Supplementary materials of

Female authorship of covid-19 research in manuscripts submitted to 11 biomedical journals: cross sectional study

Angèle GAYET-AGERON*, Khaoula BEN MESSAOUD*, Mark RICHARDS, Sara SCHROTER

Table of contents

Section 3: Sensitivity analyses

Table F. Sensitivity analyses of multivariate models for first, last and corresponding authors' gender with accuracy for gender determination above (≥) 95%. Models adjusted for the level of missing gender determination at country level.

Table H. Sensitivity analyses of multivariable models for first, last and corresponding authors' gender from multiple imputation models.

Figure A. Selection of manuscripts submitted to 11 BMJ Publishing Group journals between January 1, 2018 and May 31, 2021.
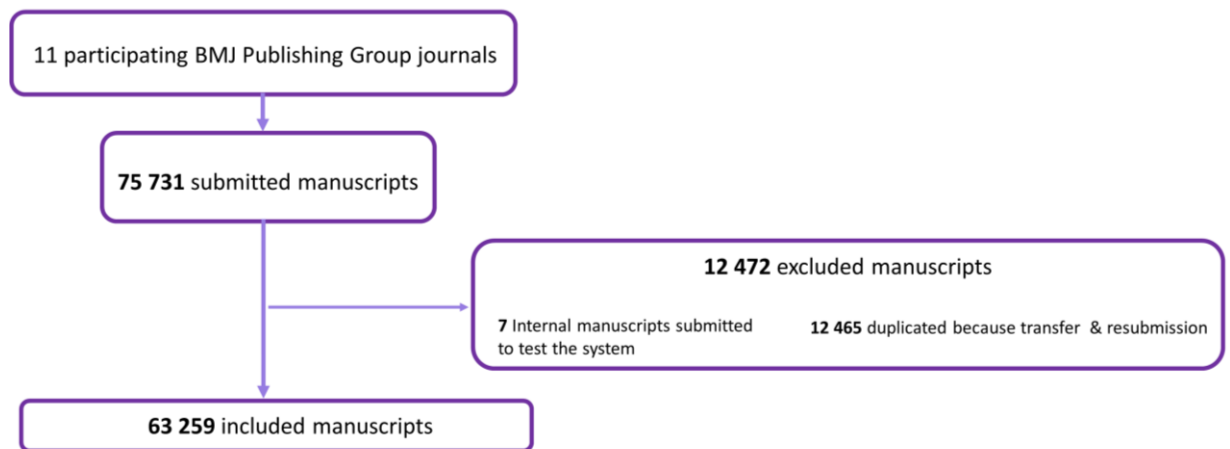
Figure B. Multi-step gender determination among authors[a] of manuscripts submitted to 11 BMJ Publishing Group journals between January 1, 2018 and May 31, 2021.

N= 453,640 Authors

Authors with complete names
N=442,766

Authors with blank names or initials
N= 10,874

Step 1: Gender API (first name)
436,937 determined (≥50% accuracy)

Step 2: Gender API (middle name)
249 determined (≥50% accuracy)

Step 3: genderize.io (first & middle names)
5,575 determined (≥50% accuracy)

