medRxiv preprint doi: https://doi.org/10.1101/2022.12.21.22283801; this version posted December 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

1 2 3	The Gender Gap in Leading Medical Journals - a Computational Audit
4	
5 6 7	Oscar Brück <sup>1</sup>
7 8 9	1 Hematoscope Lab, Helsinki University Hospital, Helsinki, Finland
10	
11	
12	
13	
14 15	ABSTRACT
16	Background:
17	Publication track record can impact careers of researchers. Therefore, monitoring
18	gender representation in medical research is required to achieve equity in academia.
19	Methods:
20	We gathered bibliometric data on original research articles published between 2010
21	and 2019 in The New England Journal of Medicine, Nature Medicine, Journal of the
22	American Medical Association, The BINJ, and The Lancet using the Web of Science
25 24	dender count and institute affiliation and research keywords
25	Findings:
26	We analyzed 10,558 articles and found that women published and were cited less than
27	men. There were fewer women as senior (24.8%) than leading authors (34.5%,
28	p<0.001). The proportion of female authors varied by country with 9.1% last authors
29	from Austria, 0.9% from Japan, and 0.0% from South Korea. The gender gap
30 21	decreased longitudinally and faster for last (-24.0 articles/year, p<0.001) than first
32	in China and Israel. Author count was associated with higher citation count $(R = 0.46)$
33	p < 0.001) as well as with male first (n=11 vs. n=10, $p < 0.001$ ) and last authors (n=11)
34	vs. $n=10$ , $p<0.001$ ). We also discovered that usage of research keywords varied by
35	gender, and it partly accounted for the difference in citation counts by gender.
36	Interpretation:
37	Gender representation has increased both at the leading and senior author levels
38	although with country-specific variability. The study frame can be easily applied to any
39 40	journal and time period to monitor changes in gender representation in science.
40 41	
42	
43	
44	
45	ADDRESS FOR CORRESPONDENCE
46	Dr. O. Brück, Hematoscope Lab, Helsinki University Hospital, Comprehensive Cancer
47 19	Center, Haartmaninkatu &, P.O. Box 700, Heisinki FIN-00290, Finland. E-mail:
4ð 49	<u>USCALDIUCK@HUS.II.</u> UKCID <u>https://orciu.org/0000-0002-7842-9419</u> .
50	LENGTH 2601 words, 3 figures and 1 table NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

perpetuity. It is made available under a CC-BY-NC 4.0 International license .

# 51 BACKGROUND

Gender equity in medical research refers to the equal representation, recognition, and 52 53 valuation of all researchers regardless of their gender. Women and other 54 underrepresented groups continue to face barriers in diverse research fields. The 55 proportion of graduating female medical students is equal to that of men<sup>1</sup>. Yet, publications are more frequently authored<sup>2,3</sup>, peer-reviewed<sup>4</sup>, and editorially evaluated 56 by men<sup>5,6</sup>. Based on a report by the Association of American Medical Colleges, the 57 58 proportion of women faculty had increased to 41% in 2019, while their proportion as 59 department chairs remained at 18%<sup>7</sup>.

60

Abrogating the gender gap in all levels of academia are official priorities of the National 61 Institutes of Health<sup>8</sup> and the European Commission<sup>9</sup>. Promoting gender equity can lead 62 to more representative and balanced research. In addition, it can help to reduce bias 63 64 and discrimination in academic careers by ensuring that all researchers are treated 65 fairly and given equal opportunities to succeed.

66

We recently discovered previously undocumented Anglocentric bias in leading medical 67 journals based on publication counts and their citation frequency<sup>10</sup>. We reasoned that 68 gender underrepresentation could be studied with a similar approach and explore 69 publication differences in research fields, scientific journals, and collaboration scope. 70

71 72

# 73 METHODS

### 74 Data collection

75 For this study, we collected data from five medical journals that were ranked highest in the Journal Citation Reports 2022 and published primarily original articles (Fig. 1A). 76 77 These were The New England Journal of Medicine (NEJM), Nature Medicine 78 (NatMed), Journal of the American Medical Association (JAMA), The BMJ, and The 79 Lancet. We included all original articles published from 2010 to 2019 totaling to 10,558 80 articles. We excluded more recent publications to avoid bias related to the COVID-19 pandemic and to ensure equal opportunities for accumulating citations. We queried for 81 82 articles in the Web of Science database by Clarivate Plc with the terms 83 "(((SO=(NATURE MEDICINE OR LANCET OR NEW ENGLAND JOURNAL OF MEDICINE OR JAMA JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION OR 84 85 BMJ BRITISH MEDICAL JOURNAL)) AND DT=(Article)) AND PY=(2010-2019))". This allowed us to download metadata for each article, including the names of the authors, 86 87 the address of the corresponding author, and the total number of citations.

