BMJ Open

Gender differences in publications related to COVID-19

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-045176
Article Type:	Original research
Date Submitted by the Author:	27-Sep-2020
Complete List of Authors:	Lerchenmüller, Carolin; University Hospital Heidelberg, Schmallenbach, Leo; University of Mannheim Jena, Anupam; Harvard University, Health Care Policy Lerchenmueller, Marc ; University of Mannheim,
Keywords:	GENERAL MEDICINE (see Internal Medicine), COVID-19, Health policy < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

Gender differences in publications related to COVID-19

Carolin Lerchenmüller, MD,[1,2,3] Leo Schmallenbach, MSc,[4]
Anupam B. Jena, MD,[5,6,7*] PhD, Marc J. Lerchenmueller, MPH, PhD[4,8*]

[1]Department of Cardiology, Angiology, Pulmonology, University Hospital Heidelberg, Germany
[2]German Center for Heart and Cardiovascular Research (DZHK), Partner Site Heidelberg/Mannheim, Germany
[3]Cardiology Division and Corrigan Minehan Heart Center, Massachusetts General Hospital, Boston, MA
[4]Area Management, University of Mannheim, Mannheim, Germany
[5]Department of Health Care Policy, Harvard Medical School Boston, MA
[6]Department of Medicine, Massachusetts General Hospital, Boston, MA
[7]National Bureau of Economic Research, Cambridge, MA
[8]School of Management, Yale University, New Haven, CT

*These authors contributed equally

Word count: 1795
Keywords: COVID-19, Health policy, Gender equity

Corresponding author:

Carolin Lerchenmueller, MD
Department of Cardiology, Angiology, Pulmonology
University Hospital Heidelberg
INF 410
69120 Heidelberg, Germany
Email: carolin.lerchenmueller@med.uni-heidelberg.de

Gender differences in publications related to COVID-19

Abstract

Objective: When concerns have been raised that the COVID-19 pandemic has disproportionally involved male scientific authors, potentially to the disadvantage of women's careers and societies' response to the pandemic, we sought to investigate the gender distribution of first authorships for publications on COVID-19.

Methods and Results: We compared gender distribution of first authorships for publications on COVID-19 from 2020 versus publications appearing in the same journals during the previous year and find that the gender gap widens by 18 percentage points with the COVID-19 pandemic. Globally, female researchers' productivity goes down in relation to their male peers across all continents.

Conclusion: The reduction in women's research output regarding COVID-19 appears particularly concerning as many disciplines informing the response to the pandemic had near equal gender shares of first authorship in 2019. Academic and funding institutions may need to consider potential remedies to mitigate the pandemic's negative effect on women's scientific productivity.

Strengths and limitations of this study

- Using a natural experiment design, this study provides evidence for a ~3.5x increase in the gender gap among leading authors on 27,821 COVID-19 publications relative to a set of 365,914 control articles appearing in the same journals the previous year
- Particularly fields that are pertinent to addressing the pandemic, like virology, infectious diseases, public health, and internal medicine, show a pronounced slump in women's productivity as the pandemic unfolds
- With fewer women contributing to COVID-19 research than one would expect, concerns mount that the academic community can offer the best ideas for addressing the pandemic
- To our knowledge, this is the most up to date and the largest study investigating gender differences in first authorships of COVID publications, including data from articles published between February 1st, and September 10th
- Our large-scale study design does not allow, however, to discern the multifactorial mechanisms leading to the decline in women's productivity, which may range from displaced work hours needed for child care to gender differences in the allocation of COVID-19 funding schemes

Introduction

It is now widely accepted that women are integral to productive and innovative science communities (*1, 2*). Nonetheless, women remain underrepresented in prestigious author positions on publications in the life sciences and medicine (*3*), are less likely to be promoted to higher academic ranks, and are paid less, despite the continuously growing number of female academics in those fields (*4*). Projections indicate that this gap will persist if targeted interventions are not implemented (*5*).

It is in this setting that concerns have been raised that research and expert reporting on the COVID-19 (Coronavirus Disease 2019, COVID) pandemic has disproportionately involved male scientific authors, at least in part because closures of academic institutions, schools, and childcare facilities likely have led to greater household responsibilities borne by female scientists (*6, 7*). Consistent with this, previous research has shown that parental leaves taken by men often result in increased productivity, while no such phenomenon can be observed for women, suggesting that working women still contribute significantly to childrearing and household tasks (*8*). These dynamics likely affect early-career female scholars disproportionally.

The proliferation of COVID related publications provide a unique window into these gendered dynamics for two reasons. First, COVID publications have been produced rapidly under unusual conditions that likely disfavor female scientists relative to usual conditions that can serve as a control, offering a natural experiment setting. Second, COVID publications are mainly produced in the life sciences and medicine where long-standing authorship standards reserve the first author position to early career investigators, which allows estimating the effect for this group in particular. In this study, we assessed the pandemic's effect on women's COVID-related scientific publishing worldwide.

Methods

We compared the gender composition of 27,821 PubMed-indexed articles published between February 1st and September 10th, 2020 that included the term "COVID" in the title or the abstract to the gender composition of 365,914 articles published in the same journals during the same period in 2019 (see **Supplementary Materials** for data and methods). We allocated 2,618 represented journals to scientific disciplines based on Clarivate journal categories and determined author gender with the Genderize database (*9*). We assessed changes in the gender composition of authors between periods (COVID and pre-COVID) within scientific discipline (e.g., medicine/internal medicine, infectious diseases, virology, etc.), to account for the possibility that the scientific areas in which COVID research has predominantly been published may at baseline (i.e. pre-COVID) have had disproportionately more male authors. We also used detailed affiliation data to determine the geographic locale of the first author.

Patient and Public Involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for design or implementation of the study. No patients were asked to advise on the interpretation or write up of the results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

Results

Our data shows that disciplines producing most COVID-relevant publications had near equal gender shares of first authorship pre-COVID. For example, in the fields of virology, general/internal medicine, and infectious diseases, the proportions of first authors who were female prior to COVID were 52%, 45%, and 49%, respectively. The share of female first authors in relevant fields like public environmental and occupational health was even higher (58%). However, significant reductions in female first authorships have occurred in almost all disciplines, including the above, since the pandemic began. Across all disciplines that published research related to COVID, we found that the average gender gap in first authorships was 7% (54% male vs 46% female first authors) in 2019, rising to 25% (63% male vs 37% female first authors) for COVID related research. The average gender gap in first authorships has thus risen by 18 percentage points in the wake of the pandemic (**Figure 1**). In the fields of virology, general/internal medicine, and infectious diseases in particular, the gender rift in first authorships has widened above average (29, 18, and 20 percentage points, respectively), while there was no gap in the control group (**Figure 1**). The previously female first author-dominated field of public environmental and occupational health now presents a gender gap of 18% (total increase from pre-COVID of 34 percentage points) (**Figure 1**).

Since the pandemic has affected countries differently, we further performed analyses based on the country affiliation of the first author. Female researchers' productivity goes down in relation to their male peers across all continents. Publishing activity by women located in Brazil, for example, is reduced by 36%, the average reduction of first authorships in Europe, where the pandemic first gathered speed, is 31% (**Figure 2**). In Canada and the United States, the reduction to date is 10% and 22%, respectively (**Figure 2**).

Discussion

Our results provide the largest and latest systematic evidence for the COVID pandemic's effect on women's publishing productivity across disciplines worldwide. In light of previous research and observations (*6-8, 10*), we suspect that the overcontribution of women to household and child rearing responsibilities that leaves less opportunity to participate in writing, submitting and publishing research related to COVID-19 leads to the reduced productivity of women as borne out by our data. Despite at an early stage and still during an ongoing pandemic, these results already suggest cause for concern, both on the individual and societal level. Our findings indicate that the current scientific/medical response to one of the most incisive global crises is overly male dominated, particularly when considering that the fields being called upon to offer scientific insights have historically had more gender balance in scientific representation. The present effects may be amplified as the pandemic wanes on and perceptions of women at earlier career stages may lead them to be less often allocated to leading roles on projects in the current circumstances.

Women are a vital part of the research and medical enterprise, impacting patient care, science and society. To avoid long-term impacts on the academic advancement and scientific contributions, the disproportionate impact of COVID-19 on early career women investigators needs serious consideration. As a first step, the problem needs to be openly and consciously discussed. Naturally, pre-existing inequities must be evaluated, and a long-term strategy has to be established to support equity and inclusion in science (*11*). But more acutely, COVID-related gender inequities need to be addressed with direct measures, monetary and non-monetary, on both the political (e.g. federal funding agencies) and institutional level. For example, modifications for grant deadlines, timelines, extensions for granted expenses, as well as additional (bridge) funding programs, are likely warranted (*12*). Extension of tenure evaluation and promotion should be considered on the individual level accounting for constraints posed by COVID. Resources for childcare should be provided for parents, additional funds for expanded childcare arrangements could help to reallocate time to regular professional duties. Financial support for postdoctoral/graduate students could help to facilitate a research set-back in a recently established laboratory.

Without policy interventions, our communities may miss out on some of the best ideas for tackling the pandemic, across scientific/medical disciplines and potentially beyond in other professions.

References

1. World Economic Forum, The Global Gender Gap Report 2017. http://www3.weforum.org/docs/WEF_GGGR_2017.pdf *(accessed 08 June 2018)*, (2017).
2. K. R. O'Brien, M. Scheffer, E. H. van Nes, R. van der Lee, How to Break the Cycle of Low Workforce Diversity: A Model for Change. *PloS one* **10**, (2015).
3. C. Lerchenmuller, M. J. Lerchenmueller, O. Sorenson, Long-Term Analysis of Sex Differences in Prestigious Authorships in Cardiovascular Research Supported by the National Institutes of Health. *Circulation* **137**, 880-882 (2018).
4. D. M. Blumenthal *et al.*, Sex Differences in Faculty Rank Among Academic Cardiologists in the United States. *Circulation* **135**, 506-517 (2017).
5. L. Holman, D. Stuart-Fox, C. E. Hauser, The gender gap in science: How long until women are equally represented? *PLoS Biol* **16**, e2004956 (2018).
6. C. Kitchener, in *The Lily / The Washington Post*. (thelily.com, 2020).
7. F. Staniscuaski *et al.*, Impact of COVID-19 on academic mothers. *Science* **368**, 724 (2020).
8. H. Antecol, K. Bedard, J. Stearns, Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies? *American Economic Review* **108**, 2420-2441 (2018).
9. V. Lariviere, C. Ni, Y. Gingras, B. Cronin, C. R. Sugimoto, Bibliometrics: global gender disparities in science. *Nature* **504**, 211-213 (2013).
10. Boston Consulting Group. Easing the covid-19 burden on working parents. https://www.bcg.com/en-us/publications/2020/helping-working-parents-ease-the-burden-of-covid-19.aspx *(accessed 14 September 2020)*, (2020).
11. J. L. Malisch *et al.*, Opinion: In the wake of COVID-19, academia needs new solutions to ensure gender equity. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 15378-15381 (2020).
12. E. M. Gibson *et al.*, How Support of Early Career Researchers Can Reset Science in the Post-COVID19 World. *Cell* **181**, 1445-1449 (2020).

Availability of data and materials

The datasets generated and analysed during the current study are available from the corresponding author.

Conflicts of interest

Dr. Jena reports receiving consulting fees unrelated to this work from Pfizer, Hill Rom Services, Bristol Myers Squibb, Novartis, Amgen, Eli Lilly, Vertex Pharmaceuticals, AstraZeneca, Celgene, Tesaro, Sanofi Aventis, Biogen, Precision Health Economics, and Analysis Group.

Funding

Support was provided by the Office of the Director, National Institutes of Health (1DP5OD017897, Dr. Jena). The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Author's contributions

LS and MJL collected data, LS, MJL and CL performed data analysis. CL, MJL, ABJ wrote the manuscript, LS edited the manuscript. All authors read and approved the final manuscript. Patients and public did not contribute to the study.

Acknowledgements

The authors acknowledge support by the German state of Baden-Württemberg through bwHPC—high performance cluster computing resources.

Supplementary Materials

Detailed methods and supporting data are available in the supplementary materials. All data and source code will be made publicly available with publication of this article.

Figure 1. Reduction in female first authorships during the COVID-19 pandemic by discipline

Heatmap depicting the percentage share of female vs. male first authorships for COVID-publications and control publications during the same period in 2019 in COVID-intensive disciplines (≥ 55 articles, descending frequency from top to bottom). Overall difference in female (compared to male) first authorships in %.

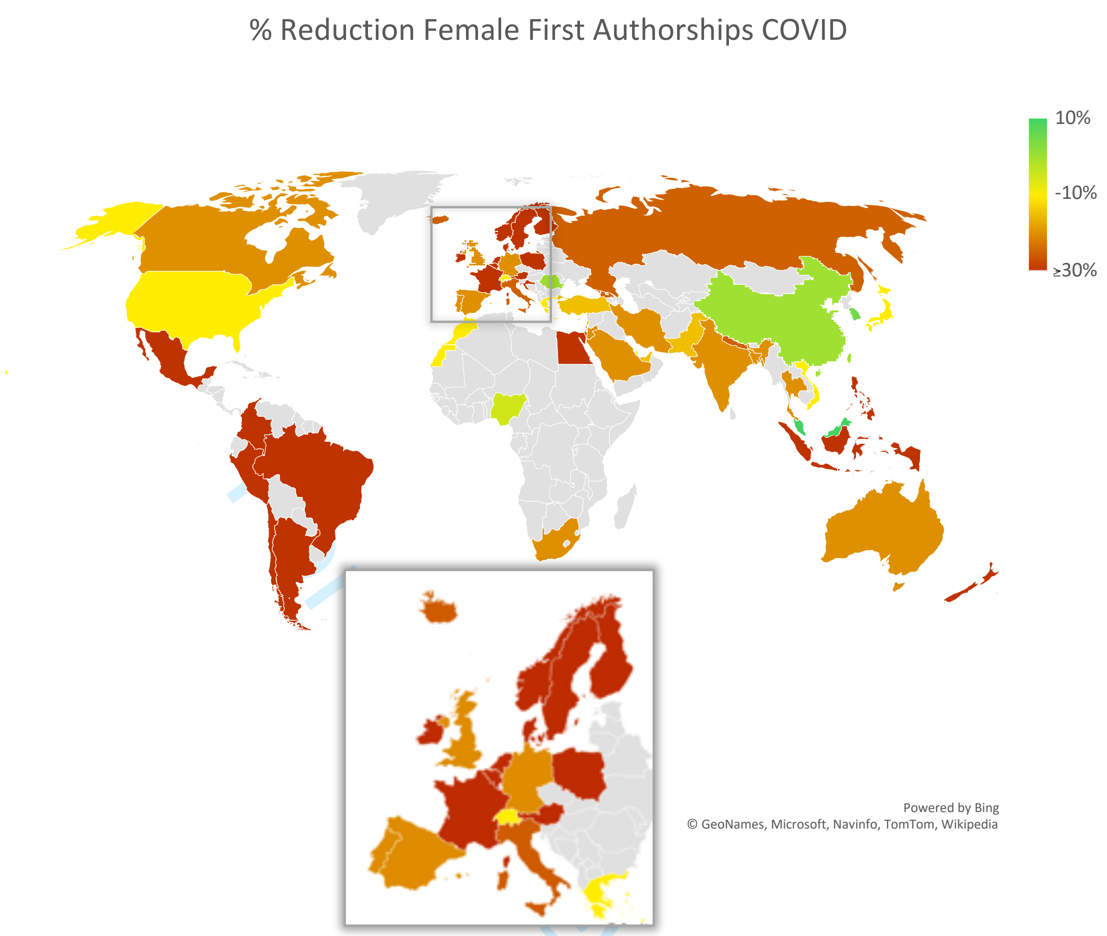

Figure 2. Reduction in female first authorships during the COVID-19 pandemic by country
Worldmap depicting the change in the gender gap in first authorship (percent female first authorship minus percent male first authorship) for COVID publications versus non-COVID publications. Included countries had at least 30 COVID publications up to September 10th, 2020.