Step 4: Author salutation
2,825 determined
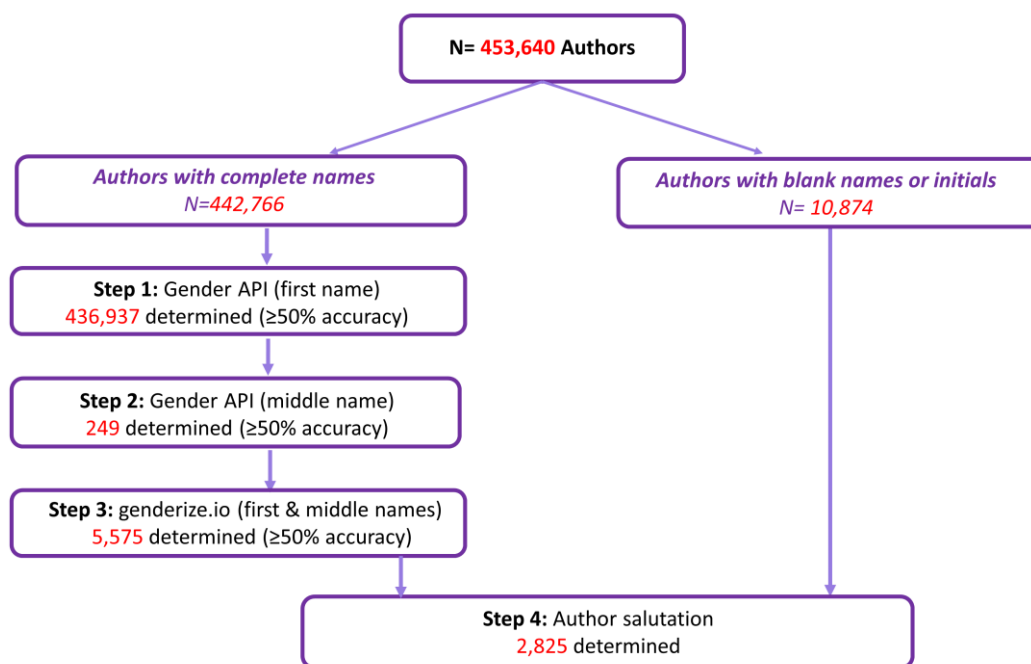
[a] No repeated authors among all co-authors of all manuscripts

Figure C. Distribution of the number of women co-authors in manuscript according to the last authors' region of affiliation[a].
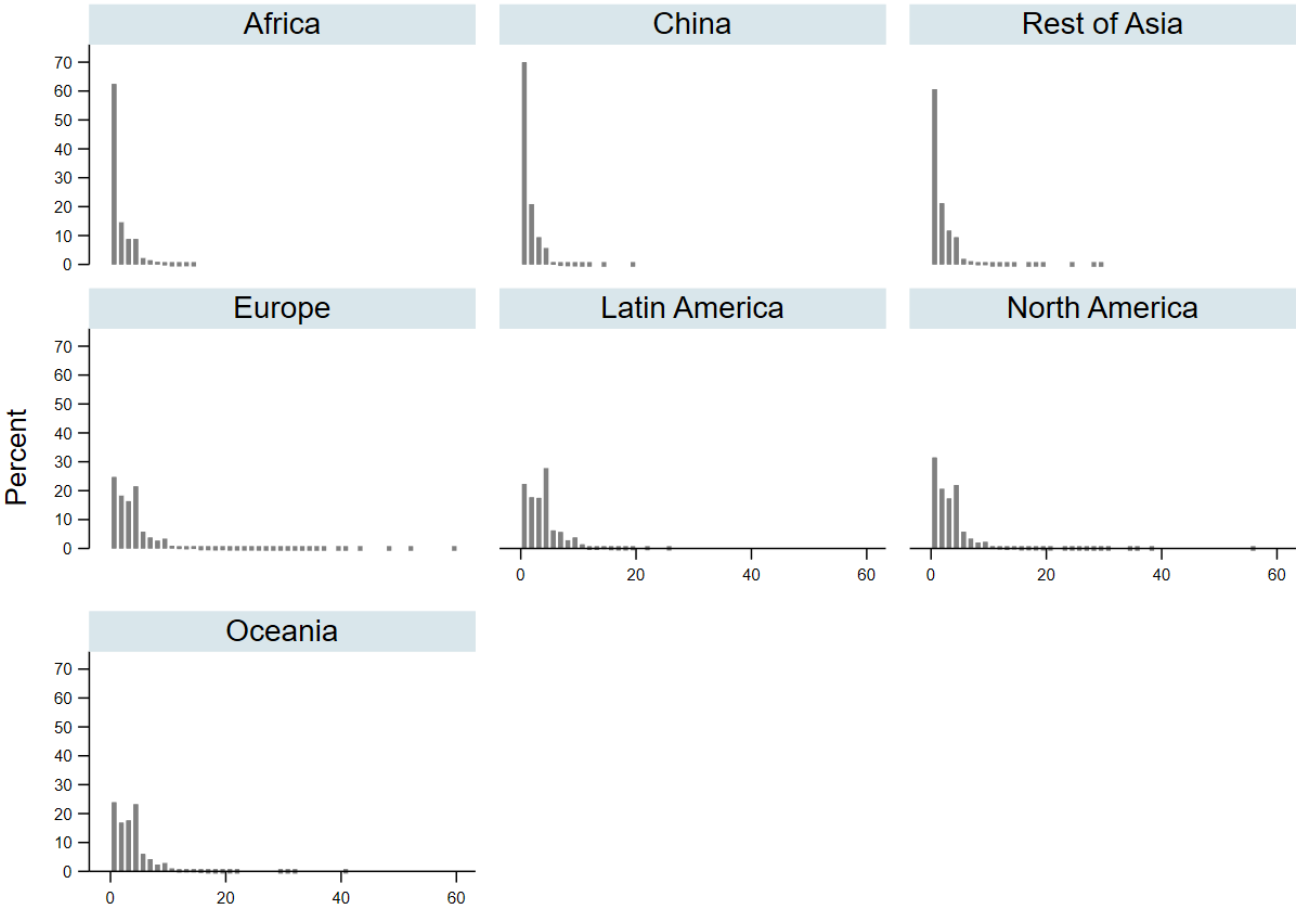
[a] We reported numbers only in manuscripts with at least 80% of co-authors who had their gender determined.

