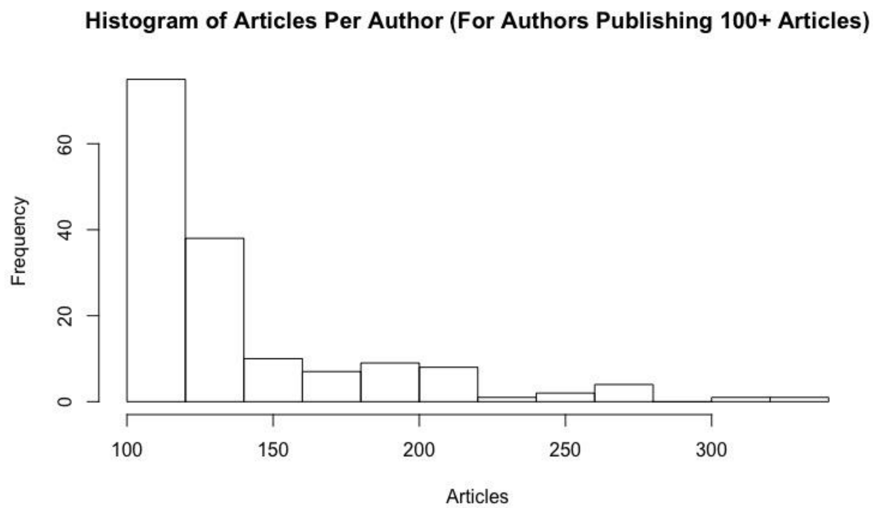


Figure S2: Percent Change in Author Positions and Number of Articles Published from 2008 to 2019

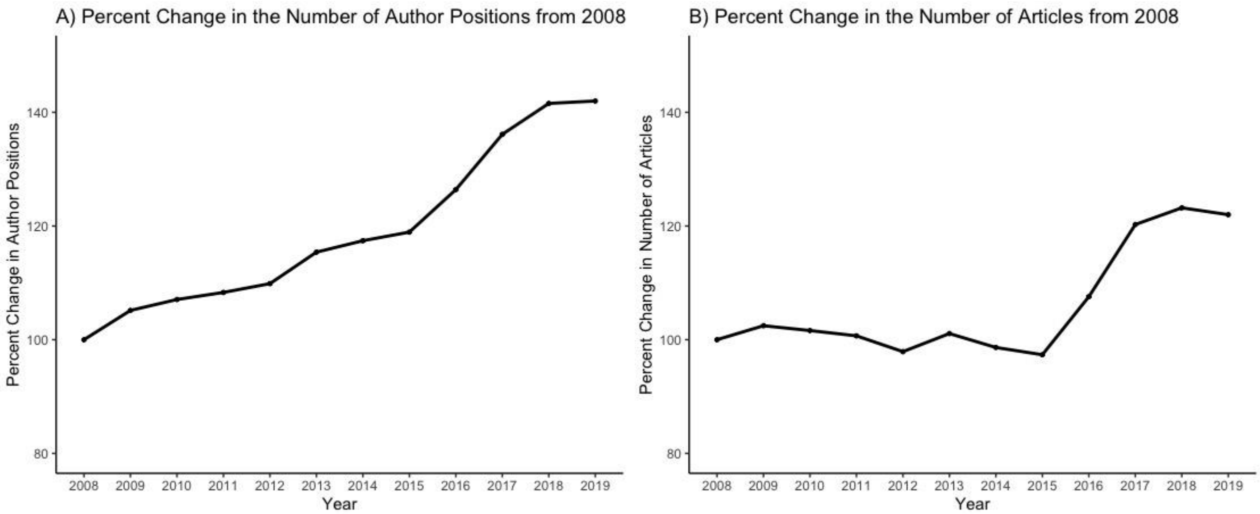


Figure S3: Change in Gini coefficient by authorship position over time limited to articles with 20 or fewer authors