Supplementary materials for

Gender differences in publications related to COVID-19

Carolin Lerchenmüller, MD, Leo Schmallenbach, MSc,
Anupam B. Jena, MD, PhD, Marc J. Lerchenmüller, MPH, PhD

Table of Contents

Additional Information on Data

We merged several databases to analyze potential gender differences in first authorships of COVID publications relative to a set of control publications in the same journals in the prior year. First, we extracted all articles from the PubMed database for which the term "COVID" appeared in the title or abstract and obtained all available article characteristics including, among others, the names of all authors, country affiliation per author, the journal ISSN (International Standard Serial Number), and time of publication (months and year). The U.S. National Library of Medicine maintains the PubMed XML database and a detailed data inventory can be found here (https://www.nlm.nih.gov/databases/download/pubmed_medline.html). We obtained the journals' major scientific discipline from the Clarivate Journal Citation Report of 2018 via the unique journal ISSNs. We used journal names as a crosswalk to identify publications that appeared a year earlier in the exact same journals as the COVID articles.

An overview of the sample creation is provided in **Figure S1**. In service of estimation accuracy, we included only journals that are listed in Clarivate. By construction that excludes all COVID publications in journals that had no publication on record in PubMed for 2019. These journals likely only came into being in 2020. We restricted our search query to articles published between February 1st and September 10th of 2019, since these months were the most productive in terms of COVID publishing and we sought to mitigate seasonal influences, like gender differences in teaching load at certain times of year.

We used the forenames recorded in PubMed to designate the gender of authors (PubMed started to systematically record forenames in 2002). We determined the probable gender of the authors through the Genderize database, an established approach that allows gender assignment for a large number of authors. At the time of initial submission, Genderize included 86,710 distinct forenames drawn from 74 countries and 81 languages. Recent tests of the accuracy and comprehensiveness of four gender assignment algorithms, using a control sample of gender-matched forenames from a US government office, found that Genderize provided the most accurate estimates of gender (*1*). Our underlying code for calling the Genderize database with a large set of forenames has been posted to Figshare (*2*). Genderize uses a variety of information, such as social media records, to assign a probability that an individual with a particular forename is a man or a woman. For example, Genderize designates the forename "Chris" as male with 93% probability based on 8,631 verified records in the database. We considered gender determined if Genderize assigned a probability greater than chance (>50%) to preserve observations, given that the early stage of the COVID pandemic limits the set of associated publications. We designated the gender for more than 80% of the authors in our set. Of the authors designated, eight out of ten authors had an assigned gender probability of 90% or more (**Figure S2**). The gender designation for female first authors is equally accurate for both control and treatment group **(Figure S3)**. As such, our main findings do not change when setting different gender designation thresholds.

Next, we compared the distribution of disciplines producing COVID research relative to the articles in the control sample (**Table S1**). Ranking the disciplines in terms of publication output, and testing a Spearman Rank correlation, we obtain a coefficient of greater 0.85. While this correlation would generally be considered strong (*3*) lending credence to our basic design, it does not consider the possibility that men and women may sort differently into these fields. However, our **Figure 1a** in the main text documents that it is primarily fields where women tend to be well represented that produce COVID research.

To execute country-level analyses, we use regular expressions to extract the full country name or country codes from affiliation data for the first author. We also ranked countries by productivity for COVID-articles and control articles, obtaining a Spearman rank correlation of 0.94, again supporting our approach of using non-COVID articles in the prior year as a control group (**Table S2**). This also mitigates concerns that countries with larger gender gaps in general produce more COVID research.

Additional information on methods

Measurement

To assess the effect of the COVID pandemic on the gender gap in publishing, we reported unadjusted differences in the percent of female first authorships versus male first authorships for COVID and non-COVID publications. This straightforward metric provides a direct and easy to understand measure of how the COVID pandemic impacts women's versus men's publication productivity.

$$\begin{aligned}\Delta GenderGap &= \{FirstAuthor_{Female} - FirstAuthor_{Male} \mid COVID\} \\ &\quad - \{FirstAuthor_{Female} - FirstAuthor_{Male} \mid Non-COVID\}\end{aligned}$$

To conduct subgroup analysis for discipline and country, we calculated the change in the gender gap based on the percent of female and male first authorships for the specific discipline and country.

Estimation

In addition to the unadjusted differences, we also provided adjusted differences in female versus male first authorship obtained from linear probability models (**Table S3**), adjusting for field of research and country. Logistic regression as an alternative estimation model has two disadvantages in our analysis. First, the large number of fixed effects when including countries and discipline dummies, for example, raises the possibility of incidental parameters bias and could prevent the convergence of some of our models. Second, logistic regressions can overestimate effect sizes as a result of the high leverage of marginal cases (i.e., identifying larger gender differences than reported in the main text), whereas linear probability models average across observations and produce more conservative results (see also **Table S4**).

We provided adjusted estimates in the supplement as one might be concerned, for example, that men are more numerous in fields that produce COVID research. This would also lower women's observed COVID productivity but not due to pandemic related constraints as hypothesized, but rather due to underlying structural differences in subspecialties. Of note, the descriptive data paint a different picture, such that women tend to be at least equal if not overrepresented in the most productive COVID disciplines.

References

1. C. N. G. detection, (http://codingnews.info/post/genderdetection (accessed 11/15/17) (2015)).
2. Lerchenmueller. Marc, *Genderize_unlimited_API_request*. (2017).
3. M. G. Pagano, Kimberlee, *Principle of Biostatistics*. (Brooks/Cole, ed. Secon Edition, 2000).

Figure S1: Sample construction for COVID articles and non-COVID (control) articles

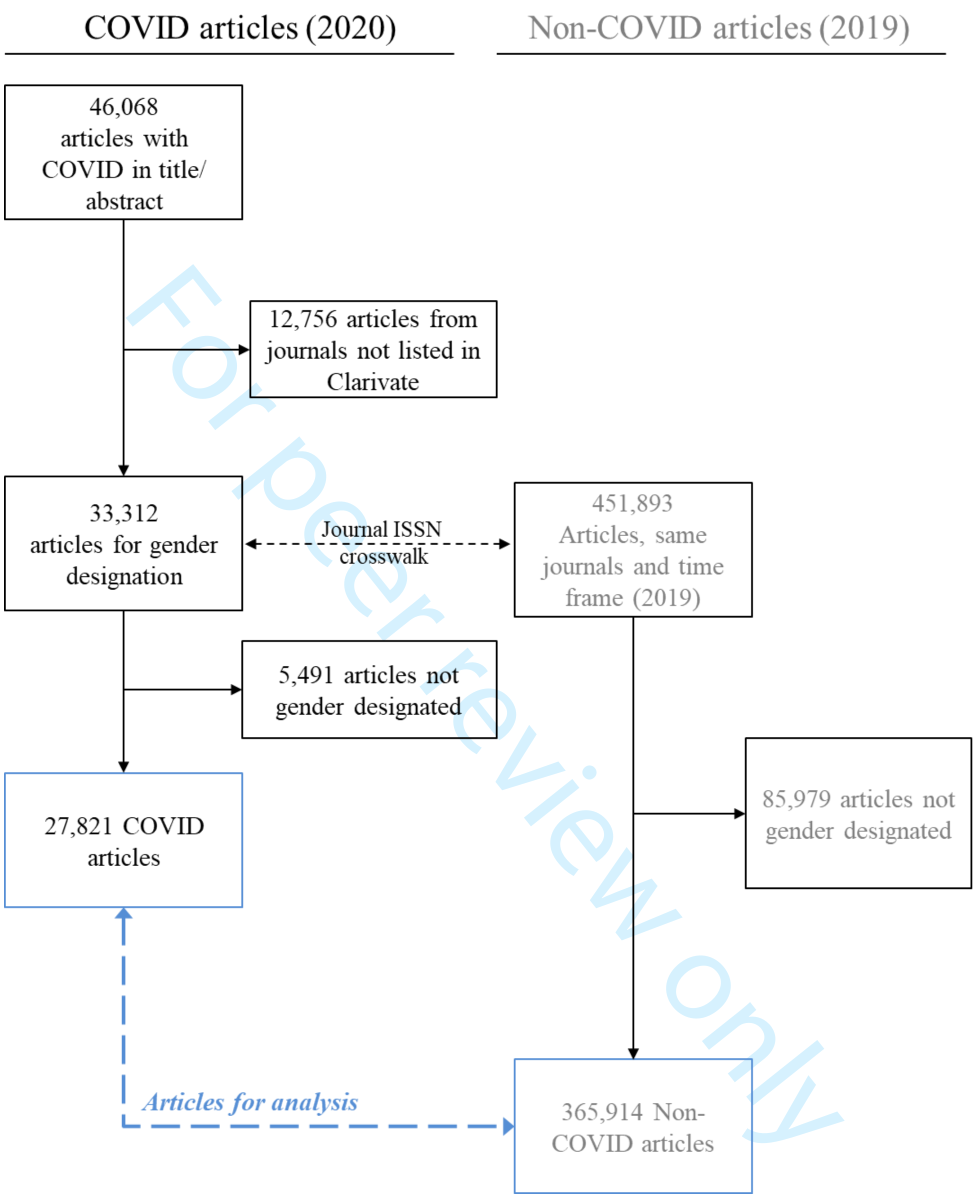

Figure S2: Gender designation accuracy for all articles in the sample (COVID articles and non-COVID articles) for female and male first authors

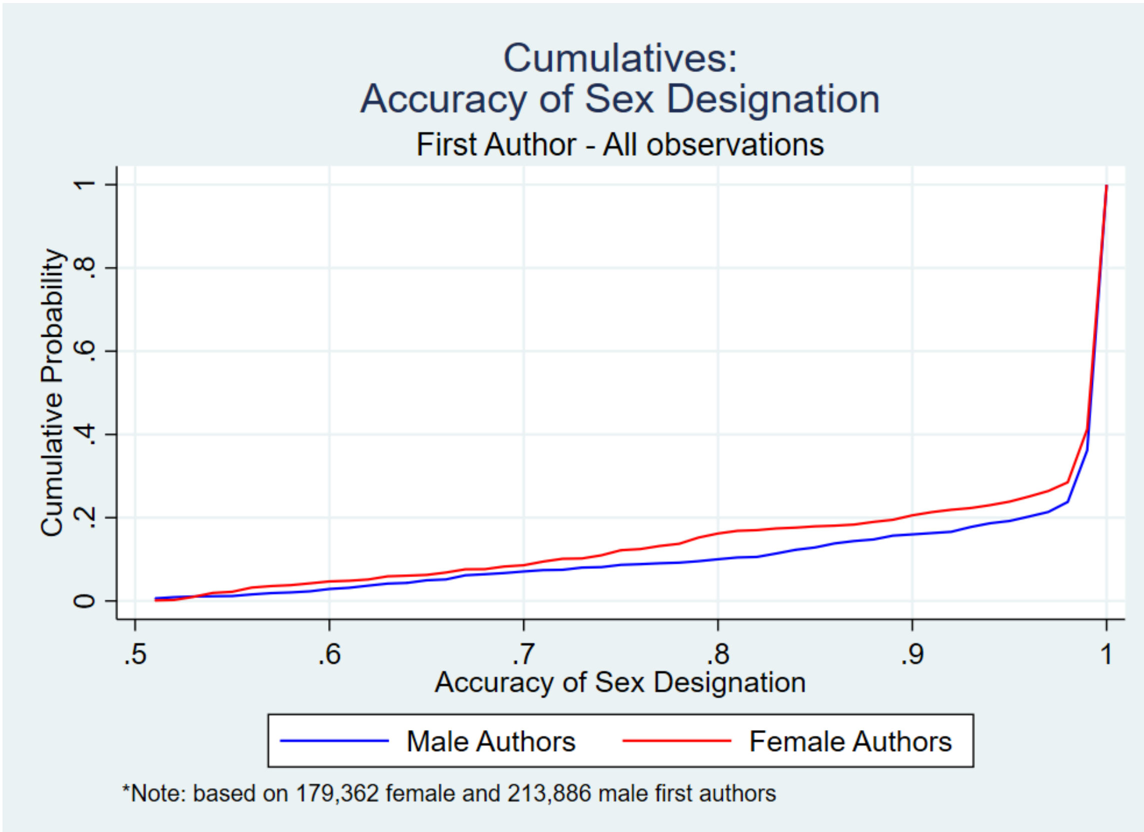

Figure S3: Gender designation accuracy for first authors for COVID articles and non-COVID (control) articles separately

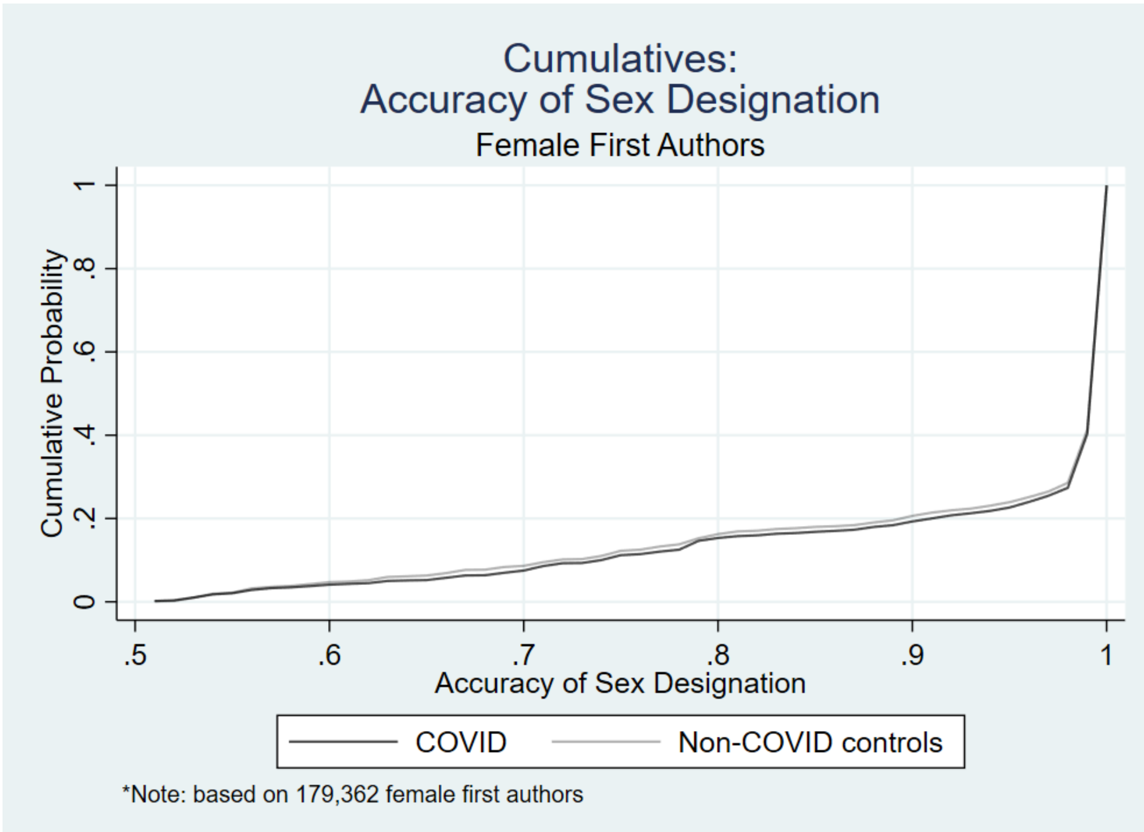

Table S1: Major disciplines for COVID articles and non-COVID (control) articles and concordance statistic