Table A: Author's gender distribution (first, last, corresponding), percentage and number of women authors, and manuscript characteristics [a,b] by time-periods and "COVID-19" topic among included manuscripts (n=63 259).

	Pre-pandemic period		Pandemic period (2020-2021)		p-value
	2018	**2019**	**"Non-COVID-19" manuscripts**	**"COVID-19" manuscripts**	
Total submission numbers	16 779	17 443	24 590	4 447	-
Total number of authors	139 581	148 182	217 597	45 981	-
Total number of authors with determined gender	114 815 (82.3%)	121 133 (81.7%)	172 282 (79.2%)	36 822 (80.1%)	-
Authors' gender[a]					
First author gender, (n missing, %)	(6 537, 19.1%)		(5 484, 22.3%)	(912, 20.5%)	
Female	12 724 (46.0)		8 583 (44.9)	1 313 (37.1)	<0.001[c]
Male	14 961 (54.0)		10 523 (55.1)	2 222 (62.9)	
Last author gender, (n missing, %)	(5 804, 17.0%)		(4 973, 20.2%)	(873, 19.6%)	
Female	8 923 (31.4)		6 118 (31.2)	996 (27.9)	<0.001[c]
Male	19 495 (68.6)		13 499 (68.8)	2 578 (72.1)	
Corresponding author gender, (n missing, %)	(5 979, 17.5%)		(5 088, 20.7%)	(845, 19.0%)	
Female	10 981 (38.9)		7 273 (37.3)	1 119 (31.1)	<0.001[c]
Male	17 262 (61.1)		12 229 (62.7)	2 483 (68.9)	
% of women authors[b]	38.5±27.7 (36.4:40.5)		38.4±28.2 (35.7:40.5)	36.5±26.4 (33.3:33.3)	-
Number of authors with determined gender[b]	6.9±5.5 (6:5)		7.1±5.8 (6:5)	8.4±7.7 (6:6)	-
Number of women authors[b]	2.7± 2.8(2:3)		2.7±2.9 (2:3)	3.2±3.7 (2:3)	-
Number of co-authors on the manuscript byline[b]	8.4±6.0 (7:5)		8.8±6.3 (7:6)	10.3±8.5 (8:8)	<0.001[d]
Editorial Decision					
Accepted	8 884 (26.0%)		4 855 (19.7 %)	568 (12.8 %)	
Rejected	24 964 (72.9%)		15 468 (62.9%)	3 134 (68.5 %)	
Under revision	354 (1.0%)		345 (1.4%)	38 (0.9%)	
No decision yet	20 (0.1%)		3 922 (16.0%)	707 (15.9%)	
Type of journal					
Generalist	19 124 (53.9)		13 951 (56.7)	3 324 (74.8)	<0.001[c]
Specialised	15 098 (44.1)		10 639 (43.3)	1 123 (25.2)	