88 89

### Preprocessing variables of interest 90

91 We measured the impact of an article by its average citation count per year. To 92 determine the number of authors, we summed the frequency of the semicolon ";" 93 delimiter between author names and then added 1. For author names, only the initials 94 of the forename were available for 538 first (5.1%), 529 second (5.0%), 780 second 95 last (7.4%) and 572 last authors (5.4%). In total, we identified 7,558 unique forenames 96 and defined the gender for 7,113 (94.1%) forenames with the genderizeR library based 97 on the genderize.io database.

98

99 To determine the geographic location (latitude and longitude coordinates) of the 100 primary institutes where the research was conducted, we applied the ggmap library 101 and Google's Geocoding API to the corresponding author's address. If we were unable

to successfully match the address, we geolocated the text-mined city and country of 102 103 the address. The combined approach resulted in successful geolocation of 10,730 out 104 of 10,732 total unique address (100.0%).

105

106 In the keyword analysis, we identified 21,820 unique keywords. As most of these were 107 rarely used, we included only keywords employed in at least 20 articles (n=583, 2.7%).

- 108
- 109 Statistical analysis

We used the Wilcoxon rank-sum test (unpaired, two-tailed) to compare two continuous 110 111 variables and the Kruskal-Wallis test to compare three or more continuous variables. 112 For categorical variables, we used the  $\chi^2$  test. We adjusted p values using the 113 Benjamini-Hochberg correction. To compare two linear regression slopes, we tested the significance of the interaction term using T-test. All statistical analyses and 114 115 visualizations were conducted using R 3.5.1. using base, tidyverse, fastDummies, 116 maps, reshape2, ggmap, data.table, countrycode, ggpubr, ggrepel, rstatix, ggdendro 117 and dendextend libraries.

- 118
- 119

#### RESULTS 120

121 The gender gap in productivity is declining but not the gap in citation count

122 First, we interrogated how authoring patterns would differ between female and male 123 researchers in original articles published in 2010-2019 in five leading medical journals 124 (Fig. 1A). Publications with a male first author were two times and with a male last 125 author three times as common compared to publications authored by a female first and 126 last author, respectively (Table 1). In addition, publications with female first authors were 46.8% more likely to have a female last author compared to publications with a 127 128 male first author ( $\chi^2$  p<0.001, Fig. 1B).

129

130 Temporally, the publication count has gradually converged between women and men 131 (Fig. 1C). We fitted a linear regression for publication count using publication year and gender as covariate. The interaction term of the regression model was significant 132 133 (coefficient -14.5, p=0.024) indicating that the yearly decline of first authorships for 134 male researcher was higher in comparison to female researchers. The decline was 135 even more distinct for last authors (coefficient -24.0, p<0.001 for the interaction term; 136 Fig. 1D). Collectively, the gender gap in top medical research has declined with an 137 average 14.5 publications/year for first authors and 24 publications/year for senior 138 authors.

139

140 Female-authored articles gathered also fewer citations (Table 1). The median citation 141 frequency by publication increased in line with rising medical journal impact factors 142 (Fig. 1E-F). However, the ascent was equal between male and female researchers for 143 both first (p=0.53 of the interaction term between publication year and gender; Fig. 1E) 144 and last authors (p=0.67; Fig. 1F) implying no convergence in accumulated citations.

- 145
- 146

#### 147 The gender gap in productivity varies by country

The proportion of women authoring in leading medical journals varied by the country 148 149 where the research had been conducted (Fig. 1G). Female authors were least common 150 in publications originating from South Korea, Japan, Singapore, Germany, and Austria, 151 whereas they were most frequent from South Africa and India. The gender gap was 152 more pronounced for second last and last authors with 24.8% female authors in

153 median across countries compared to first and second authors with 34.5% female 154 authors (p<0.001). Gender underrepresentation was highest in South Korea with 0.0% 155 (0/28), Japan with 0.9% (1/106) and Austria with 9.1% (3/33) female last authors.