Discpline	Non-COVID			COVID			Total		
	Total	%	Rank	Total	%	Rank	Total	%	Rank
MEDICINE GENERAL & INTERNAL	20237	5.55%	4	2909	10.69%	1	23146	5.91%	3
PUBLIC ENVIRONMENTAL & OCCUPATIONAL HEAL	13059	3.58%	8	2147	7.89%	2	15206	3.88%	7
SURGERY	20514	5.63%	3	1726	6.34%	3	22240	5.68%	4
IMMUNOLOGY	9435	2.59%	11	1303	4.79%	4	10738	2.74%	10
CARDIAC & CARDIOVASCULAR SYSTEMS	11951	3.28%	9	1003	3.69%	5	12954	3.31%	9
PHARMACOLOGY & PHARMACY	14833	4.07%	6	887	3.26%	6	15720	4.01%	6
ONCOLOGY	18644	5.11%	5	857	3.15%	7	19501	4.98%	5
PSYCHIATRY	7773	2.13%	15	781	2.87%	8	8554	2.18%	15
INFECTIOUS DISEASES	2189	0.60%	40	756	2.78%	9	2945	0.75%	36
PEDIATRICS	8208	2.25%	13	692	2.54%	10	8900	2.27%	13
ENVIRONMENTAL SCIENCES	14000	3.84%	7	654	2.40%	11	14654	3.74%	8
CLINICAL NEUROLOGY	8093	2.22%	14	646	2.37%	12	8739	2.23%	14
BIOCHEMISTRY & MOLECULAR BIOLOGY	22672	6.22%	2	637	2.34%	13	23309	5.95%	2
VIROLOGY	1744	0.48%	46	627	2.30%	14	2371	0.61%	41
RADIOLOGY NUCLEAR MEDICINE & MEDICAL IMAC	5875	1.61%	17	620	2.28%	15	6495	1.66%	17
DERMATOLOGY	4267	1.17%	25	591	2.17%	16	4858	1.24%	23
MEDICINE RESEARCH & EXPERIMENTAL	5344	1.47%	20	583	2.14%	17	5927	1.51%	20
HEALTH CARE SCIENCES & SERVICES	3719	1.02%	31	551	2.02%	18	4270	1.09%	27
GASTROENTEROLOGY & HEPATOLOGY	5628	1.54%	18	549	2.02%	19	6177	1.58%	18
NEUROSCIENCES	9627	2.64%	10	520	1.91%	20	10147	2.59%	11
RESPIRATORY SYSTEM	3149	0.86%	34	509	1.87%	21	3658	0.93%	34
MULTIDISCIPLINARY SCIENCES	27786	7.62%	1	475	1.75%	22	28261	7.21%	1
HEMATOLOGY	3685	1.01%	32	449	1.65%	23	4134	1.05%	30
UROLOGY & NEPHROLOGY	5492	1.51%	19	445	1.63%	24	5937	1.52%	19
ANESTHESIOLOGY	1950	0.53%	44	436	1.60%	25	2386	0.61%	40
ENDOCRINOLOGY & METABOLISM	5240	1.44%	21	434	1.59%	26	5674	1.45%	21
RHEUMATOLOGY	2592	0.71%	36	327	1.20%	27	2919	0.74%	37
EMERGENCY MEDICINE	1701	0.47%	47	289	1.06%	28	1990	0.51%	46
OPHTHALMOLOGY	4503	1.23%	24	286	1.05%	29	4789	1.22%	25
OBSTETRICS & GYNECOLOGY	3564	0.98%	33	266	0.98%	30	3830	0.98%	33
NURSING	3747	1.03%	30	255	0.94%	31	4002	1.02%	32
MICROBIOLOGY	4589	1.26%	23	245	0.90%	32	4834	1.23%	24
GERIATRICS & GERONTOLOGY	1569	0.43%	51	240	0.88%	33	1809	0.46%	49
CELL BIOLOGY	6582	1.80%	16	227	0.83%	34	6809	1.74%	16
DENTISTRY ORAL SURGERY & MEDICINE	3126	0.86%	35	198	0.73%	35	3324	0.85%	35
CRITICAL CARE MEDICINE	1486	0.41%	54	188	0.69%	36	1674	0.43%	52
OTORHINOLARYNGOLOGY	1650	0.45%	50	185	0.68%	37	1835	0.47%	48
ORTHOPEDICS	2081	0.57%	42	146	0.54%	38	2227	0.57%	43
MEDICAL LABORATORY TECHNOLOGY	1116	0.31%	59	142	0.52%	39	1258	0.32%	57
SPORT SCIENCES	2180	0.60%	41	122	0.45%	40	2302	0.59%	42
ETHICS	478	0.13%	73	118	0.43%	41	596	0.15%	69
BIOTECHNOLOGY & APPLIED MICROBIOLOGY	4070	1.12%	28	115	0.42%	42	4185	1.07%	29
PATHOLOGY	1663	0.46%	48	115	0.42%	43	1778	0.45%	50
ECONOMICS	178	0.05%	88	112	0.41%	44	290	0.07%	79
PSYCHOLOGY MULTIDISCIPLINARY	2423	0.66%	37	101	0.37%	45	2524	0.64%	38
NUTRITION & DIETETICS	4124	1.13%	27	100	0.37%	46	4224	1.08%	28
GENETICS & HEREDITY	4248	1.16%	26	94	0.35%	47	4342	1.11%	26
PSYCHOLOGY CLINICAL	1770	0.49%	45	94	0.35%	48	1864	0.48%	47
BIOLOGY	1654	0.45%	49	89	0.33%	49	1743	0.44%	51
VETERINARY SCIENCES	2357	0.65%	38	85	0.31%	50	2442	0.62%	39

Spearman Rank Correlation - all disciplines	
coefficient (rs)	0.851
N	136
T statistic	18.765
DF	134
p-value	0.000

Spearman Rank Correlation - top 50 disciplines	
coefficient (rs)	0.690
N	50
T statistic	6.605
DF	48
p-value	0.000

Table S2: Major countries for COVID articles and non-COVID (control) articles and concordance statistics

Country (First Author)	**Non-COVID**			**COVID**			**Total**		
	Total	%	Rank	Total	%	Rank	Total	%	Rank
United States	101455	24.50%	1	8461	25.63%	1	109916	24.58%	1
Italy	17731	4.28%	5	3414	10.34%	2	21145	4.73%	4
China	50503	12.20%	2	2919	8.84%	3	53422	11.95%	2
United Kingdom	23817	5.75%	3	2687	8.14%	4	26504	5.93%	3
India	11992	2.90%	9	1575	4.77%	5	13567	3.03%	9
Spain	11762	2.84%	10	1070	3.24%	6	12832	2.87%	10
Canada	13661	3.30%	8	979	2.97%	7	14640	3.27%	8
France	10382	2.51%	13	969	2.94%	8	11351	2.54%	13
Brasil	10616	2.56%	12	877	2.66%	9	11493	2.57%	12
Germany	18279	4.41%	4	741	2.24%	10	19020	4.25%	5
Australia	13921	3.36%	7	732	2.22%	11	14653	3.28%	7
Iran	6232	1.50%	15	714	2.16%	12	6946	1.55%	15
Turkey	5751	1.39%	16	548	1.66%	13	6299	1.41%	16
Singapore	2005	0.48%	31	466	1.41%	14	2471	0.55%	27
Japan	16637	4.02%	6	378	1.15%	15	17015	3.81%	6
Switzerland	5004	1.21%	18	359	1.09%	16	5363	1.20%	18
South Korea	11499	2.78%	11	333	1.01%	17	11832	2.65%	11
Hong Kong	1600	0.39%	36	291	0.88%	18	1891	0.42%	34
Netherlands	7692	1.86%	14	285	0.86%	19	7977	1.78%	14
Taiwan	5473	1.32%	17	282	0.85%	20	5755	1.29%	17
Israel	3320	0.80%	22	275	0.83%	21	3595	0.80%	22
Pakistan	1638	0.40%	35	263	0.80%	22	1901	0.43%	33
Greece	2213	0.53%	27	254	0.77%	23	2467	0.55%	28
Belgium	3298	0.80%	23	250	0.76%	24	3548	0.79%	23
Saudi Arabia	2040	0.49%	30	245	0.74%	25	2285	0.51%	30
Mexico	2507	0.61%	24	213	0.65%	26	2720	0.61%	24
Poland	4543	1.10%	20	169	0.51%	27	4712	1.05%	20
Ireland	1700	0.41%	34	167	0.51%	28	1867	0.42%	35
Egypt	2147	0.52%	29	143	0.43%	29	2290	0.51%	29
Sweden	4579	1.11%	19	140	0.42%	30	4719	1.06%	19
Portugal	2485	0.60%	25	140	0.42%	31	2625	0.59%	25
Malaysia	1334	0.32%	40	126	0.38%	32	1460	0.33%	39
Austria	2404	0.58%	26	113	0.34%	33	2517	0.56%	26
South Africa	1441	0.35%	37	107	0.32%	34	1548	0.35%	37
Bangladesh	307	0.07%	65	97	0.29%	35	404	0.09%	63
Peru	383	0.09%	62	96	0.29%	36	479	0.11%	55
Denmark	3558	0.86%	21	90	0.27%	37	3648	0.82%	21
Colombia	740	0.18%	45	88	0.27%	38	828	0.19%	45
New Zealand	1783	0.43%	33	80	0.24%	39	1863	0.42%	36
Georgia	1080	0.26%	43	76	0.23%	40	1156	0.26%	43
Argentina	1250	0.30%	41	74	0.22%	41	1324	0.30%	41
Chile	1083	0.26%	42	74	0.22%	42	1157	0.26%	42
Indonesia	635	0.15%	48	74	0.22%	43	709	0.16%	48
United Arab Emirates	505	0.12%	53	72	0.22%	44	577	0.13%	52
Lebanon	619	0.15%	49	71	0.22%	45	690	0.15%	49
Nigeria	577	0.14%	50	70	0.21%	46	647	0.14%	50
Norway	2195	0.53%	28	67	0.20%	47	2262	0.51%	31
Morocco	267	0.06%	67	64	0.19%	48	331	0.07%	66
Jordan	386	0.09%	61	59	0.18%	49	445	0.10%	59
Thailand	709	0.17%	47	58	0.18%	50	767	0.17%	46

Spearman Rank Correlation - all countries	
coefficient (rs)	0.94
N	165
T statistic	34.61
DF	163
p-value	0.000

Spearman Rank Correlation - top 50 countries	
coefficient (rs)	0.86
N	50
T statistic	11.50
DF	48
p-value	0.000

Table S3: Hierarchical linear probability model for the likelihood of female first authorship for COVID articles versus non-COVID (control) articles

Dependent variable: First Author Female	(1)	(2)	(3)	(4)	(5)
COVID	**-0.089*** (0.00)**	**-0.089*** (0.00)**	**-0.090*** (0.00)**	**-0.091*** (0.00)**	**-0.096*** (0.00)**
number of authors		0.002*** (0.00)	0.002*** (0.00)	0.003*** (0.00)	0.003*** (0.00)
publication month fixed effects (8)			Included	Included	Included
discipline fixed effects (136)				Included	Included
country fixed effects (165)					Included
constant	0.462*** (0.00)	0.449*** (0.00)	0.444*** (0.00)	0.305*** (0.02)	0.146 (0.13)
R-squared	0.002	0.003	0.003	0.036	0.052
Adjusted R-squared	0.002	0.003	0.003	0.036	0.051
Observations	393,248	393,248	393,207	393,207	368,599

Note: standard errors in brackets, * p < 0.05, ** p< 0.01, *** p < 0.001

Table S4: Hierarchical logit regression for the likelihood of female first authorship for COVID articles versus non-COVID (control) articles

Dependent variable: First Author Female	(1)	(2)	(3)	(4)	(5)
COVID	**0.693*** (0.01)**	**0.693*** (0.01)**	**0.689*** (0.01)**	**0.677*** (0.01)**	**0.662*** (0.01)**
number of authors		1.009*** (0.00)	1.009*** (0.00)	1.011*** (0.00)	1.011*** (0.00)
publication month fixed effects (8)			Included	Included	Included
discipline fixed effects (136)				Included	Included
country fixed effects (165)					Included
constant	0.860*** (0.00)	0.813*** (0.00)	0.798*** (0.01)	0.441*** (0.04)	0.220** (0.12)
observations	393,248	393,248	393,207	393,202	368,543

Note: Coefficients reported as odds ratios, standard errors in brackets, * p < 0.05, ** p < 0.01, *** p < 0.001

Longitudinal Analyses of Gender Differences in First Authorship Publications Related to COVID-19

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-045176.R1
Article Type:	Original research
Date Submitted by the Author:	01-Mar-2021
Complete List of Authors:	Lerchenmüller, Carolin; University Hospital Heidelberg, Schmallenbach, Leo; University of Mannheim Jena, Anupam; Harvard University, Health Care Policy Lerchenmueller, Marc ; University of Mannheim,
Primary Subject Heading:	Health policy
Secondary Subject Heading:	Health services research, Medical publishing and peer review
Keywords:	GENERAL MEDICINE (see Internal Medicine), COVID-19, Health policy < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

Longitudinal Analyses of Gender Differences in First Authorship Publications Related to COVID-19

Carolin Lerchenmüller, MD,[1,2] Leo Schmallenbach, MSc,[3]
Anupam B. Jena, MD, PhD[4,5,6*], Marc J. Lerchenmueller, MPH, PhD[3*]

[1]Department of Cardiology, Angiology, Pulmonology, University Hospital Heidelberg, Germany
[2]German Center for Heart and Cardiovascular Research (DZHK), Partner Site Heidelberg/Mannheim, Germany
[3]Area Management, University of Mannheim, Mannheim, Germany
[4]Department of Health Care Policy, Harvard Medical School Boston, MA
[5]Department of Medicine, Massachusetts General Hospital, Boston, MA
[6]National Bureau of Economic Research, Cambridge, MA

*These authors contributed equally

Word count: 3837
Keywords: COVID-19, Health policy, Gender equity

Corresponding author:

Carolin Lerchenmueller, MD
Department of Cardiology, Angiology, Pulmonology
University Hospital Heidelberg
INF 410
69120 Heidelberg, Germany
Email: carolin.lerchenmueller@med.uni-heidelberg.de

Abstract

Objective: Concerns have been raised that the COVID pandemic has shifted research productivity to the disadvantage of women in academia, particularly in early career stages. In this study, we aimed to assess the pandemic's effect on women's' COVID-related publishing over the first year of the pandemic.