[a]Categorical variables: n (%), [b]Continuous and discrete variables: mean± SD (median: IQR), [c]p-value from mixed-effect logistic regression models in univariate; [d]p-value from mixed-effects Poisson regression model in univariate.

Table B: Association between study groups and chance of having a woman first author by journal specialty.

	N	Pre-pandemic	Female first author (OR[a] (95% CI)[b], p value)	
			Pandemic	
			"COVID-19" manuscripts	"Non COVID-19" manuscripts
Journal				
Generalist	36 399	1.00	0.74 (0.68-0.81), p<0.001	1.05 (1.00-1.11), p=0.05
Speciality 1	1 355	1.00	0.78 (0.48-1.27), p=0.32	1.13 (0.87-1.47), p=0.37
Speciality 2	6 534	1.00	0.58 (0.40-0.84), p=0.004	1.01 (0.90-1.15), p=0.82
Speciality 3	4 824	1.00	0.51 (0.33-0.78), p=0.002	0.91 (0.79-1.04), p=0.18
Speciality 4	5 524	1.00	0.79 (0.51-1.23), p=0.30	0.96 (0.84-1.08), p=0.48
Speciality 5	1 832	1.00	0.43 (0.27-0.69), p<0.001	0.90 (0.72-1.13), p=0.40
Speciality 6	1 576	1.00	0.65 (0.45-0.95), p=0.03	0.86 (0.67-1.12), p=0.27
Speciality 7	840	1.00	0.76 (0.27-2.12), p=0.60	0.73 (0.52-1.01), p=0.06
Speciality 8	4 375	1.00	0.55 (0.39-0.77), p<0.001	0.95 (0.82-1.10), p=0.48

[a] From mixed-effects logistic regression model with the country of first author as random factor and an interaction term between the journal specialty (generalist, specialties 1 to 8 to preserve journal confidentiality) and the groups. [b]This analysis was conducted in response to BMJ peer-review, so the CIs were not adjusted for multiple comparisons as it was done for other analyses

Table C. Distribution of geographical origin of first, last, and corresponding authors among groups (n=63 259).