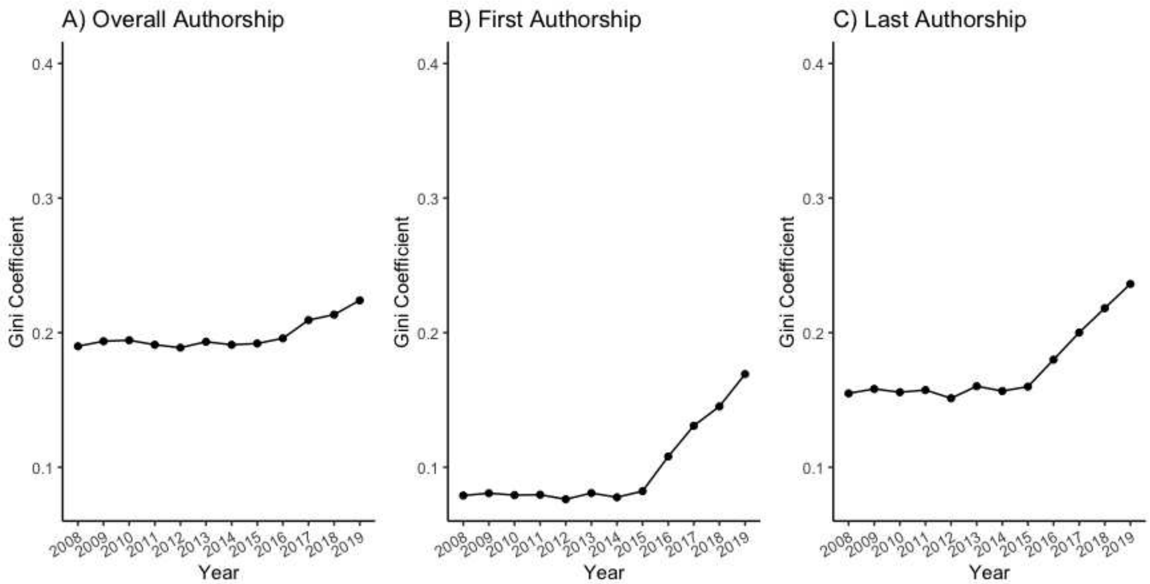


Figure S4: Change in Gini coefficient by authorship position over time limited to articles without a consortium as last author

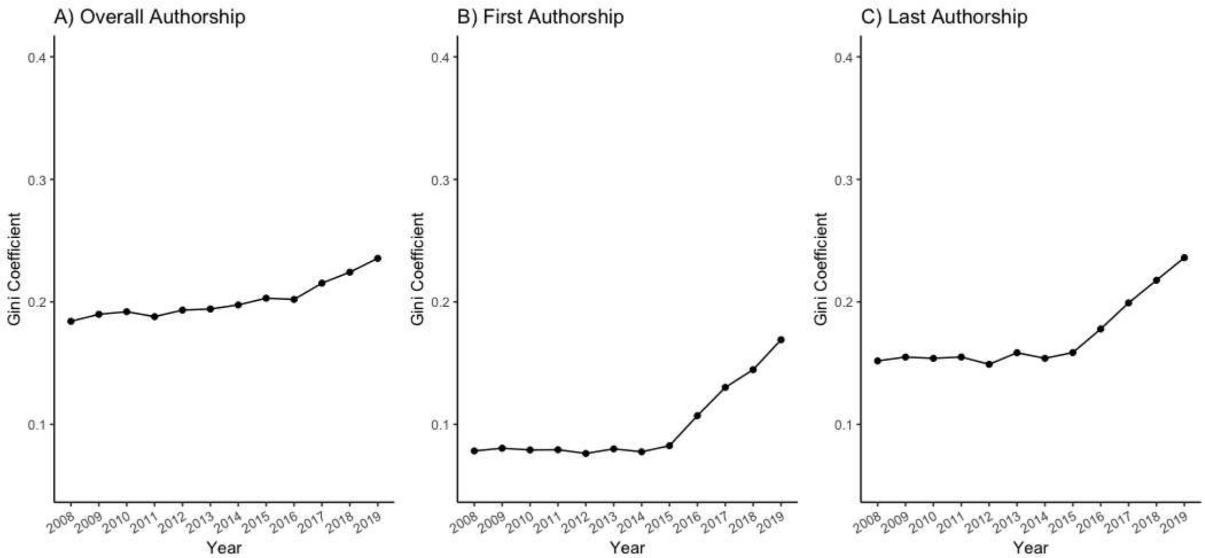


Figure S5: Change in Gini coefficient by authorship position over time in 10 random samplings of 20,000 articles per year