Methods and Results: We compared the gender distribution of first authorships for 42,898 publications on COVID-19 from February 1st, 2020 to January 31st, 2021 to 483,232 publications appearing in the same journals during the same period the year prior. We found that the gender gap – the percentage of articles on which men versus women were first authors – widened by 14 percentage points during the COVID pandemic, despite many pertinent research fields showing near equal proportions of men and women first authors publishing in the same fields before the pandemic. Longitudinal analyses revealed that the significant initial expansions of the gender gap began to trend backwards to expected values over time in many fields. As women may have been differentially affected depending on their geography, we also assessed the gender distribution of first authorships grouped by countries and continents. While we observed a significant reduction of the shares of women first authors in almost all countries, longitudinal analyses confirmed a resolving trend over time.

Conclusion: The reduction in women's COVID-related research output appears particularly concerning as many disciplines informing the response to the pandemic had near equal gender shares of first authorship in the year prior to the pandemic. The acute productivity drain with the onset of the pandemic magnifies deep rooted obstacles on the way to gender equity in scientific contribution.

Strengths and limitations of this study

- The COVID pandemic is an exogenous source of variation that allows the examination of differential effects of the pandemic on women's and men's publishing activity
- We used a retrospective cohort design, comparing author gender for COVID articles to articles in similar fields published during the year prior to the pandemic
- Data on affiliations, publishing journals and dates enable analyses of gender differences in publication rates by geography, scientific discipline, and over time
- This large-scale archival study did not allow disentangling the mechanisms that underpin gender differences in publishing rates associated with the pandemic
- The methodology relied on a probabilistic algorithm to assign gender to thousands of authors, bearing a residual risk of gender misclassification

Introduction

Women are integral to productive and innovative science communities.[1 2] Nonetheless, women remain underrepresented in prestigious author positions on publications in the life sciences and medicine,[3] are less likely to be promoted to higher academic ranks, and are paid less, despite the continuously growing number of women academics.[4] Projections indicate that this gap will persist if targeted interventions are not implemented.[5]

It is in this setting that concerns have been raised that research and expert reporting on the COVID (Coronavirus Disease 2019, COVID-19) pandemic has disproportionately involved men as scientific authors. For example, women submitted fewer manuscripts overall, were less available for peer review,[6-8] and attended fewer funding panel meetings.[7] Also, women first authorship was significantly reduced on preprints and publications about COVID in the US[9 10] and globally.[11]

It has been suggested that this might, at least in part, be due to an exacerbation of pre-existing work-family conflicts, especially for early-career mothers in academia.[12] With lockdown measures to prevent uncontrolled spread of the coronavirus came not only remote working, but also closures of childcare services like daycares and schools transferring teaching responsibilities often to mothers, without the possibility to involve family members in childcare that belong to the aging population who are particular vulnerable to severe illness from the coronavirus.[12-16] Evidence from Germany showed, for example, that women not only took over the physical load of increased childcare and household responsibilities, but also the mental load associated with taking care of the family during a pandemic.[17] Even prior to the pandemic, research has shown that childrearing and household work were tasks largely taken care of by women, thereby impacting women's academic careers more than men's. For example, parental leaves taken by men often result in increased productivity, while no such phenomenon can be observed for women.[18]

The proliferation of COVID related publications provide a unique window into these gendered dynamics for two reasons. First, COVID publications have been produced rapidly under unusual conditions that likely disfavor women scientists relative to usual conditions that can serve as a control. Second, COVID publications are mainly, though not exclusively, produced in the life sciences and medicine where long-standing authorship norms reserve the first author position to early career investigators leading the project, which allows estimating the repercussions for this group in particular. Therefore, we assessed the pandemic's effect on women's COVID-related scientific publishing over the first year of the pandemic by analyses of first authorships in a longitudinal approach. We further performed analyses to quantify the effect per scientific specialty and country affiliation, as women may have been differentially affected across specialties and geographic areas.

Methods

Study Design and Data

We use a retrospective cohort design, comparing the gender composition on the author byline of 42,898 PubMed indexed life science articles that included the term "COVID" in the title and/ or abstract and that were published between February 2020 and January 2021 to a set of 483,232 control articles published in the same journals a year earlier (see Supplement Materials for details on data and methods). The onset of the COVID-19 pandemic in early 2020 (and ensuing countermeasures like lockdowns, remote work etc.) serve as source of exogenous variation that affects authors of COVID articles but not authors who published research in the same journals prior to the COVID outbreak, i.e., our control group. To assess possible effect stratification across research areas, we allocated 3,426 represented journals to scientific disciplines based on the disciplines provided in the Clarivate Journal Citation Report using unique International Standard Serial Numbers (ISSNs) as a crosswalk. We restricted our analysis to disciplines with at least 50 publications per reported time period to provide more precise estimates. We obtained detailed affiliation records to determine the geographic locale of first authors for country-specific analyses. We restricted this analysis to countries with at least 50 publications, and continents with at least 10 publications per reported time point, to increase precision of estimates (see Figure S1 for details on the sample construction).

We further made use of a long-standing authorship norm in the life sciences, according to which the first author is usually the junior author who executed the research, while the last author is generally the senior author who funded and may have conceived of the research. To designate the probable gender of thousands of these authors in our dataset, we use the genderize.io database that draws on a number of official sources, like Social Security Administration records and social media profiles, to assign a probability that a given forename is more likely held by men or women. For our analysis, we only included cases where the algorithm assigned a 90% or greater probability to the individual being of a specific gender (see also Figures S2–S4). Overall, our applied inclusion criteria did not introduce tangible selection bias in terms of fields or countries represented (see Tables S1–S2).

Outcome measures

We calculated the gender gap in academic authorships as the absolute percentage point difference between men and women authors. For example, if men and women accounted for 55% and 45% of first author positions, respectively, the absolute gender difference would be 10 percentage points. We also offered parametric analyses in the supplement, analyzing the effect of authoring during the pandemic versus not (a binary independent variable) on the likelihood that the first author was a woman versus a man (our outcome) (Table S3–S5).

Sensitivity analyses

We conducted additional sensitivity analyses (Table S6-S7), including varying the confidence with which gender could be inferred, excluding articles with group authorships, rerunning our analyses for the full set of articles (i.e., without applying sampling restrictions), and comparing sampled to non-sampled observations.

Patient and Public Involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for design or implementation of the study. No patients were asked to advise on the interpretation or write up of the results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

Results

One-year gender differences in first and last authorships related to COVID publications

On average, men accounted for 54.9% and women for 45.1% of first authorships in 483,232 articles, published before the pandemic (between February 2019 and January 2020), for an absolute gender gap of 9.8 percentage points. In contrast, men and women accounted for 62.3% and 37.7% of first authorships on COVID-related publications, for an absolute gender gap of 24.6 percentage points. The gender gap therefore widened by approximately 14.8 percentage points in disciplines related to the COVID pandemic (Figure 1). Smaller changes were observed in last authorships. For example, while the gender gap in last authorships before the pandemic was expectedly much larger (approximately 36 percentage points), the effect of the pandemic on last author publishing by women was less pronounced (albeit statistically significant given the large sample size). On average, women accounted for 31.9% of last authorships in articles published before the pandemic, compared with 30.5% of last authorships for COVID-related publications, representing a widening of the last author gender gap by approximately 2.8 percentage points (Figure 1). In line with previous data, our results therefore indicate that junior women investigators were disproportionally affected by the pandemic.[9-11 13 15] As women may have been differently affected depending on their field of study or depending on geography, we further sought to investigate gender differences particularly in first authorships according to field of study and country.

Longitudinal analysis of gender differences in first authorships on COVID publications, by scientific discipline

As the first author gender gap in publications generally varies across fields,[19] we calculated the first author gender gap by discipline. We allocated the journals that published on COVID to scientific fields based on the disciplines represented in the Clarivate journal citation report.[20] As depicted in the first column of the heatmap in Figure 2, disciplines producing most COVID-relevant publications had shares of first authorship pre-COVID by women of 45% (i.e., an average gender gap of approximately 10 percentage points across all disciplines). During the pandemic, however, publications related to COVID had an average one-year gender gap of approximately 24 percentage points (i.e. the share of first authorships from women for publications related to COVID was 38%), a deviation of 14 percentage points (Figure 2). This effect was most prominent in the first months of COVID publishing, from February to May of 2020, when the gender gap rose to 36 percentage points (corresponding to a share of women first authorships of 32%). In the following months from June to September 2020, the share of female first authors slowly increased again to an average of 37%, and to 41% from October 2020 to January 2021, reducing the gender gap to 26 and 18 percentage points, respectively. However, this still represented a significant deviation from the pre-pandemic gender gap of 8 percentage points.

Interestingly, many of the disciplines that produced most COVID publications had equal or near-equal gender shares in the year before the pandemic. For example, in the fields of Virology, Immunology, Infectious Diseases, and General/Internal Medicine, the shares of women first authors prior to COVID were 50%, 52%, 48%, and 44%, respectively. The share of women first authors in other relevant fields like Public, Environmental and Occupational Health was even higher (59%). In many relevant fields, the gender rift in first authorships for COVID-related publications widened significantly above the pre-COVID average. For example, while women were more likely to be first authors on publications within Public, Environmental and Occupational Health (difference of 18 percentage points) before the pandemic, the gap changed by 30 percentage points so that women were now less likely to publish research within this field as first authors of COVID-related publications (gender gap of 12 percentage points). In Biochemistry and Molecular Biology, the gender gap in first authorships increased by 34 percentage points, from 2 percentage points pre-COVID to 36 percentage points for COVID publications. In Virology, the gender gap increased by 26 percentage points, from equal shares (no gap) to 26 percentage points for COVID-publications. Other COVID-relevant fields were much less prone to changes in the first author gender gap, for example Surgery, and Cardiac & Cardiovascular Systems, in which the gender gap prior to COVID was 44 percentage points compared with 46 percentage points for COVID-related studies, respectively (meaning women's first authorship shares were as low as 28 and 27%, respectively) (Figure 2).

Longitudinal analyses of gender differences in first authorships on COVID publications, by affiliated geographic area

Since the pandemic has affected countries differently, we further performed analyses based on the country affiliation of the first author. Women's research productivity went down in almost all countries (Figure 3a). For example, in the United States, which accounted for ~26% of all COVID publications between February 2020 and January 2021, women's first authorship share decreased from a share of 44% to a share of 41% (corresponding to a widening of the gender gap by 6 percentage points). In European countries that were hit earlier by the pandemic than the US, women's representation in authorships was also more affected. In Italy, for example, women's share of first authorships decreased from 49% before the pandemic to 35% for COVID-related publications, an increase in the gender gap of 28 percentage points because of the pandemic, with the overall number of publications from Italy accounting for 10% of total publications on COVID. The increase in the first authorship gender gap was also substantive in Brazil (30 percentage points), and Mexico (35 percentage points), Australia (14 percentage points), and India (22 percentage points). Only very few countries showed no change in the first authorship gender gap, including China (no change), South Korea (decreased by 3 percentage points), or Taiwan (decreased by 2 percentage points).

We further performed a granular time-resolved (per two-months) analysis of women's first authorship shares grouped by continents. Our data showed that the largest reduction in women's first authorship shares happened early in last spring (April and May 2020). In Europe, for example, the gender gap increased by 18 percentage points, in North America by 8 percentage points, in Latin America by 28 percentage points, in Australia and Oceania by 15 percentage points, in Africa by 18 percentage points, and in Asia by 7 percentage points (Figure 3b). Similar to our analysis of the difference in first authorship shares over time by field, we found that the gap began to close again over time and seemed to have reached baseline levels in North America and Oceania. In all the other continents, the gender gap has yet to reach the levels expected from the year prior to the pandemic, with Africa and Latin America being farthest from the baseline (12 percentage points, respectively) (Figure 3b), while COVID related research output has been relatively stable from April 2020 to January 2021 within each continent.

Discussion

Our results provide evidence for the COVID pandemic's effect on women's publishing productivity across disciplines, worldwide, and over time. In line with our hypothesis, we found that the relative increase in the gender gap was more pronounced for women in the first author position. In light of previous research and observations [13 15 18 21], we suspected that the overcontribution of women to household and child rearing responsibilities - that leaves less opportunity to participate in writing, submitting and publishing research related to COVID - led to reduced productivity of early career women investigators. However, aside from time constraints that disproportionally changed for women in an earlier stage of their career, other reasons are possible for the significant difference of women's productivity with regards to COVID. For example, since COVID emerged as a high-profile, and very publishable subject, it is possible that it was easier for men, that are still more likely to be in leadership and well-funded positions in academia to pick up the topic quickly. Also, women were less likely to function as expert reviewers on articles related to COVID,[6-8] known to potentially exacerbate a pre-existing gender bias in the peer-review and publishing process.[22-24] These might contribute to a vicious cycle that hindered access to COVID publishing especially for women in the early stages of their career.[11] The exact determination of underlying mechanisms, however, warrant future research that might also benefit from longer time series data.

Interestingly, we found that decreased publishing activity was specifically significant in fields that had a relatively equal share between women and men as first authors prior to the pandemic. Those were also fields that had a high overall productivity among COVID articles. This is a reason for concern insofar as the current scientific/medical response to one of the most incisive global crises could be overly dominated by men and missing expert voices by women that would usually be a vital part of this research.

Applying longitudinal analyses, we found that the much-increased gender gap in relevant fields was particularly noticeable early on and continuously trended back towards the baseline since then. However, recovery remains slow, with potential reasons being the extent of the impact, the fact that many fields with an above average share of women first authors were affected, but also because some of the most affected fields require in-person work, for example in wet labs (e.g. Biochemistry, Molecular Biology, Microbiology).

Analyzing the change in the gender gap globally, we similarly found that the shares of women first authors declined across almost all continents early in the course of the pandemic in spring of 2020. Since then, the gender gap slowly began to close again towards the expected baseline and even reached baseline in North America and Oceania. Asia is the only continent where no significant reduction in women's first authorships were noted. In China, there was no change in gender composition among COVID publication when compared to control publications. In Taiwan and South Korea, a small increase in women first author shares was noted. Given that these countries together accounted for 4.2% of COVID publications in our dataset, and that gender designation algorithms tend to offer lower probability gender designations for Asian forenames, we are reluctant to conclusively interpret these findings. Of note, we applied a uniform probability threshold of 90% for designating an author's gender as a conservative measure.

Even though our longitudinal analysis leads us to be cautiously optimistic that the impact of the pandemic on women's COVID research activity might have been temporary, we speculate that the absence of many expert women voices during the initial response to the pandemic impacted the individual researchers, but also society as a whole. While our study focused on COVID publications, the dynamics reported here may be amplified in research beyond COVID. Effects of the pandemic on early-stage or ongoing projects are likely to show with a time-delay and potentially have long-lasting consequences jeopardizing efforts toward equity in academia. For example, women at earlier career stages have not been able to allocate enough time to their research, manuscript and grant writing, were bound to remote working instead of in-person work, were potentially less likely to be allocated to leading roles on projects given the circumstances, etc.