| | Pre-pandemic 2018-2019 | Pandemic 2020-2021 | | | | | | | | p-value |
| | | Non-COVID-19-related manuscripts | | | | COVID-19-related manuscripts | | | | |
		Jan-May 20	Jun-Sep 20	Oct 20-Jan 21	Feb-May 21	Jan-May 20	Jun-Sep 20	Oct 20-Jan 21	Feb-May 21	
Number of submissions	34 222	7 055	6 042	5 803	5 690	1 102	1 334	987	1024	-
Geographical original										
First author										<0.001[a]
Africa	855 (2.5)	238 (3.4)	252 (4.2)	253 (4.4)	203 (3.6)	16 (1.5)	36 (2.7)	36 (3.7)	29 (2.8)	
North America	5 136 (15.0)	982 (13.9)	897 (14.9)	787 (13.6)	751 (13.2)	94 (8.5)	143 (10.7)	129 (13.1)	155 (15.1)	
Latin America	670 (2.0)	146 (2.1)	108 (1.8)	101 (1.7)	96 (1.7)	18 (1.6)	35 (2.6)	24 (2.4)	42 (4.1)	
Europe	13 581 (39.7)	2 569 (36.4)	2 059 (34.1)	1 961 (33.8)	1 976 (34.7)	326 (29.6)	622 (46.6)	434 (44.0)	440 (43.0)	
Oceania	1 897 (5.5)	355 (5.0)	350 (5.8)	285 (4.9)	243 (4.3)	19 (1.7)	34 (2.6)	27 (2.7)	27 (2.6)	
China	6 479 (18.9)	1 604 (22.7)	1 364 (22.6)	1 465 (25.3)	1 483 (26.1)	474 (43.0)	263 (19.7)	160 (16.2)	135 (13.2)	
Rest of Asia	5 604 (16.4)	1 161 (16.5)	1 012 (16.8)	951 (16.4)	938 (16.5)	155 (14.1)	201 (15.1)	177 (17.9)	196 (19.1)	
Last author										<0.001[a]
Africa	800 (2.3)	207 (2.9)	235 (3.9)	229 (4.0)	187 (3.3)	17 (1.5)	36 (2.7)	32 (3.2)	29 (2.8)	
North America	5 510 (16.1)	1 041 (14.8)	956 (15.8)	860 (14.8)	790 (13.9)	100 (9.1)	161 (12.1)	139 (14.1)	157 (15.3)	
Latin America	591 (1.7)	140 (2.0)	105 (1.7)	97 (1.7)	95 (1.7)	17 (1.5)	27 (2.0)	23 (2.3)	41 (4.0)	
Europe	13 780 (40.3)	2 630 (37.3)	2 093 (34.6)	1 987 (34.2)	2 020 (35.5)	339 (30.8)	631 (47.3)	453 (45.9)	449 (43.8)	
Oceania	1 942 (5.7)	378 (5.4)	369 (6.1)	283 (4.9)	263(4.6)	18 (1.6)	32 (2.4)	28 (2.8)	28 (2.7)	
China	6 192 (18.1)	1 543 (21.9)	1 306 (21.6)	1 433 (24.7)	1 442 (25.3)	451 (40.9)	251 (18.8)	153 (15.5)	132 (12.9)	
Rest of Asia	5 407 (15.8)	1 116 (15.8)	978 (16.2)	914 (15.8)	893 (15.7)	160 (14.5)	196 (14.7)	159 (16.1)	188 (18.4)	
Corresponding author										<0.001[a]
Africa	830 (2.4)	233 (3.3)	241 (4.0)	248 (4.3)	199 (3.5)	15 (1.4)	36 (2.7)	31 (3.1)	29 (2.8)	
North America	5 318 (15.5)	1 018 (14.4)	929 (15.4)	828 (14.3)	778 (13.7)	98 (8.9)	155 (11.6)	139 (14.1)	157 (15.3)	
Latin America	641 (1.9)	139 (2.0)	105 (1.7)	98 (1.7)	98 (1.7)	15 (1.4)	32 (2.4)	23 (2.3)	41 (4.0)	
Europe	13 651 (39.9)	2 592 (36.7)	2 074 (34.3)	1 973 (34.0)	1 986 (34.9)	333 (30.2)	623 (46.7)	438 (44.4)	448 (43.8)	
Oceania	1 904 (5.6)	362 (5.1)	359 (5.9)	281 (4.8)	245 (4.3)	21 (1.9)	36 (2.7)	27 (2.7)	27 (2.6)	
China	6 334 (18.5)	1 573 (22.3)	1 324 (21.9)	1 444 (24.9)	1 458 (25.6)	460 (41.7)	250 (18.7)	155 (15.7)	128 (12.5)	
Rest of Asia	5 544 (16.2)	1 138 (16.1)	1 010 (16.7)	931 (16.0)	926 (16.3)	160 (14.5)	202 (15.1)	174 (17.6)	194 (18.9)	

[a]p-value from Chi-2 test

Table D. Region of affiliation for first, last, and corresponding authors according to gender determination (male, female or missing) (n=63 259).