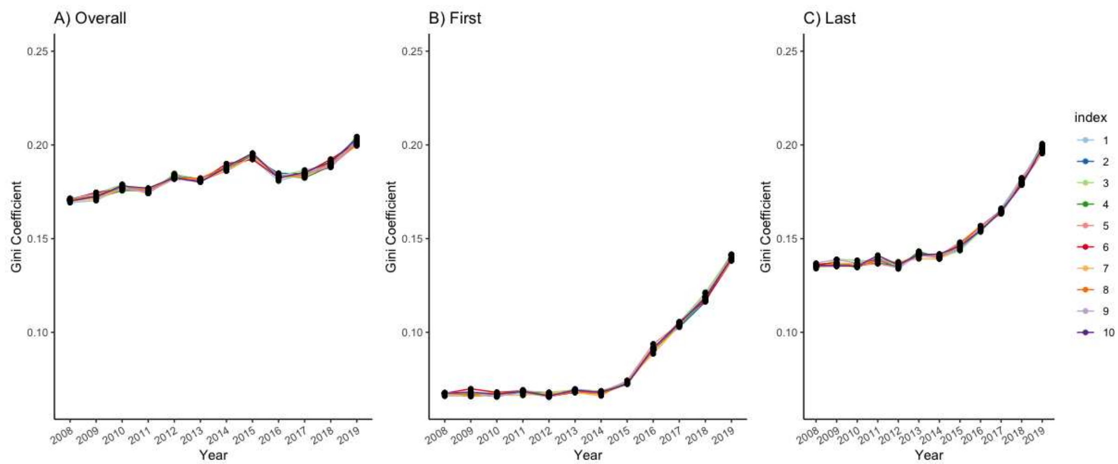


Figure S6: Change in Gini coefficient by authorship position over time in the entire collection excluding articles published in the cross-specialty journals

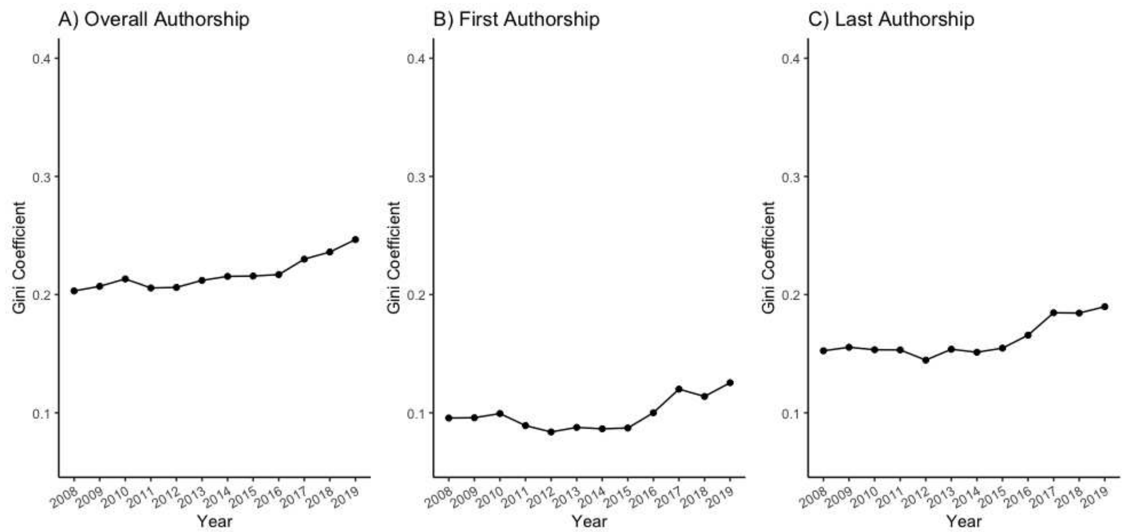


Figure S7: Gini coefficient by authorship position separated by specialty limiting to 1000 articles per year per specialty