To avoid long-term impacts on the academic advancement and scientific contributions, the disproportionate impact of the pandemic on early career women investigators needs serious consideration and immediate actions. In a first step we would suggest for an open discourse about how the pandemic has highlighted systemic and structural barriers preventing gender equity in academia.[25] Naturally, pre-existing inequities must be evaluated, and a long-term strategy has to be established to support equity in science. [26] But more acutely, COVID-related gender inequities need to be addressed with direct measures, monetary and non-monetary, on both the political (e.g. federal funding agencies) and institutional level. For example, financial support for postdoctoral/graduate students could help to facilitate a research set-back in a recently established laboratory. Modifications for grant deadlines, timelines, extensions for granted expenses, as well as additional (bridge) funding programs, are likely warranted for early career mentored/independent investigators.[27] Extension of tenure evaluation and promotion should be considered on the individual level accounting for constraints posed by COVID for junior faculty (for both men and women). Resources for childcare should be provided for parents, additional funds for expanded childcare arrangements could help to reallocate time to regular professional duties. However, it is as necessary to normalize the increased stress of living through a pandemic that not only affects professional obligations and goals, but also other family members and no penalty should be awarded for caregivers but measures mentioned above should rather allow for extra quality-time.[28]

Our study had several limitations. One limitation is that part of our large-scale study design was based on field-association by Clarivate Journal categories, which bears a potential risk of misclassification or inclusion of articles from journals that might not follow the norm of author ordering with regards to contribution that we assume for the life sciences and medicine. For example, one of the most affected fields in our data analysis is Public, Environmental and Occupational Health, where such norms might not be generalizable. However, previous research about the topic in Public Health, for example, also applied said authorship order norms and given the high relevance of the field and results, we decided to present the data.[29] Next, we relied on authors first names to designate their likely gender, which bears the risk of gender misclassification, particularly across different geographies. We attempted to minimize this risk by applying a 90% probability requirement, however, a certain level of uncertainty remains.[30] Also, by design of the gender designation algorithms grouping into two categories, namely 'man' and 'women' we cannot separate out an effect for scholars who are non-binary, transgender men and women. Along those lines, by virtue of the large-scale nature of our study, we acknowledge that we cannot draw conclusions for researchers on the individual level as well as confirm the assumed career stage. While our study focused on gender disparities for COVID-related research, it is important to note that, beyond gender diversity,[31] ethnic and cultural diversity benefit science on multiple levels.[32] With our analyses, however, we cannot comment on how the pandemic might have affected ethnic and cultural diversity with regards to COVID-related research or if populations under-represented in academic life sciences were similarly affected. Lastly, in this observational study, we cannot causally decipher the underlying mechanisms leading to women being underrepresented on COVID-related research, also precluding a definitive explanation for the trend back to baseline over time and what the potentially successful measures were or could be.

In conclusion, we found that women first authors have been underrepresented in COVID-related research, particularly at the beginning of the pandemic, despite having nearly equal first authorship shares as men in pertinent fields prior to the pandemic.

Figure Captions

Figure 1. Overall gender differences in first and last authorships related to COVID publications.
Gender gap in first authorships for COVID publications (red) compared to the gender gap for control publications appearing in the same journals a year earlier (blue).

Figure 2. Time-resolved gender differences in first authorship shares on COVID publications, by scientific discipline
Heatmap depicting the gender gap in first authorships for COVID publications and control publications from the same disciplines appearing in the same journals during the same period in the year prior to the pandemic. Fields sorted in descending order by number of publications. Red indicates an overrepresentation of women first authors, white indicates gender parity, blue indicates an overrepresentation of men first authors (in percentage points).

Figure 3. Difference in first authorship gender gap, by country/geographical area
a. World map depicting the deviation in the gender gap in first authorships for COVID publications when compared to the expected gender gap derived from control publications from the same countries appearing in the same journals a year earlier. Red indicates an overrepresentation of women first authors, white indicates gender parity, blue indicates an overrepresentation of men first authors (in percentage points).
b. Time-resolved deviation in the gender gap in first authorships for COVID publications when compared to the expected gender gap derived from control publications from the same geographical area appearing in the same journals a year earlier (in percentage points).

References

1. World Economic Forum. The Global Gender Gap Report 2017. *http://www3weforumorg/docs/WEF_GGGR_2017pdf (accessed 08 June 2018)*
2. O'Brien KR, Scheffer M, van Nes EH, et al. How to Break the Cycle of Low Workforce Diversity: A Model for Change. *PloS one* 2015;10(7) doi: ARTN e0133208
10.1371/journal.pone.0133208
3. Lerchenmuller C, Lerchenmueller MJ, Sorenson O. Long-Term Analysis of Sex Differences in Prestigious Authorships in Cardiovascular Research Supported by the National Institutes of Health. *Circulation* 2018;137(8):880-82. doi: 10.1161/CIRCULATIONAHA.117.032325
4. Blumenthal DM, Olenski AR, Yeh RW, et al. Sex Differences in Faculty Rank Among Academic Cardiologists in the United States. *Circulation* 2017;135(6):506-17. doi: 10.1161/CIRCULATIONAHA.116.023520
5. Holman L, Stuart-Fox D, Hauser CE. The gender gap in science: How long until women are equally represented? *PLoS Biol* 2018;16(4):e2004956. doi: 10.1371/journal.pbio.2004956 [published Online First: 2018/04/20]
6. Squazzoni F, Bravo, G., Grimaldo, F., Garcia-Costa, D., Farjam, M., Mehmani, B. Only Second-Class Tickets for Women in the COVID-19 Race. A Study on Manuscript Submissions and Reviews in 2329 Elsevier Journals. *SSRN Journal* 2020 [published Online First: October 16, 2020]

7. Krukowski RA, Jagsi R, Cardel MI. Academic Productivity Differences by Gender and Child Age in Science, Technology, Engineering, Mathematics, and Medicine Faculty During the COVID-19 Pandemic. *J Womens Health (Larchmt)* 2020 doi: 10.1089/jwh.2020.8710 [published Online First: 2020/11/21]
8. Mogensen MA, Lee, C.I., Carlos, R.C. The Impact of the COVID-19 Pandemic on Journal Scholarly Activity Among Female Contributors. *Journal of the American College of Radiology* 2021 doi: https://doi.org/10.1016/j.jacr.2021.01.011
9. Andersen JP, Nielsen MW, Simone NL, et al. COVID-19 medical papers have fewer women first authors than expected. *Elife* 2020;9 doi: 10.7554/eLife.58807 [published Online First: 2020/06/17]
10. Kibbe MR. Consequences of the COVID-19 Pandemic on Manuscript Submissions by Women. *JAMA Surg* 2020;155(9):803-04. doi: 10.1001/jamasurg.2020.3917 [published Online First: 2020/08/05]
11. Pinho-Gomes AC, Peters S, Thompson K, et al. Where are the women? Gender inequalities in COVID-19 research authorship. *BMJ Glob Health* 2020;5(7) doi: 10.1136/bmjgh-2020-002922 [published Online First: 2020/06/13]
12. Yildirim TM, Eslen-Ziya H. The Differential Impact of COVID-19 on the Work Conditions of Women and Men Academics during the Lockdown. *Gend Work Organ* 2020 doi: 10.1111/gwao.12529 [published Online First: 2020/09/10]
13. Staniscuaski F, Reichert F, Werneck FP, et al. Impact of COVID-19 on academic mothers. *Science* 2020;368(6492):724. doi: 10.1126/science.abc2740 [published Online First: 2020/05/16]
14. Oleschuk M. Gender Equity Considerations for Tenure and Promotion during COVID-19. *Can Rev Sociol* 2020;57(3):502-15. doi: 10.1111/cars.12295 [published Online First: 2020/08/12]
15. Kitchener C. Women academics seem to be submitting fewer papers during coronavirus . 'Never seen anything like it,' says one editor. *The Lily / The Washington Post* 2020 April 24, 2020.
16. Myers KR, Tham WY, Yin Y, et al. Unequal effects of the COVID-19 pandemic on scientists. *Nat Hum Behav* 2020;4(9):880-83. doi: 10.1038/s41562-020-0921-y [published Online First: 2020/07/17]
17. Czymara CS, Langenkamp A, Cano T. Cause for concerns: gender inequality in experiencing the COVID-19 lockdown in Germany. *European Societies* 2020:1-14. doi: 10.1080/14616696.2020.1808692
18. Antecol H, Bedard K, Stearns J. Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies? *American Economic Review* 2018;108(9):2420-41. doi: 10.1257/aer.20160613
19. Jagsi R, Guancial EA, Worobey CC, et al. The "gender gap" in authorship of academic medical literature--a 35-year perspective. *The New England journal of medicine* 2006;355(3):281-7. doi: 10.1056/NEJMsa053910 [published Online First: 2006/07/21]
20. Clarivate. 2018 Journal Citation Reports 2018 [Available from: https://clarivate.com/webofsciencegroup/wp-content/uploads/sites/2/2019/10/Crv_JCR_Full-Marketing-List_A4_2018_v4.pdf accessed 20 February 2020.
21. BCG. Easing the covid-19 burden on working parents 2020 [Available from: https://www.bcg.com/en-us/publications/2020/helping-working-parents-ease-the-burden-of-covid-19.aspx accessed 14 September 2020.

22. Clark J, Horton R. What is The Lancet doing about gender and diversity? *Lancet* 2019;393(10171):508-10. doi: 10.1016/S0140-6736(19)30289-2 [published Online First: 2019/02/12]
23. Holmes A, Hardy S. Gender bias in peer review–opening up the black box. *London, UK: London School of Economics* 2019
24. Pinho-Gomes AC. The Time Is Ripe for Addressing Gender Inequalities in the Authorship of Scientific Papers. *Am J Public Health* 2021;111(1):15-16. doi: 10.2105/AJPH.2020.306028 [published Online First: 2020/12/17]
25. Coe IR, Wiley R, Bekker LG. Organisational best practices towards gender equality in science and medicine. *Lancet* 2019;393(10171):587-93. doi: 10.1016/S0140-6736(18)33188-X [published Online First: 2019/02/12]
26. Malisch JL, Harris BN, Sherrer SM, et al. Opinion: In the wake of COVID-19, academia needs new solutions to ensure gender equity. *Proceedings of the National Academy of Sciences of the United States of America* 2020;117(27):15378-81. doi: 10.1073/pnas.2010636117 [published Online First: 2020/06/20]
27. Gibson EM, Bennett FC, Gillespie SM, et al. How Support of Early Career Researchers Can Reset Science in the Post-COVID19 World. *Cell* 2020;181(7):1445-49. doi: 10.1016/j.cell.2020.05.045 [published Online First: 2020/06/14]
28. Spinelli M, Lionetti F, Pastore M, et al. Parents' Stress and Children's Psychological Problems in Families Facing the COVID-19 Outbreak in Italy. *Front Psychol* 2020;11:1713. doi: 10.3389/fpsyg.2020.01713 [published Online First: 2020/07/29]
29. Bittante C, Beatrice G, Carletti L, et al. Gender disparity in authorships of manuscripts on the COVID-19 outbreak. *Z Gesundh Wiss* 2020:1-2. doi: 10.1007/s10389-020-01323-x [published Online First: 2020/08/25]
30. Lerchenmueller MJ, Sorenson O, Jena AB. Gender differences in how scientists present the importance of their research: observational study. *Bmj* 2019;367:l6573. doi: 10.1136/bmj.l6573 [published Online First: 2019/12/18]
31. Nielsen MW, Alegria S, Borjeson L, et al. Opinion: Gender diversity leads to better science. *Proceedings of the National Academy of Sciences of the United States of America* 2017;114(8):1740-42. doi: 10.1073/pnas.1700616114 [published Online First: 2017/02/24]
32. Science benefits from diversity. *Nature* 2018;558(7708):5. doi: 10.1038/d41586-018-05326-3 [published Online First: 2019/05/12]

Availability of data and materials

All data and source code generated and analyzed during the current study will be made publicly available with publication of this article.

Conflicts of interest

Dr. Jena reports receiving consulting fees unrelated to this work from Pfizer, Hill Rom Services, Bristol Myers Squibb, Novartis, Amgen, Eli Lilly, Vertex Pharmaceuticals, AstraZeneca, Celgene, Tesaro, Sanofi Aventis, Biogen, Precision Health Economics, and Analysis Group. Also unrelated to this work, Dr. Lerchenmueller reports serving on the board of AaviGen.

Funding

Support was provided by the Office of the Director, National Institutes of Health (1DP5OD017897, Dr. Jena). The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Author's contributions

LS and MJL collected data, LS, MJL and CL performed data analysis. CL, MJL, ABJ wrote the manuscript, LS edited the manuscript. All authors read and approved the final manuscript. Patients and public did not contribute to the study.

Acknowledgements

The authors acknowledge support by the German state of Baden-Württemberg through bwHPC—high performance cluster computing resources. The authors also thank Tobias Engel and Felix Fieberg for supporting data collection and literature review.

Supplementary Materials

Detailed methods and supporting data are available in the supplementary materials.

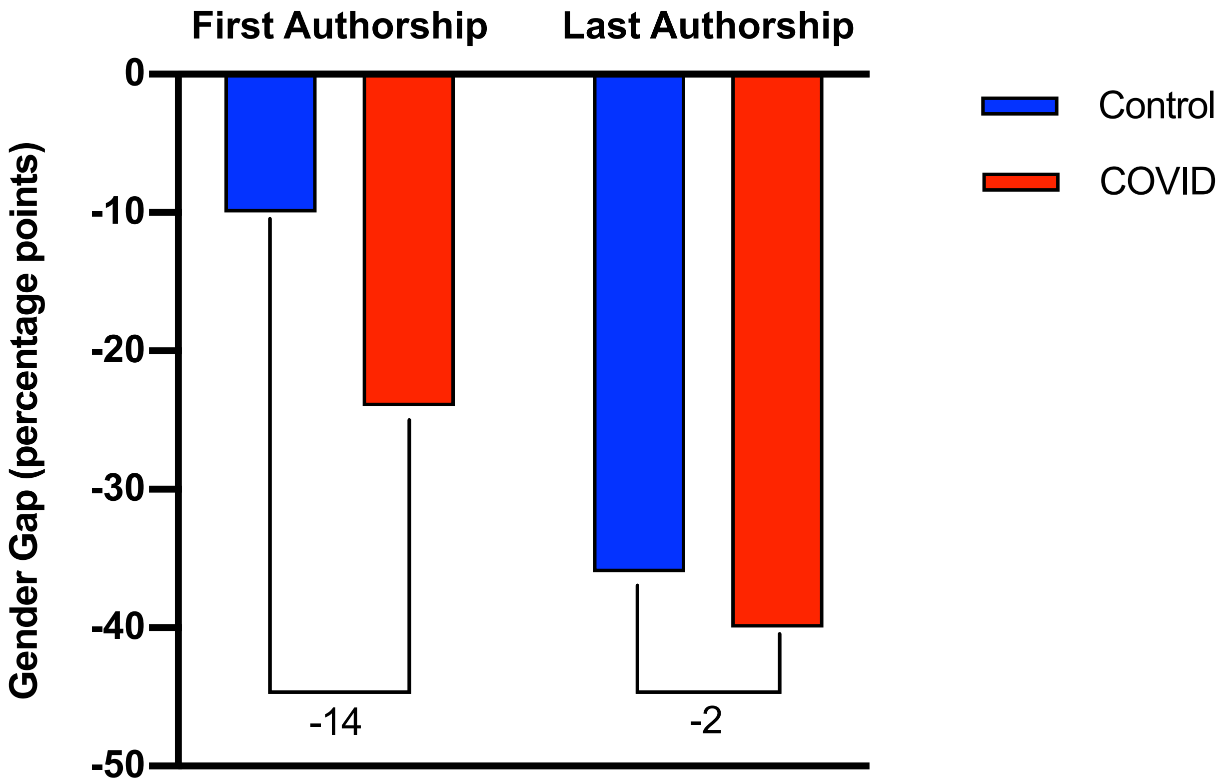

Figure 1. Overall gender differences in first and last authorships related to COVID publications.