Affiliation region	First author[a]			Last author[a]			Corresponding author[a]		
	Men	Women	Missing gender	Men	Women	Missing gender	Men	Women	Missing gender
Africa	1 195 (62.3)	485 (25.3)	238 (12.4)	1 140 (64.3)	395 (22.3)	237 (13.4)	1 200 (64.4)	435 (23.4)	227 (12.2)
China	3 790 (28.2)	1 949 (14.5)	7 688 (57.3)	4 697 (36.4)	1 323 (10.3)	6 883 (53.3)	4 816 (36.7)	1 359 (10.3)	6 951 (53.0)
Europe	11 005 (45.9)	11 457 (47.8)	1 506 (6.3)	15 014 (61.6)	7 902 (32.4)	1 466 (6.0)	12 539 (52.0)	10 102 (41.9)	1 477 (6.1)
Latin America	542 (43.7)	646 (52.1)	52 (4.2)	605 (53.3)	474 (41.7)	57 (5.0)	606 (50.8)	536 (45.0)	50 (4.2)
North America	4 145 (45.7)	3 755 (41.4)	1 174 (12.9)	5 649 (58.2)	2 958 (30.4)	1 107 (11.4)	4 822 (51.2)	3 387 (35.9)	1 211 (12.9)
Oceania	1 193 (36.9)	1 644 (50.8)	400 (12.4)	1 672 (50.0)	1 340 (40.1)	329 (9.9)	1 323 (40.6)	1 562 (47.9)	377 (11.6)
Rest of Asia	5 836 (56.1)	2 684 (25.8)	1 875 (18.0)	6 795 (67.9)	1 645 (16.4)	1 571 (15.7)	6 668 (64.9)	1 992 (19.4)	1 619 (15.7)

[a]Categorical variables: n (%)

Table E. Multivariable models for first (model 1), last (model 2), and corresponding (model 3) authors' gender[a] among "pre-pandemic", "COVID-19", and "pandemic non-COVID-19" manuscripts. Models adjusted for the level of missing gender determination at country level.

Independent variables	First author (model 1)[a] n=43 319			Last author (model 2)[a] n=43 297			Corresponding author (model 3)[a] n=50 896		
	OR[b]	(98.75% CI)[c]	p-value[d]	OR[b]	(98.75% CI)[c]	p-value[d]	OR[b]	(98.75% CI)[c]	p-value[d]
Groups (ref=Pre-Pandemic)			<0.001			0.08			<0.001
Pandemic "COVID-19"	0.67	(0.60-0.75)	<0.001	0.93	(0.82-1.04)	0.10	0.72	(0.65-0.80)	<0.001
Pandemic "non-COVID-19"	0.99	(0.94-1.05)	0.87	1.03	(0.97-1.09)	0.23	1.01	(0.96-1.06)	0.62
First author' gender (ref=Male)	-	-	-	1.84	(1.74-1.94)	<0.001	-	-	-
Female									
Last author' gender (ref=Male)	1.84	(1.74-1.94)	<0.001	-	-	-	-	-	-
Female									
Author Type (ref= Last)	-	-	-	-	-	-			<0.001
First							1.71	(1.61-1.81)	<0.001
Other							1.26	(1.16-1.38)	<0.001
Number of authors on the article byline	0.99	(0.99-1.00)	0.27	0.98	(0.98-0.99)	<0.001	0.99	(0.98-0.99)	<0.001
Type of journal (ref=Generalist)									
Specialist	1.02	(0.67-1.56)	0.91	0.99	(0.53-1.86)	0.98	0.96	(0.54-1.71)	0.87

[a]Manuscripts with complete data for variables used in models (model 1, model 2 and model 3) per groups; all three models were adjusted for the level of missing gender determination at country level (not reported); [b]OR: adjusted odds ratio from mixed-effects logistic regression models; [c]98.75% CI: 98.75% confidence interval. [d]p-values should be interpreted below the 0.0125 cut-off with only p<0.0125 yielding a CI not crossing the null value.