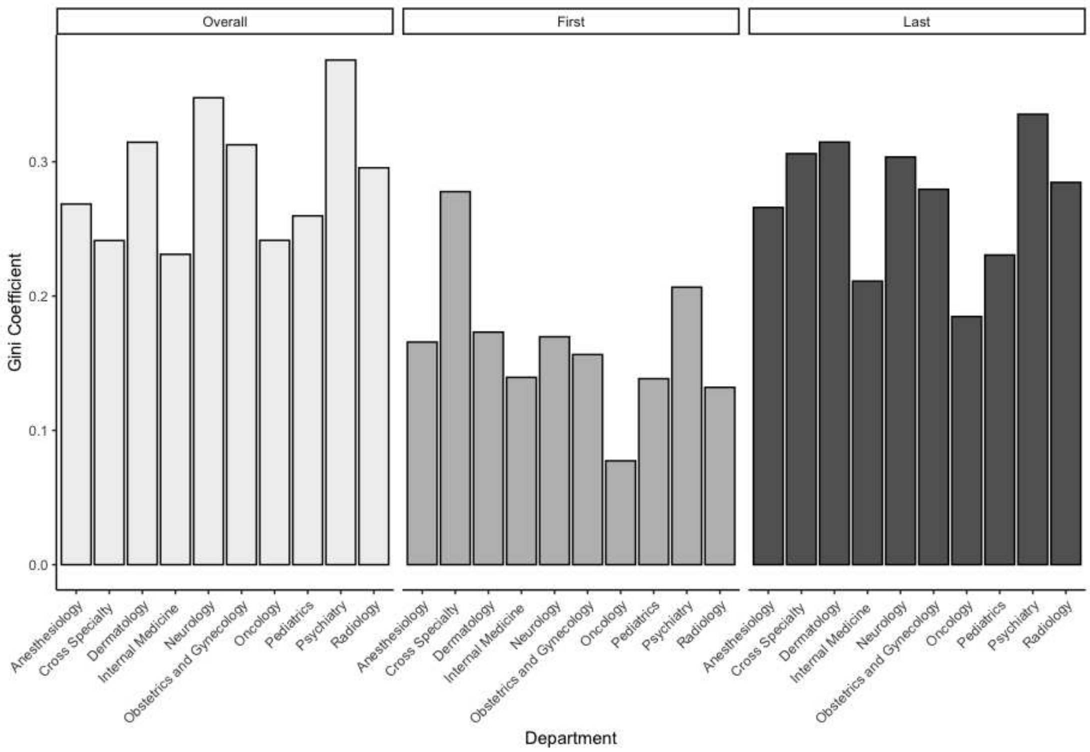


Figure S8: Change in Gini coefficient by authorship position over time limiting to 1000 articles per year per specialty