Gender gap in first authorships for COVID publications (red) compared to the gender gap for control publications appearing in the same journals a year earlier (blue).

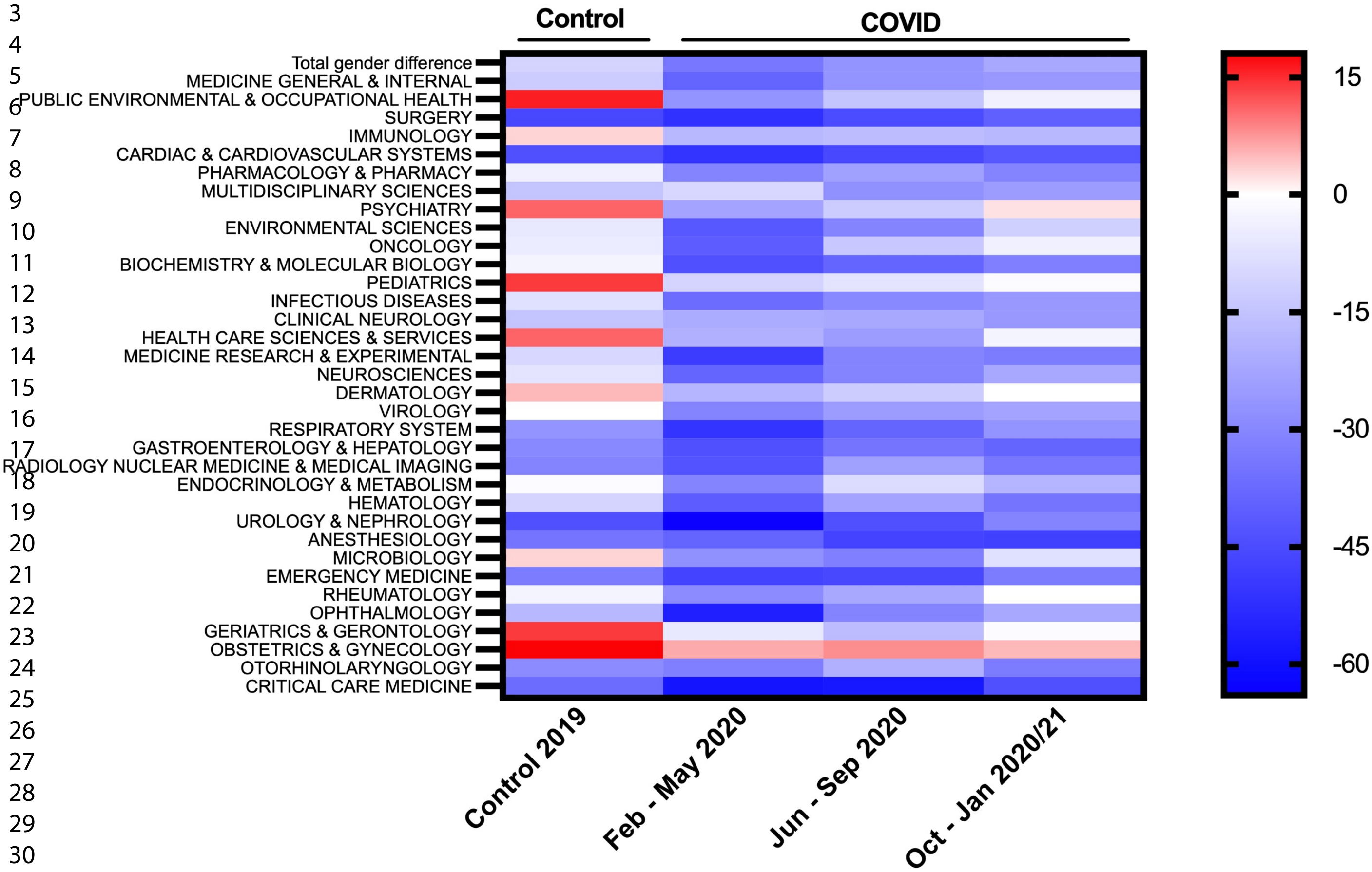

Figure 2. Time-resolved gender differences in first authorship shares on COVID publications, by scientific discipline

Heatmap depicting the gender gap in first authorships for COVID publications and control publications from the same disciplines appearing in the same journals during the same period in the year prior to the pandemic. Fields sorted in descending order by number of publications. Red indicates an overrepresentation of women first authors, white indicates gender parity, blue indicates an overrepresentation of men first authors (in percentage points).

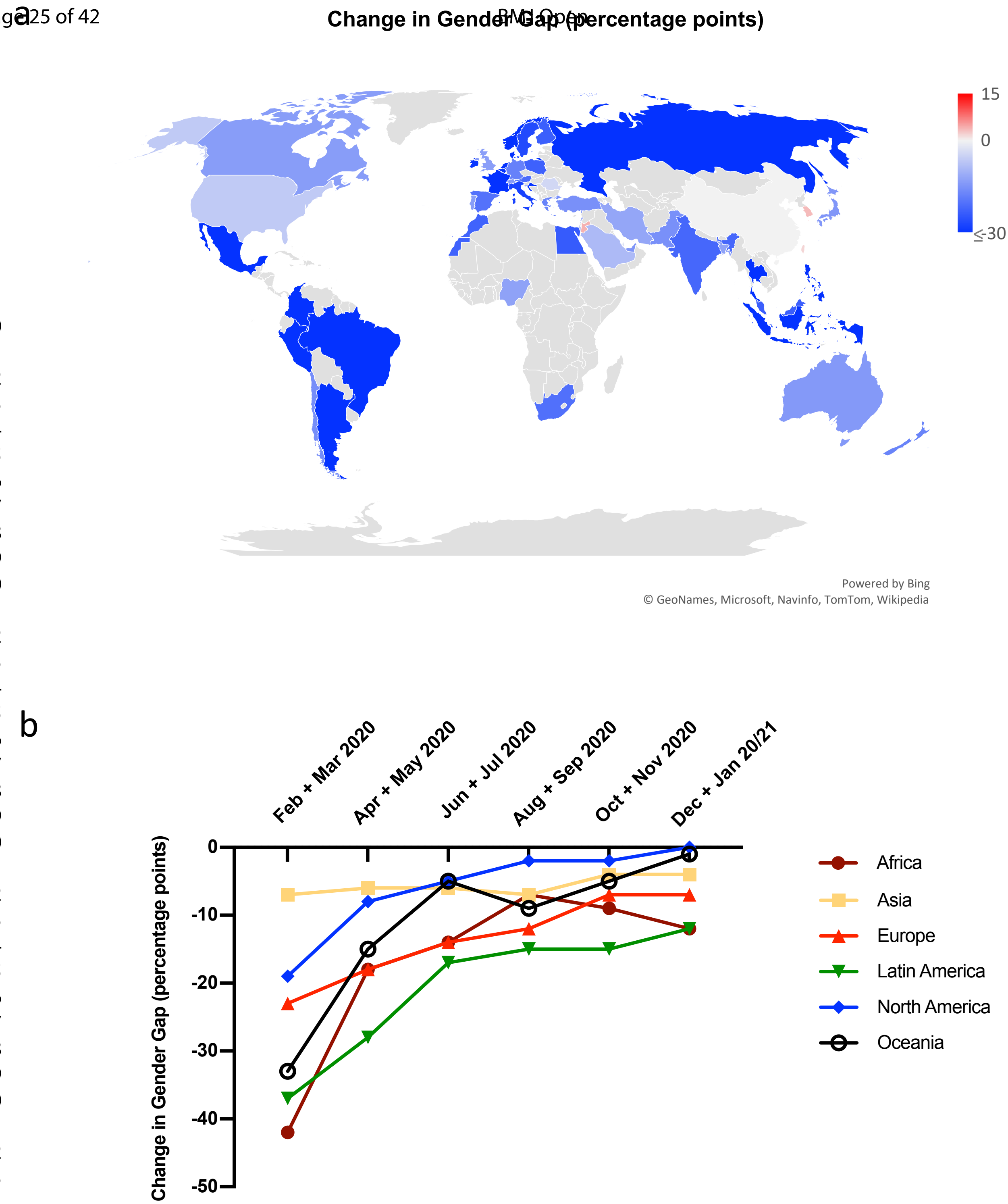

Figure 3. Difference in first authorship gender gap, by country/geographical area

a. World map depicting the deviation in the gender gap in first authorships for COVID publications when compared to the expected gender gap derived from control publications from the same countries appearing in the same journals a year earlier. Red indicates an overrepresentation of women first authors, white indicates gender parity, blue indicates an overrepresentation of men first authors (in percentage points). **b.** Time-resolved deviation in the gender gap in first authorships for COVID publications when compared to the expected gender gap derived from control publications from the same continent appearing in the same journals a year earlier (in percentage points).

Supplementary materials for

Longitudinal Analyses of Gender Differences in First Authorship Publications Related to COVID-19

Carolin Lerchenmüller, MD, Leo Schmallenbach, MSc,
Anupam B. Jena, MD, PhD, Marc J. Lerchenmüller, MPH, PhD

Table of Contents

Additional Information on Data

We merged several databases to analyze potential gender differences in first authorships of COVID publications relative to a set of control publications in the same journals and within the same time period one year earlier. First, we extracted all articles from the PubMed database for which the term "COVID" appeared in the title or abstract and obtained all available article characteristics including, among others, the names of all authors, country affiliation per author, the journal ISSN (International Standard Serial Number), and time of publication (months and year). The U.S. National Library of Medicine maintains the PubMed XML database and a detailed data inventory can be found online (https://www.nlm.nih.gov/databases/download/pubmed_medline.html). We obtained the journals' major scientific discipline from the Clarivate Journal Citation Report of 2018 via the unique journal ISSNs. We used journal names as a crosswalk to identify publications that appeared a year earlier in the exact same journals as the COVID articles.

An overview of the sample creation is provided in **Figure S1**. In service of estimation accuracy, we included only journals that are listed in Clarivate. By construction that excludes all COVID publications in journals that had no publication on record in PubMed for 2019. These journals likely only came into being in 2020. We restricted our search query to articles published between February 1st of 2020 and January 31st of 2021, since these months were the most productive in terms of COVID publishing and we sought to mitigate seasonal influences, like gender differences in teaching load at certain times of year.

We used the forenames recorded in PubMed to designate the gender of authors (PubMed started to systematically record forenames in 2002). We determined the probable gender of the authors through the Genderize database, an established approach that allows gender assignment for a large number of authors. At the time of initial submission, Genderize included 86,710 distinct forenames drawn from 74 countries and 81 languages. Recent tests of the accuracy and comprehensiveness of four gender assignment algorithms, using a control sample of gender-matched forenames from a US government office, found that Genderize provided the most accurate estimates of gender (*1*). Our underlying code for calling the Genderize database with a large set of forenames has been posted to Figshare (*2*). Genderize uses a variety of information, such as social media records, to assign a probability that an individual with a particular forename is a man or a woman. For example, Genderize designates the forename "Chris" as male with 93% probability based on 8,631 verified records in the database. We considered gender determined if Genderize assigned a probability of greater than 90%. Applying this threshold, we designated the gender for more than 72% of the authors in our dataset. However, there is variation across author origins (**Figure S2**). For example, we designated the gender for 84% of authors with an affiliation from North America and for 52% of authors with an affiliation from Asia. The lower accuracy for authors from Asia is a common challenge in name-based gender designation and a limitation to our analysis of authors from these countries. Yet, there is no difference in the accuracy of gender designation across men and women authors (**Figure S3**) or COVID and non-COVID articles (**Figure S4**). Hence, there is no reason to be concerned that the gender designation would systematically bias our results. Additionally, our main findings do not change when setting different gender designation thresholds.

Next, we compared the distribution of disciplines producing COVID research relative to the articles in the control sample (**Table S1**). Ranking the disciplines in terms of publication output, and testing a Spearman Rank correlation, we obtain a coefficient of greater 0.80. While this correlation would generally be considered strong (*3*) lending credence to our basic design, it does not consider the possibility that men and women may sort differently into these fields. However, our **Figure 2** in the main text documents that it is primarily fields where women tend to be well represented that produce COVID research.

To execute country-level analyses, we use regular expressions to extract the full country name or country codes from affiliation data for the first author. We also ranked countries by productivity for COVID-articles and control articles, obtaining a Spearman rank correlation of 0.94, again supporting our approach of using non-COVID articles in the prior year as a control group (**Table S2**). This also mitigates concerns that countries with larger gender gaps in general produce more COVID research.

Additional information on methods

Measurement

To assess the effect of the COVID pandemic on the gender gap in publishing, we reported unadjusted differences in the percent of women first authorships versus male first authorships for COVID and non-COVID publications. This straightforward metric provides a direct and easy to understand measure of how the COVID pandemic impacts women's versus men's publication productivity.

$$\begin{aligned}\Delta GenderGap \; &= \{FirstAuthor_{Female} - \; FirstAuthor_{Male} \mid COVID\} \\ &\quad - \{FirstAuthor_{Female} - \; FirstAuthor_{Male} \mid Non - COVID\}\end{aligned}$$

To conduct subgroup analysis for discipline and country, we calculated the change in the gender gap based on the percent of first authorships by men and women for the specific discipline and country.

Estimation

In addition to the unadjusted differences, we also provided adjusted differences in first authorships from women and men obtained from linear probability models (**Table S3**), adjusting for the number of authors on a publication, the month of publication, the field of research and country. We run the same analysis for last authorships from women and men (**Table S4**). Both regression analyses support the descriptive evidence presented in Figure 1 of the main text. Logistic regression as an alternative estimation model has two disadvantages in our analysis. First, the large number of fixed effects when including countries and discipline dummies, for example, raises the possibility of incidental parameters bias and could prevent the convergence of some of our models. Second, logistic regressions can overestimate effect sizes as a result of the high leverage of marginal cases (i.e., identifying larger gender differences than reported in the main text), whereas linear probability models average across observations and produce more conservative results (see also **Table S5**).

We provided adjusted estimates in the supplement as one might be concerned, for example, that men are more numerous in fields that produce COVID research. This would also lower women's observed COVID productivity but not due to pandemic related constraints as hypothesized, but rather due to underlying structural differences in subspecialties. Of note, the descriptive data paint a different picture, such that women tend to be at least equal if not overrepresented in the most productive COVID disciplines.

We conducted four robustness checks to establish the reliability of our findings (**Table S6**). In the first two robustness checks, we vary the threshold applied to the accuracy of the gender designation. In Model 1, we consider all authors, for which gender was assigned with a probability higher than chance (>50%). In Model 2, we only consider authors, for which the gender designation accuracy was reported with 100%. Both models show very similar estimates for the decrease in women authorship on COVID publications (8.2%-points vs. 9.0%-points). Next, we excluded articles from the analysis, for which collective authorship was indicated in PubMed. This concerns roughly 8% of articles but excluding them does not alter the effect estimate. Last, we reran the analysis on the full sample, that is including COVID articles published in journals, which are not listed in Clarivate's journal citation report and for which the first author's gender could be designated. As we do not know the disciplines these journals fall into, we include journal instead of discipline fixed effects in this last model specification. Again, the results are consistent with our previous analysis. Accordingly, a descriptive comparison of the articles in- and excluded from the analysis shows that they are near identical with respect to the representation of women first and last authors (**Table S7**).