Table F. Sensitivity analyses of multivariable models for first, last, and corresponding authors' gender with accuracy for gender determination above (≥) 95%. Models adjusted for the level of missing gender determination at country level.

Independent variables	First author (model 1)[a] n=35 726			Last author (model 2)[a] n=35 692			Corresponding author (model 3)a n=43 996		
	OR[b]	(98.75% CI)[c]	p-value[d]	OR[b]	(98.75% CI)[c]	p-value[d]	OR[b]	(98.75% CI)[c]	p-value[d]
Groups (ref=Pre-Pandemic)			<0.001			0.05			<0.001
Jan-May 20 "COVID-19"	0.54	(0.41-0.71)	<0.001	0.71	(0.52-0.96)	0.005	0.56	(0.44-0.73)	<0.001
Jan-May 20 "non-COVID-19"	0.97	(0.88-1.06)	0.37	1.00	(0.91-1.11)	0.94	0.99	(0.90-1.08)	0.70
Jun-Sep 20 "COVID-19"	0.63	(0.52-0.77)	<0.001	0.82	(0.65-1.02)	0.02	0.69	(0.57-0.83)	<0.001
Jun-Sep 20 "non-COVID-19"	1.03	(0.93-1.14)	0.52	0.98	(0.88-1.09)	0.67	1.00	(0.91-1.10)	0.95
Oct 20-Jan 21 "COVID-19"	0.69	(0.55-0.87)	<0.001	0.95	(0.74-1.20)	0.57	0.72	(0.58-0.90)	<0.001
Oct 20-Jan 21 "non-COVID-19"	1.02	(0.92-1.13)	0.60	1.02	(0.91-1.14)	0.65	1.05	(0.96-1.16)	0.19
Feb-May 21 "COVID-19"	0.80	(0.64-1.01)	0.02	1.10	(0.87-1.39)	0.31	0.79	(0.64-0.98)	0.005
Feb-May 21 "non-COVID-19"	1.01	(0.91-1.13)	0.73	1.04	(0.93-1.17)	0.35	1.00	(0.91-1.10)	0.99
First author' gender (ref=Male)	-	-	-				-	-	-
Female				1.94	(1.83-2.07)	<0.001			
Last author' gender (ref=Male)				-	-	-	-	-	-
Female	1.93	(1.82-2.06)	<0.001						
Author Type (ref= Last)	-	-	-	-	-	-			<0.001
First							1.73	(1.63-1.84)	<0.001
Other							1.31	(1.19-1.45)	<0.001
Number of authors on byline	0.99	(0.99-1.00)	0.22	0.98	(0.98-0.99)	<0.001	0.99	(0.98-0.99)	<0.001
Type of journal (ref=Generalist)									
Specialist	0.99	(0.64-1.57)	0.99	1.03	(0.57-1.89)	0.89	0.95	(0.54-1.68)	0.83

[a] Manuscripts with complete data for variables used in models (model 1, model 2 and model 3) per groups; all three models were adjusted for the level of missing gender determination at country level (not reported); [b]OR: adjusted odds ratio from mixed-effects logistic regression models; [c]98.75% CI: 98.75% confidence interval. [d]p-values should be interpreted below the 0.0125 cut-off with only p<0.0125 yielding a CI not crossing the null value.

Table G. Multivariable models for the number of women co-authors across groups among manuscripts[a] with last authors' affiliation in Europe, Latin America, North America, or Oceania (model 4A), and in Africa, China, or rest of Asia (model 4B), among "pre-pandemic", "COVID-19", and pandemic "non-COVID-19" manuscripts. Models adjusted for the level of missing gender determination at country level.