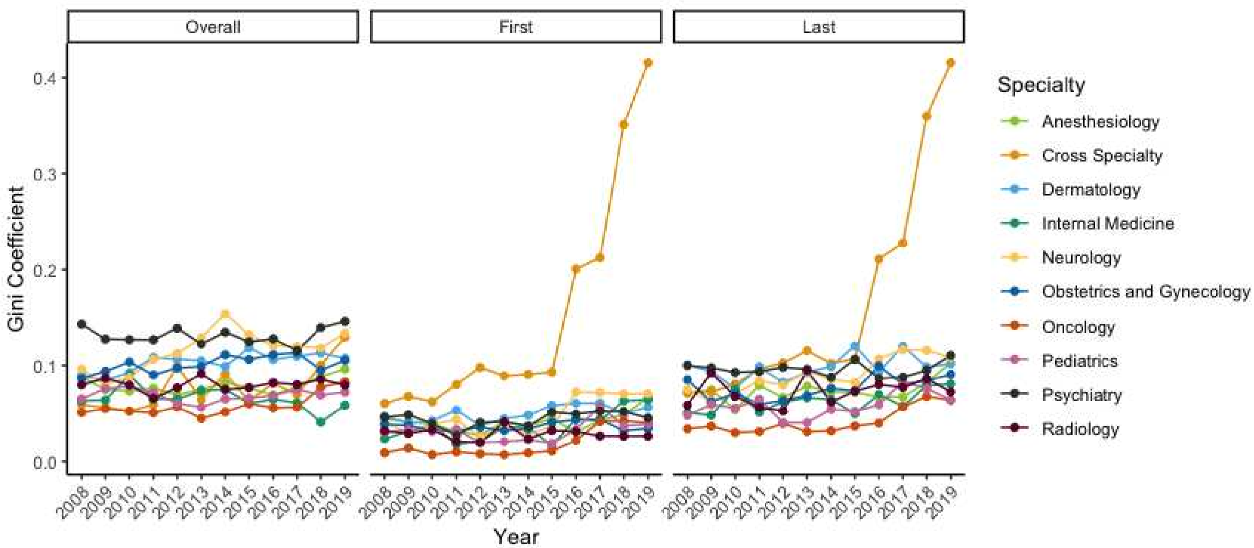


Table S1: Journals and Impact Factors by Specialty

Department	Category	Impact Factor	Department	Category	Impact Factor
Anesthesiology	Br J Anaesth	6.24	Internal Medicine	Ann. Intern. Med.	17.2
	Anesthesiology	5.79		JAMA Intern Med	16.54
	Pain	5.45		PLoS Med.	11.86
	Anaesthesia	4.74		J Cachexia Sarcopenia Muscle	9.7
	J Neurosurg Anesthesiol	4.03		BMC Med	8.1
	Anesth. Analg.	4.01		J. Intern. Med.	7.98
	Eur J Anaesthesiol	3.57		Mayo Clin. Proc.	6.86
	Reg Anesth Pain Med	3.52		CMAJ	6.78
	Clin J Pain	3.49		Nat Rev Dis Primers	6.39
	Eur J Pain	3.02		Cochrane Database Syst Rev	6.26
	Pain Physician	2.84		Am. J. Med.	5.55
	Minerva Anestesiol	2.62		Ann Fam Med	4.92
	Pain Pract	2.5		Transl Res	4.65
	Acta Anaesthesiol Scand	2.44		Dtsch Arztebl Int	4.26
	Curr Opin Anaesthesiol	2.37		Palliat Med	4.22
Dermatology	J. Am. Acad. Dermatol.	7	Neurology	Lancet Neurol	26.28
	J. Invest. Dermatol.	6.29		Nat Rev Neurol	20.26
	JAMA Dermatol	5.82		Acta Neuropathol.	12.21
	Pigment Cell Melanoma Res	5.17		Brain	10.29
	Br. J. Dermatol.	4.71		JAMA Neurol	10.03

	Contact Derm.	4.34		Ann. Neurol.	9.89
	J. Dermatol. Sci.	3.73		Alzheimers Dement	9.48
	Acta Dermatovenerol Alp Pannonica Adriat	3.65		Sleep Med Rev	8.96
	J Eur Acad Dermatol Venereol	3.53		Neurology	8.32
	Wound Repair Regen	3.04		Neuro-oncology	7.79
	J Dtsch Dermatol Ges	2.87		Neuroscientist	7.39
	Int Wound J	2.85		J. Neurol. Neurosurg. Psychiatry	7.349
	Skin Pharmacol Physiol	2.76		Mov. Disord.	7.07
	Am J Clin Dermatol	2.76		Brain Pathol.	6.62
	Photodermatol Photoimmunol Photomed	2.66		Alzheimers Res Ther	6.15
Obstetrics and Gynecology	Hum. Reprod. Update	11.75	Oncology	CA Cancer J Clin	187.04
	Am. J. Obstet. Gynecol.	5.23		Nat. Rev. Cancer	37.15
	Obstet Gynecol	5.22		Lancet Oncol.	33.9
	BJOG	5.05		Cancer Cell	27.41
	Hum. Reprod.	5.02		J. Clin. Oncol.	24.01
	Gynecol. Oncol.	4.96		Nat Rev Clin Oncol	20.69
	Ultrasound Obstet Gynecol	4.71		Cancer Discov	20.01
	Fertil. Steril.	4.45		JAMA Oncol	16.56
	Pregnancy Hypertens	3.93		J. Natl. Cancer Inst.	12.59
	Mol. Hum. Reprod.	3.59		Ann. Oncol.	11.86
	Maturitas	3.26		Leukemia	11.7