References

1. C. N. G. detection, (http://codingnews.info/post/genderdetection (accessed 11/15/17) (2015)).
2. Lerchenmueller. Marc, *Genderize_unlimited_API_request*. (2017).
3. M. G. Pagano, Kimberlee, *Principle of Biostatistics*. (Brooks/Cole, ed. Secon Edition, 2000).

Figure S1: Sample construction for COVID articles and non-COVID (control) articles

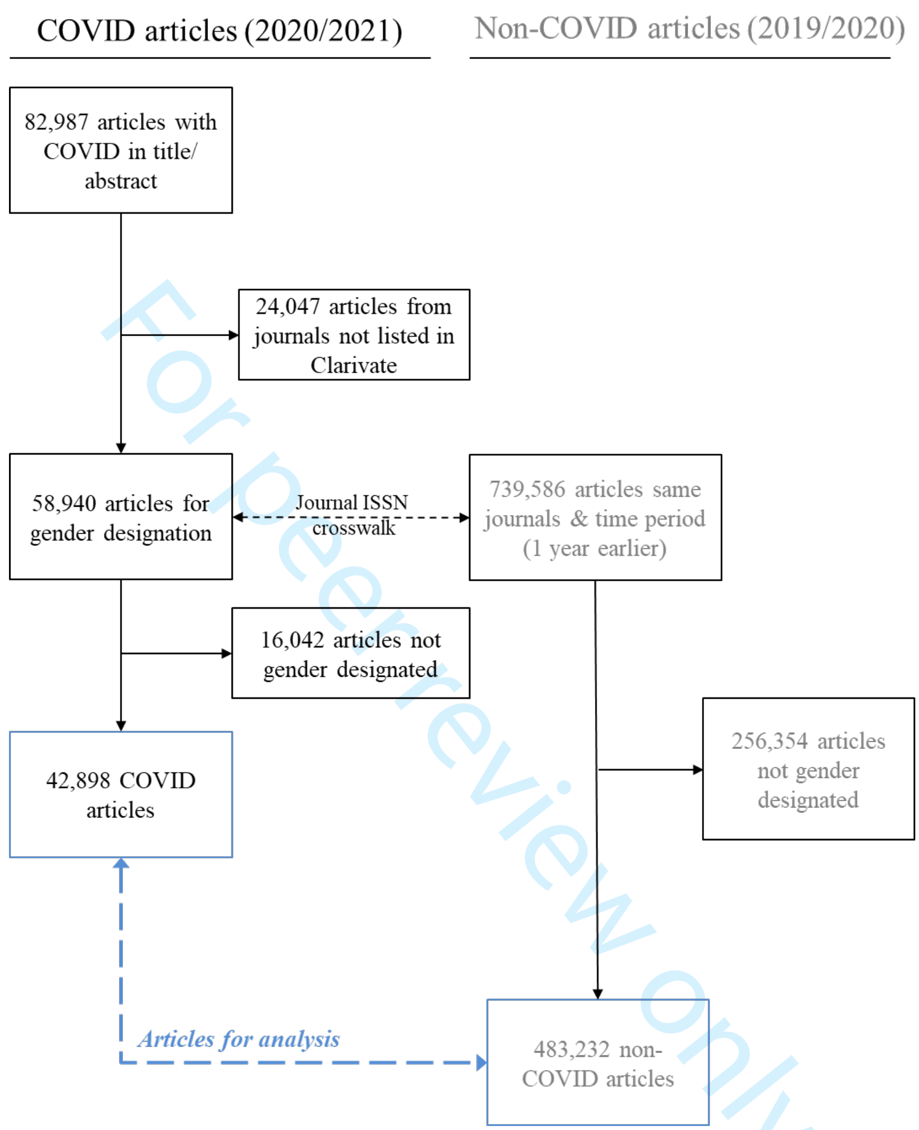

Figure S2: Gender designation accuracy for first authors from North America, Latin America, Europe, Asia, Oceania, and Africa separately

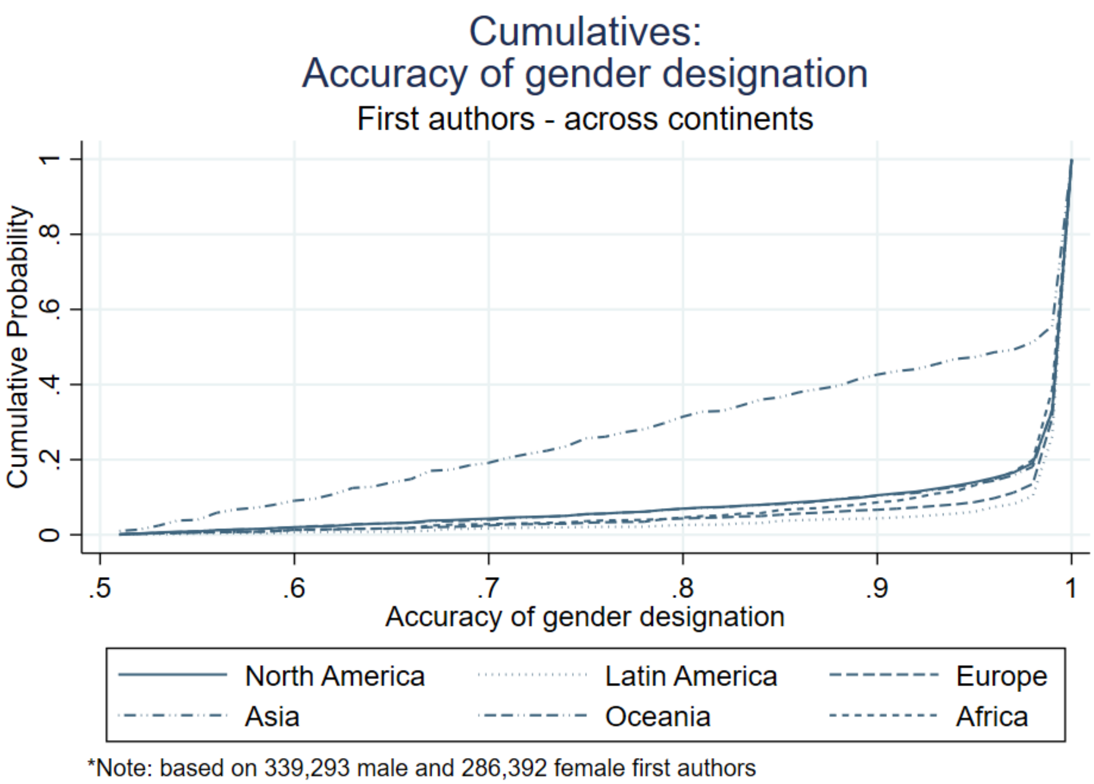

Figure S3: Gender designation accuracy for all articles in the sample (COVID articles and non-COVID articles) for women and men first authors

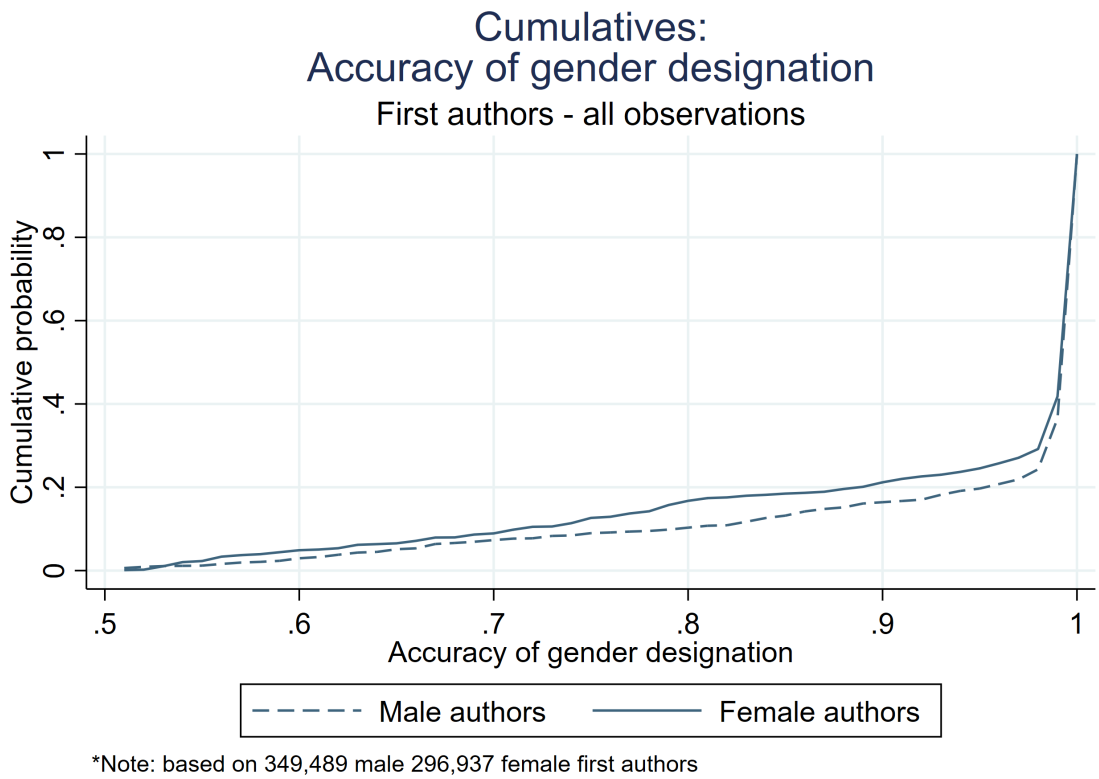

Figure S4: Gender designation accuracy for women first authors for COVID articles and non-COVID (control) articles separately

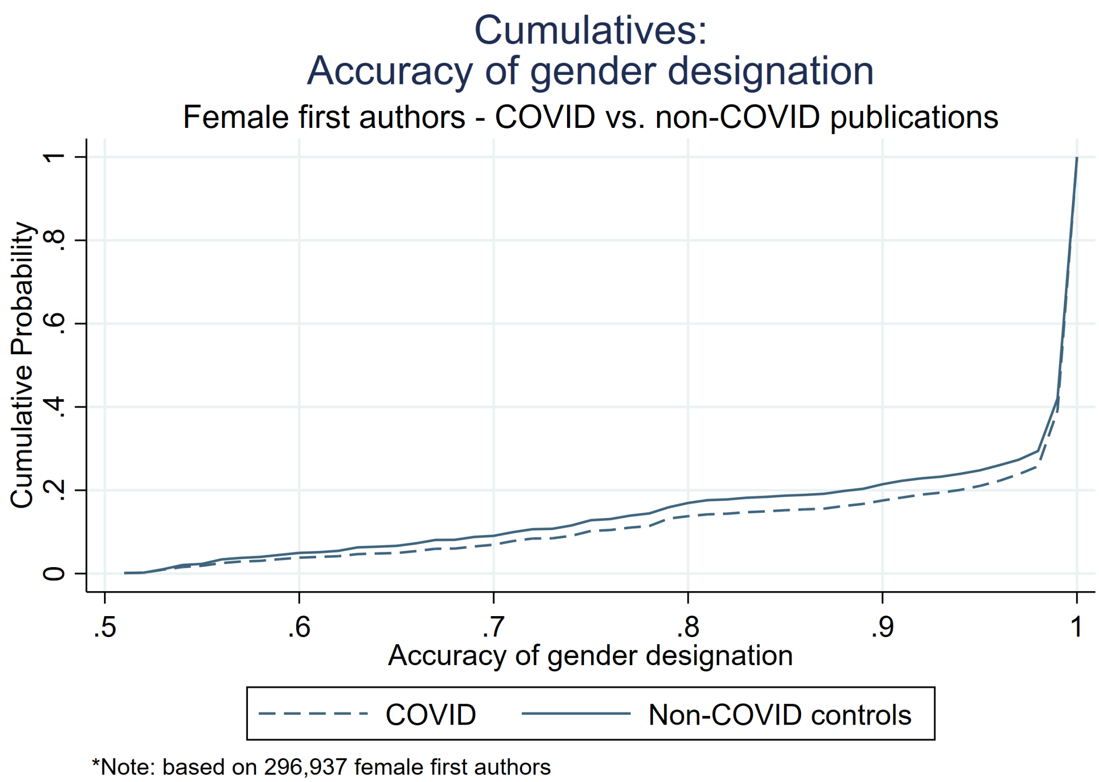

Table S1: Major disciplines for COVID articles and non-COVID (control) articles and concordance statistic

Discpline	Non-COVID			COVID			Total		
	Total	%	Rank	Total	%	Rank	Total	%	Rank
MEDICINE GENERAL & INTERNAL	23,163	4.79%	5	4,229	9.88%	1	27,392	5.21%	4
PUBLIC ENVIRONMENTAL & OCCUPATIONAL HEALTH	17,860	3.70%	8	3,530	8.24%	2	21,390	4.07%	6
SURGERY	26,457	5.48%	3	2,254	5.26%	3	28,711	5.46%	3
IMMUNOLOGY	11,151	2.31%	12	1,925	4.50%	4	13,076	2.49%	11
CARDIAC & CARDIOVASCULAR SYSTEMS	16,196	3.35%	9	1,479	3.45%	5	17,675	3.36%	9
PHARMACOLOGY & PHARMACY	18,408	3.81%	7	1,439	3.36%	6	19,847	3.77%	8
MULTIDISCIPLINARY SCIENCES	38,524	7.97%	1	1,435	3.35%	7	39,959	7.60%	1
PSYCHIATRY	9,890	2.05%	14	1,333	3.11%	8	11,223	2.13%	14
ENVIRONMENTAL SCIENCES	19,607	4.06%	6	1,278	2.98%	9	20,885	3.97%	7
ONCOLOGY	23,207	4.80%	4	1,159	2.71%	10	24,366	4.63%	5
BIOCHEMISTRY & MOLECULAR BIOLOGY	30,224	6.25%	2	1,118	2.61%	11	31,342	5.96%	2
PEDIATRICS	10,280	2.13%	13	1,105	2.58%	12	11,385	2.16%	13
INFECTIOUS DISEASES	3,143	0.65%	41	1,001	2.34%	13	4,144	0.79%	37
CLINICAL NEUROLOGY	9,614	1.99%	15	922	2.15%	14	10,536	2.00%	15
HEALTH CARE SCIENCES & SERVICES	5,209	1.08%	29	921	2.15%	15	6,130	1.17%	27
MEDICINE RESEARCH & EXPERIMENTAL	5,645	1.17%	25	887	2.07%	16	6,532	1.24%	25
NEUROSCIENCES	14,611	3.02%	10	835	1.95%	17	15,446	2.94%	10
DERMATOLOGY	5,452	1.13%	26	777	1.81%	18	6,229	1.18%	26
VIROLOGY	2,198	0.45%	47	771	1.80%	19	2,969	0.56%	43
RESPIRATORY SYSTEM	4,074	0.84%	35	732	1.71%	20	4,806	0.91%	32
GASTROENTEROLOGY & HEPATOLOGY	6,367	1.32%	23	719	1.68%	21	7,086	1.35%	22
RADIOLOGY NUCLEAR MEDICINE & MEDICAL IMAGING	7,092	1.47%	20	715	1.67%	22	7,807	1.48%	19
ENDOCRINOLOGY & METABOLISM	6,468	1.34%	22	628	1.47%	23	7,096	1.35%	21
HEMATOLOGY	4,449	0.92%	31	618	1.44%	24	5,067	0.96%	31
UROLOGY & NEPHROLOGY	7,369	1.52%	18	581	1.36%	25	7,950	1.51%	18
ANESTHESIOLOGY	2,100	0.43%	48	519	1.21%	26	2,619	0.50%	44
MICROBIOLOGY	6,086	1.26%	24	499	1.17%	27	6,585	1.25%	24
NURSING	4,602	0.95%	30	484	1.13%	28	5,086	0.97%	30
EMERGENCY MEDICINE	1,958	0.41%	52	459	1.07%	29	2,417	0.46%	48
RHEUMATOLOGY	3,146	0.65%	40	455	1.06%	30	3,601	0.68%	40
PSYCHOLOGY MULTIDISCIPLINARY	3,271	0.68%	38	413	0.96%	31	3,684	0.70%	39
OPHTHALMOLOGY	5,303	1.10%	27	412	0.96%	32	5,715	1.09%	28
GERIATRICS & GERONTOLOGY	1,818	0.38%	55	392	0.92%	33	2,210	0.42%	52
OBSTETRICS & GYNECOLOGY	4,388	0.91%	32	353	0.82%	34	4,741	0.90%	33
CELL BIOLOGY	8,490	1.76%	16	352	0.82%	35	8,842	1.68%	16
OTORHINOLARYNGOLOGY	2,098	0.43%	49	303	0.71%	36	2,401	0.46%	49
CRITICAL CARE MEDICINE	2,018	0.42%	50	282	0.66%	37	2,300	0.44%	50
DENTISTRY ORAL SURGERY & MEDICINE	3,212	0.66%	39	275	0.64%	38	3,487	0.66%	41
ECONOMICS	495	0.10%	81	274	0.64%	39	769	0.15%	74
BIOTECHNOLOGY & APPLIED MICROBIOLOGY	4,021	0.83%	36	246	0.57%	40	4,267	0.81%	34
NUTRITION & DIETETICS	5,241	1.08%	28	229	0.53%	41	5,470	1.04%	29
PSYCHOLOGY CLINICAL	2,392	0.50%	43	210	0.49%	42	2,602	0.49%	45
ETHICS	674	0.14%	75	195	0.46%	43	869	0.17%	71
MEDICAL LABORATORY TECHNOLOGY	1,217	0.25%	63	185	0.43%	44	1,402	0.27%	61
SPORT SCIENCES	3,065	0.63%	42	182	0.43%	45	3,247	0.62%	42
PATHOLOGY	2,011	0.42%	51	181	0.42%	46	2,192	0.42%	53
GENETICS & HEREDITY	7,184	1.49%	19	178	0.42%	47	7,362	1.40%	20
ORTHOPEDICS	2,345	0.49%	44	168	0.39%	48	2,513	0.48%	46
PERIPHERAL VASCULAR DISEASE	1,700	0.35%	57	133	0.31%	49	1,833	0.35%	56
MATHEMATICS INTERDISCIPLINARY APPLICATIONS	70	0.01%	120	126	0.29%	50	196	0.04%	104