Independent variables	Percentage of women co-authors (model 4A)[b] n=32 196			Percentage of women co-authors (model 4B)[b] n=7 972		
	IRR	(98.75% CI)[c]	p-value[d]	IRR	(98.75% CI)[c]	p-value[d]
Groups (ref=Pre-Pandemic)			<0.001			0.002
Pandemic "COVID-19"	0.95	(0.93-0.98)		0.97	(0.90-1.05)	0.37
Pandemic "non-COVID-19"	1.01	(1.00-1.03)	<0.001	0.94	(0.90-0.98)	<0.001
			0.02			
Type of journal (ref=Generalist)						
Specialist	0.94	(0.69-1.29)	0.63	1.04	(0.68-1.60)	0.81

[a] Manuscripts with complete data for variables used in models 4A and 4B. Models adjusted for the level of missing gender determination at country level (not reported); [b]IRR: incidence rate ratio from mixed-effects Poisson regression model (nested random factors on the intercept, last author country nested within journal); [c]98.75% CI: 98.75% confidence interval. [d]p-values should be interpreted below the 0.0125 cut-off with only p<0.0125 yielding a CI not crossing the null value.

Table H. Sensitivity analyses of multivariable models for first, last and corresponding authors' gender from multiple imputation models.

Independent variables	First author (model 1)[a] n=62 751 OR	(98.75% CI)[b]	p-value[c]	Last author (model 2)[a] n=62 326 OR	(98.75% CI)[b]	p-value[c]	Corresponding author (model 3)[a] n=62 766 OR	(98.75% CI)[b]	p-value[c]
Groups (ref=Pre-Pandemic)			<0.001			<0.001			<0.001
Jan-May 20 "COVID-19"	0.59	(0.50-0.69)	<0.001	0.70	(0.59-0.84)	<0.001	0.55	(0.46-0.65)	<0.001
Jan-May 20 "non-COVID-19"	0.99	(0.93-1.05)	0.64	1.04	(0.98-1.11)	0.22	1.01	(0.95-1.07)	0.83
Jun-Sep 20 "COVID-19"	0.67	(0.59-0.76)	<0.001	0.81	(0.70-0.93)	0.003	0.66	(0.57-0.75)	<0.001
Jun-Sep 20 "non-COVID-19"	1.06	(0.99-1.13)	0.09	1.03	(0.96-1.11)	0.43	1.04	(0.97-1.12)	0.22
Oct 20-Jan 21 "COVID-19"	0.71	(0.60-0.83)	<0.001	0.99	(0.85-1.17)	0.95	0.76	(0.66-0.89)	<0.001
Oct 20-Jan 21 "non-COVID-19"	1.05	(0.98-1.12)	0.19	1.11	(1.02-1.19)	0.006	1.08	(1.00-1.16)	0.03
Feb-May 21 "COVID-19"	0.88	(0.75-1.02)	0.10	1.09	(0.93-1.28)	0.28	0.82	(0.70-0.95)	0.01
Feb-May 21 "non-COVID-19"	1.05	(0.98-1.12)	0.21	1.08	(0.99-1.16)	0.05	1.05	(0.97-1.13)	0.20
First author' gender (ref=Male)									
Female	-	-	-	1.90	(1.82-1.99)	<0.001	-	-	-
Last author' gender (ref=Male)									
Female	2.17	(2.08-2.27)	<0.001	-	-	-	-	-	-
Author Type (ref= Last)									<0.001
First	-	-	-	-	-	-	1.57	(1.51-1.64)	<0.001
Other	-	-	-	-	-	-	1.20	(1.12-1.28)	<0.001
Number of authors on byline	0.99	(0.99-0.99)	0.02	0.98	(0.98-0.99)	<0.001	0.99	(0.99-0.99)	<0.001
Type of journal (ref=Generalist)									
Specialist	0.99	(0.65-1.51)	0.96	0.91	(0.45-1.82)	0.78	0.88	(0.45-1.72)	0.71

[a] OR: adjusted odds ratio from mixed-effects logistic regression models; [b] 98.75% CI: 98.75% confidence interval. [c] p-values should be interpreted below the 0.0125 cut-off with only p<0.0125 yielding a CI not crossing the null value.