	Reprod. Biomed. Online	3.25		Clin. Cancer Res.	9.62
	Clin Perinatol	3.233		Biochim. Biophys. Acta	9.45
	Semin. Perinatol.	3.185		Semin. Cancer Biol.	9.14
	J Gynecol Oncol	3.14		Cancer Res.	9.12
Pediatrics	JAMA Pediatr	10.25	Psychiatry	World Psychiatry	26.56
	J Am Acad Child Adolesc Psychiatry	6.44		JAMA Psychiatry	15.31
	Pediatrics	5.71		Am J Psychiatry	14.18
	Pediatr Diabetes	4.27		Mol. Psychiatry	13.2
	Arch. Dis. Child. Fetal Neonatal Ed.	4.1		Lancet Psychiatry	11.59
	J Adolesc Health	3.97		Biol. Psychiatry	11.41
	J. Pediatr.	3.87		Psychother Psychosom	8.96
	Pediatr Allergy Immunol	3.78		Schizophr Bull	7.58
	Pediatr Crit Care Med	3.5		J. Neurol. Neurosurg. Psychiatry	7.35
	Pediatr Obes	3.4		Acta Psychiatr Scand	6.79
	Semin Fetal Neonatal Med	3.33		Neuropsychopharmacology	6.4
	Eur Child Adolesc Psychiatry	3.3		J Am Acad Child Adolesc Psychiatry	6.44
	Arch. Dis. Child.	3.27		Br J Psychiatry	6.35
	Clin Perinatol	3.23		J Child Psychol Psychiatry	6.23
	Semin. Perinatol.	3.19		Addiction	5.79
Radiology, Nuclear Medicine, and Imaging	JACC Cardiovasc Imaging	10.19	Cross-Specialty	N. Engl. J. Med.	72.41

	Radiology	7.3	Lancet	47.83
	Eur. J. Nucl. Med. Mol. Imaging	7.28	JAMA	44.41
	Circ Cardiovasc Imaging	6.8	BMJ	20.79
	J. Nucl. Med.	6.65		
	Neuroimage	5.84		
	J Cardiovasc Magn Reson	5.6		
	Semin Radiat Oncol	5.36		
	Invest Radiol	5.2		
	Int. J. Radiat. Oncol. Biol. Phys.	5.13		
	Ultrasound Obstet Gynecol	4.71		
	Clin Nucl Med	4.56		
	Hum Brain Mapp	4.53		
	Radiother Oncol	4.33		
	Med Image Anal	4.19		

Adapted from Hart KL, Perlis RH. Trends in Proportion of Women as Authors of Medical Journal Articles, 2008-2018. JAMA Intern Med. 2019 Sep 1;179(9):1285.

Table S2: Gini Coefficient by authorship position separated by specialty

Specialty	Overall Authorship	First Authorship	Last Authorship
Internal Medicine	0.30	0.20	0.28
Pediatrics	0.37	0.23	0.34
Anesthesiology	0.36	0.25	0.36
Oncology	0.42	0.20	0.37
Psychiatry	0.43	0.25	0.39
Obstetrics and Gynecology	0.44	0.27	0.40
Cross Specialty	0.34	0.38	0.41
Dermatology	0.41	0.25	0.41
Neurology	0.46	0.26	0.42
Radiology	0.47	0.28	0.46

Table S3: Gini Coefficients for each sensitivity analysis by author position

	Entire Collection*	Articles with 20 or fewer authors	Articles without a consortium as last author	Excluding articles published in the 4 cross-specialty journals
GC for entire author block	0.49	0.47	0.47	0.48
GC for first author position	0.30	0.29	0.29	0.28
GC for last author position	0.44	0.44	0.43	0.44

*The GCs for the primary analysis of the entire cohort are included for reference

Table S4: Spearman rank order correlations between year and Gini coefficient for each random sampling of 20,000 articles