Spearman Rank Correlation - all disciplines	
coefficient (rs)	0.807
N	148
T statistic	16.537
DF	146
p-value	0.000

Spearman Rank Correlation - top 50 disciplines	
coefficient (rs)	0.738
N	50
T statistic	7.588
DF	48
p-value	0.000

Table S2: Major countries for COVID articles and non-COVID (control) articles and concordance statistics

Country (First Author)	Non-COVID			COVID			Total		
	Total	%	Rank	Total	%	Rank	Total	%	Rank
United States	120,478	26.65%	1	11,066	26.12%	1	131,544	26.75%	1
Italy	22,670	5.01%	5	4,309	10.15%	2	26,979	5.49%	4
United Kingdom	27,994	6.19%	3	3,157	8.13%	3	31,151	6.33%	2
India	11,388	2.52%	12	1,778	7.43%	4	13,166	2.68%	12
Spain	15,273	3.38%	9	1,672	4.19%	5	16,945	3.45%	9
China	29,078	6.43%	2	1,377	3.77%	6	30,455	6.19%	3
Canada	16,654	3.68%	7	1,324	3.11%	7	17,978	3.66%	7
France	13,837	3.06%	10	1,313	3.02%	8	15,150	3.08%	10
Germany	24,413	5.40%	4	1,161	2.64%	9	25,574	5.20%	5
Brasil	13,219	2.92%	11	1,160	2.64%	10	14,379	2.92%	11
Australia	16,132	3.57%	8	1,052	2.50%	11	17,184	3.49%	8
Iran	6,757	1.49%	15	878	2.07%	12	7,635	1.55%	14
Turkey	6,134	1.36%	18	851	1.97%	13	6,985	1.42%	16
Japan	18,952	4.19%	6	499	1.30%	14	19,451	3.96%	6
Netherlands	9,702	2.15%	13	425	1.24%	15	10,127	2.06%	13
Switzerland	6,455	1.43%	16	422	1.00%	16	6,877	1.40%	17
Singapore	1,823	0.40%	36	405	0.99%	17	2,228	0.45%	32
Israel	3,813	0.84%	22	378	0.96%	18	4,191	0.85%	22
Saudi Arabia	2,020	0.45%	33	341	0.96%	19	2,361	0.48%	30
Greece	2,605	0.58%	27	323	0.86%	20	2,928	0.60%	26
Belgium	4,132	0.91%	21	314	0.82%	21	4,446	0.90%	21
Pakistan	1,769	0.39%	38	286	0.80%	22	2,055	0.42%	36
Mexico	3,186	0.70%	24	281	0.74%	23	3,467	0.71%	24
Egypt	2,541	0.56%	28	254	0.71%	24	2,795	0.57%	28
Poland	6,318	1.40%	17	253	0.66%	25	6,571	1.34%	18
Hong Kong	1,285	0.28%	41	248	0.63%	26	1,533	0.31%	41
Ireland	2,100	0.46%	31	230	0.59%	27	2,330	0.47%	31
Austria	3,313	0.73%	23	203	0.58%	28	3,516	0.71%	23
South Korea	7,400	1.64%	14	199	0.52%	29	7,599	1.55%	15
Sweden	5,841	1.29%	19	192	0.45%	30	6,033	1.23%	19
Bangladesh	318	0.07%	60	154	0.45%	31	472	0.10%	55
Portugal	3,007	0.67%	25	151	0.36%	32	3,158	0.64%	25
Denmark	4,599	1.02%	20	143	0.35%	33	4,742	0.96%	20
South Africa	1,771	0.39%	37	142	0.35%	34	1,913	0.39%	38
United Arab Emirates	532	0.12%	51	129	0.34%	35	661	0.13%	50
Colombia	869	0.19%	45	120	0.33%	36	989	0.20%	45
Chile	1,520	0.34%	40	114	0.29%	37	1,634	0.33%	40
Taiwan	2,101	0.46%	30	112	0.27%	38	2,213	0.45%	33
Norway	2,812	0.62%	26	106	0.26%	39	2,918	0.59%	27
Malaysia	1,224	0.27%	42	94	0.25%	40	1,318	0.27%	42
Peru	346	0.08%	58	93	0.22%	41	439	0.09%	56
Argentina	1,739	0.38%	39	89	0.20%	42	1,828	0.37%	39
Romania	1,142	0.25%	43	84	0.20%	43	1,226	0.25%	43
Russia	2,083	0.46%	32	83	0.20%	44	2,166	0.44%	34
Lebanon	679	0.15%	49	82	0.20%	45	761	0.15%	49
New Zealand	2,011	0.44%	34	80	0.19%	46	2,091	0.43%	35
Nigeria	501	0.11%	52	75	0.19%	47	576	0.12%	52
Indonesia	289	0.06%	64	73	0.17%	48	362	0.07%	60
Jordan	345	0.08%	59	68	0.17%	49	413	0.08%	58
Morocco	304	0.07%	63	66	0.17%	50	370	0.08%	59

Spearman Rank Correlation - all countries	
coefficient (rs)	0.93
N	167
T statistic	32.32
DF	165
p-value	0.000

Spearman Rank Correlation - top 50 countries	
coefficient (rs)	0.85
N	50
T statistic	10.97
DF	48
p-value	0.000

Table S3: Hierarchical linear probability model for the likelihood of women first authorship for COVID articles versus non-COVID (control) articles

Dependent variable: First Author Female	(1)	(2)	(3)	(4)	(5)
COVID	**-0.074*** (0.00)**	**-0.074*** (0.00)**	**-0.075*** (0.00)**	**-0.086*** (0.00)**	**-0.089*** (0.00)**
number of authors		0.002*** (0.00)	0.002*** (0.00)	0.002*** (0.00)	0.002*** (0.00)
publication month fixed effects (12)			Included	Included	Included
discipline fixed effects (148)				Included	Included
country fixed effects (167)					Included
constant	0.451*** (0.00)	0.442*** (0.00)	0.454*** (0.00)	0.269*** (0.02)	0.036 (0.10)
R-squared	0.002	0.002	0.002	0.043	0.060
Adjusted R-squared	0.002	0.002	0.002	0.043	0.059
Observations	526,130	526,130	526,130	526,130	491,912

Note: standard errors in brackets, * p < 0.05, ** p< 0.01, *** p < 0.001

Table S4: Hierarchical linear probability model for the likelihood of women last authorship for COVID articles versus non-COVID (control) articles

Dependent variable: Last Author Female	(1)	(2)	(3)	(4)	(5)
COVID	**-0.014***** **(0.00)**	**-0.014***** **(0.00)**	**-0.015***** **(0.00)**	**-0.033***** **(0.00)**	**-0.037***** **(0.00)**
number of authors		-0.002*** (0.00)	-0.002*** (0.00)	-0.001*** (0.00)	-0.001*** (0.00)
publication month fixed effects (12)			Included	Included	Included
discipline fixed effects (148)				Included	Included
country fixed effects (167)					Included
constant	0.319*** (0.00)	0.332*** (0.00)	0.342*** (0.00)	0.232*** (0.02)	0.081 (0.09)
R-squared	0.000	0.001	0.001	0.043	0.060
Adjusted R-squared	0.000	0.001	0.001	0.043	0.059
Observations	539,103	539,103	539,103	539,103	504,148

Note: standard errors in brackets, * p < 0.05, ** p< 0.01, *** p < 0.001

Table S5: Hierarchical logit regression for the likelihood of women first authorship for COVID articles versus non-COVID (control) articles

Dependent variable: First Author Female	(1)	(2)	(3)	(4)	(5)
COVID	**0.738*** (0.01)**	**0.737*** (0.01)**	**0.733*** (0.01)**	**0.690*** (0.01)**	**0.678*** (0.01)**
number of authors		1.006*** (0.00)	1.006*** (0.00)	1.009*** (0.00)	1.009*** (0.00)
publication month fixed effects (12)			Included	Included	Included
discipline fixed effects (148)				Included	Included
country fixed effects (167)					Included
constant	0.821*** (0.00)	0.790*** (0.00)	0.829*** (0.01)	0.365*** (0.03)	0.117*** (0.06)
observations	526,130	526,130	526,130	526,112	491,837

Note: Coefficients reported as odds ratios, standard errors in brackets, * p < 0.05, ** p < 0.01, *** p < 0.001

Table S6: Robustness checks

Dependent variable: First Author Female	Accuracy of gender designation > 50% (1)	Accuracy of gender designation = 100% (2)	Exluding collective authorships (3)	Full sample (4)
COVID	**-0.082***** **(0.00)**	**-0.090***** **(0.00)**	**-0.089***** **(0.00)**	**-0.079***** **(0.00)**
number of authors	0.002*** (0.00)	0.002*** (0.00)	0.003*** (0.00)	0.002*** (0.00)
publication month fixed effects (12)	Included	Included	Included	Included
discipline fixed effects (148)	Included	Included	Included	Included
country fixed effects (167)	Included	Included	Included	
journal fixed effects (5,101)				Included
constant	0.095 (0.10)	0.040 (0.10)	0.031 (0.10)	0.215* (0.10)
R-squared	0.049	0.067	0.060	0.091
Adjusted R-squared	0.048	0.066	0.059	0.082
Observations	607,598	443,711	483,308	507,653

Note: standard errors in brackets, * p < 0.05, ** p< 0.01, *** p < 0.001

Table S7: Descriptive statistics of articles included in the analysis versus articles not included in the analysis

	Included in analysis		Excluded from analysis		t-test	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Difference**	**t-statistic**
First Author Female	0.38	0.48	0.38	0.49	0.00	0.45
Last Author Female	0.32	0.47	0.33	0.47	0.01	2.27
Publication Month	7.35	3.15	7.51	3.25	0.15	5.35
Number of Authors	6.42	8.56	5.75	6.25	-0.68	-10.76
North America	0.29	0.45	0.24	0.43	-0.05	-12.70
Europe	0.35	0.48	0.28	0.45	0.07	-16.15
Asia	0.19	0.39	0.23	0.42	0.04	9.69
Latin America	0.05	0.21	0.04	0.20	0.00	-1.44
Oceania	0.03	0.16	0.02	0.14	-0.01	-4.22
Africa	0.02	0.14	0.03	0.17	0.01	6.20
Observations	**42,898**		**17,445**		**60,343**	

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No.	Recommendation	Page No.	Relevant text from manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1	Longitudinal analyses
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1	Widening gender gap in productivity with COVID
Introduction				
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2	COVID impact on productivity
Objectives	3	State specific objectives, including any prespecified hypotheses	2	Gendered productivity drain
Methods				
Study design	4	Present key elements of study design early in the paper	3	COVID as exogenous shock
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	2	Longitudinal case-control design
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	3	COVID publications (Feb 20 to Jan 21) as cases versus publications in same journals a year earlier (controls)
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case		
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7	Absolute percentage point gender difference in authorship
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6	PubMed, Clarivate JCR, Genderize.io
Bias	9	Describe any efforts to address potential sources of bias	6/7	Case-control design, sensitivity analyses
Study size	10	Explain how the study size was arrived at	7	Population of PubMed articles

Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why		Statistical Software Stata
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6/7	Descriptive and parametric analyses
		(*b*) Describe any methods used to examine subgroups and interactions	6	Discipline and Country
		(*c*) Explain how missing data were addressed	Supplement	Exclusion plus sensitivities
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy		N/A
		(*e*) Describe any sensitivity analyses	7	Full sample and subgroup testing
Results				
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7	42,898 COVID articles (cases) 483,232 control articles
		(b) Give reasons for non-participation at each stage		N/A
		(c) Consider use of a flow diagram	Supplement	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6	COVID pandemic as exposure, gender of first and last authors
		(b) Indicate number of participants with missing data for each variable of interest		N/A
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)		N/A
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time		
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	9	Gender gap 9.8 percentage points (54.9% men vs 45.1% women)
		Cross-sectional study—Report numbers of outcome events or summary measures		
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	9	See 15.
		(*b*) Report category boundaries when continuous variables were categorized		N/A

Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	9-12	Discipline and country analyses
Discussion				
Key results	18	Summarise key results with reference to study objectives	13	Acute increase in gender gap early 2020 with longitudinal reversion
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	16	Focus on COVID research, probabilistic gender designation
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13/14	Confirming acute increase in gender gap, new longitudinal findings
Generalisability	21	Discuss the generalisability (external validity) of the study results	16	Applies to COVID research, likely with external validity
Other information				
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	20	Office of the Director, National Institutes of Health (1DP5OD017897, Dr. Jena). The funding sources had no role in study

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.