	Overall		**First Authorship**		**Last Authorship**	
	rho	**P**	**rho**	**P**	**rho**	**P**
Primary Analysis*	0.99	<0.001	0.75	0.007	0.85	<0.001
Model 1	0.88	< 0.001	0.92	< 0.001	0.90	< 0.001
Model 2	0.90	< 0.001	0.92	< 0.001	0.92	< 0.001
Model 3	0.88	< 0.001	0.98	< 0.001	0.92	< 0.001
Model 4	0.87	< 0.001	0.86	0.001	0.90	< 0.001
Model 5	0.88	< 0.001	0.89	< 0.001	0.92	< 0.001
Model 6	0.94	< 0.001	0.81	0.002	0.94	< 0.001
Model 7	0.91	< 0.001	0.84	0.001	0.98	< 0.001
Model 8	0.91	< 0.001	0.83	0.002	0.98	< 0.001
Model 9	0.91	< 0.001	0.80	0.003	0.92	< 0.001
Model 10	0.88	< 0.001	0.85	0.001	0.97	< 0.001

*The correlation coefficients for the primary analysis of the entire cohort are included for reference. Models 1-10 refer to models derived from randomly samplings of 20,000 articles for each year within the study period and repeating the correlation between year and GC for each author position.

Table S5: Gini coefficients for each gender separated by author position

	Authors identified as men	**Authors identified as women**
GC for entire author block	0.53	0.45
GC for first author position	0.33	0.28
GC for last author position	0.47	0.43

Table S6: Spearman rank order correlations by author position by gender

	Authors identified as men		Authors identified as women	
Author Position	**rho**	**P**	**rho**	**P**
Entire author block	0.97	<0.001	0.99	<0.001
First author position	0.85	<0.001	0.85	<0.001
Last author position	0.83	0.002	0.96	<0.001

Supplemental Results:

- *Controlling for Author Block Length, Consortia, and Number of Articles Published per Year*

Table S2 presents the GC for each authorship position (overall, first author, and last author) for each sensitivity analysis.

Our data indicated that there was a significant increase in the number of author positions from 2008 to 2019 (160795 author positions in 2008 and 228324 author positions in 2019, Figure S2A; r(10)= 0.97, $p < 0.001$). Therefore, we conducted a sensitivity analysis to control for the extreme right tail of author block lengths by repeating the primary analyses solely among articles with 20 or fewer authors. This threshold was chosen to exclude articles with the largest author blocks, while also keeping at least 95% of articles for analysis (n excluded = 10188 articles). The results of this analysis were qualitatively similar to the primary results (Figure S3). Looking within articles with 20 or fewer authors (n articles = 302,034), there was a significant positive correlation between year and GC for the overall authorship position (rho = 0.68, $p = 0.02$), for the first author position (rho =0.78, $p = 0.005$), and for the last author position (rho = 0.83, $p = 0.001$).

Alternatively, inclusion of consortiums as authors may also influence the number of authorship positions, as some authors may be included within the consortium but not in the main author block. Results were generally unchanged from those in the primary analysis. In groups without a consortium listed as the last author (n articles = 299,235), there was a significant positive correlation between year and GC for the overall authorship position (rho = 0.97, $p < 0.001$), the first author position (rho =0.75, $p = 0.007$), and the last author position (rho = 0.86, $p < 0.001$; Figure S4).

Additionally, there was a significant increase in the number of publications per year from 2008 to 2019 (24800 articles in 2008 and 30256 articles in 2019; r(10) = 0.75, $p = 0.005$), with the most substantial increase in the number of articles occurring between 2015 and 2017 (Figure S2B). Therefore, to account potential influences of the increase in the number of articles published during the study period, we repeated the primary analysis 10 times, each time randomly sampling 20,000 articles from each year without replacement and calculated the GC for each author position in each year. The results are presented in Figure S5 and Table S3. These results did not differ from the primary results.

- *Controlling for Variance Across Specialties*

Because there was an increase in GC within the high-impact cross-specialty journals over the study period, two additional sensitivity analyses were completed. The first sensitivity analysis repeated the primary analysis after excluding articles published in the 4 cross-specialty journals. There was a significant positive correlation between year and GC for overall authorship (rho = 0.92, $p < 0.001$) and the last author position (rho =0.69, $p = 0.02$), but not for the first author position (rho = 0.48 $p = 0.12$; Figure S6).

The second sensitivity analysis sought to control for the smaller number of articles published in the cross-specialty journals as compared to the other specialties, sampling 1000 articles per specialty per year. In this reduced collection, the cross-specialty journals continued to have higher GCs for the first and last author position over time, as compared to the other specialties (Figure S7 and Figure S8).