156

157 Next, we examined longitudinal national trends in gender representation. The genderassociated difference in publication number decreased rapidly in UK, Canada, and 158 159 Belgium both for first and last authors (Supplementary Fig. 1A-B). Yet, the gender gap 160 remained evident at the end of the follow-up in many countries, such as in the USA, 161 UK, and China, and particularly at the last author level. In opposite to the general trend, the gender gap rose for last authors of publications from Israel and both for first and 162 163 last authors from China. Collectively, these findings imply that the longitudinal trend 164 could guide in customizing national measures to mitigate gender underrepresentation. 165

166

167 Collaboration scope varies by gender and is associated with publication and citation 168 count

169 Next, we interrogated the significance of collaboration scope over gender-related 170 differences in citation count. The number of co-authors (R 0.46, p<0.001) correlated with citation count. Male first and last authors published in average with one author 171 172 more than their female colleagues (Table 1). To measure the independent impact of 173 author count and gender on citations, we applied a linear regression model using these 174 and their interaction term as covariates. In parallel, we examined the association of 175 gender and citations in subgroups of categorized authorship count. Both approaches 176 revealed that author count and gender are independent but weak predictors of future 177 citation count (Supplementary Table 1, Supplementary Fig. 2).

178

179 When examining the temporal evolution of authorship patterns, we observed a stable 180 median increase of 9 additional authors per article during the 10-year follow-up 181 (Supplementary Fig. 3A). Publications with more than 11 authors doubled in that time 182 and these accumulated faster citations per article (2.99 citations/article/year vs. 1.19 citations/article/year; Fig. 2A-B). The inclination was also 55.3% steeper for men as 183 184 first author and 94.1% for men as last authors during the 10-year follow-up compared 185 to corresponding female authors (Supplementary Fig. 3B). Instead, the slope between 186 first and last authors did not differ when comparing separately women and men (Supplementary Fig. 3B). 187

188

To study national differences in collaboration scope, we compared the number of 189 190 authors for publications by gender and by country (Fig. 2C). While the number of 191 authors did not differ for most countries, almost all variation in the number of authors 192 were related to the geographical origin of the research. Publications from USA, UK, 193 Denmark, and Sweden shared the least authors (Fig. 2C). On the opposite, publications from France, Germany, China, and South Korea were associated with 194 195 more numerous authors/article (Fig. 2C). Together, the findings emphasize that the 196 geographical origin of the research has a more pronounced association with 197 collaboration scope than gender.

198 199

200 The thematic and journalistic disparity between female and male medical researchers 201 Previous reports have suggested differences in funding, mentorship training, and 202 household and caregiving responsibilities between women and men to account for the gender gap in medical publishing<sup>2,11</sup>. We hypothesized whether the areas of research 203

medRxiv preprint doi: https://doi.org/10.1101/2022.12.21.22283801; this version posted December 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

would differ by gender. We observed a clear correlation between keyword-associated 204 citation frequency and the proportion of male first (R 0.40, p<0.001) and last authors 205 206 (R 0.40, p<0.001; Fig. 3A). Keywords associated with highest publication citations 207 were related to phase II-III trials, oncology, immunotherapy, chemotherapy and 208 antibody-based therapy, and were enriched with publications authored by men, especially at the senior author level. On the contrary, the 20 keywords associated with 209 210 least citations were predominantly associated with female authorship, notably in the 211 context of first authorship (Fig. 3A). The keywords covered healthcare-related themes 212 such as patient involvement, insurance, guality-of-care, and access.

213

214 Beside gender-associated distinctions in research subfields, we sought to understand whether publications sharing the same keyword would differ by their accumulated 215 citation count. Out of 583 keywords, we included 579 for comparing first author and 216 217 582 for last author gender comparison as these keywords were used by both genders. For first authors, 78/579 compared to 5/579 keywords resulted in higher citation count 218 219 (adjusted p<0.05) when employed by men vs. women, respectively. For last authors, 220 the corresponding proportion was 32/582 for men and 8/582 for women (adjusted 221 p<0.05). Collectively, the findings indicate that within the same fields of research, 222 publications authored by women accumulate fewer citations compared to publications 223 authored by men.

224

225 To conclude, we investigated the proportion of female-authored publications by 226 journals. The absolute difference in the proportion of first authors between the five 227 journals was 14.8% for first and 10.2% for last authors (Fig. 3B). Female first and 228 second authors were least frequent in articles published in NEJM and Lancet, whereas 229 second last and last authors were least common in *NatMed*. According to this analysis, 230 BMJ and JAMA were the most representative journals considering all four authorship 231 positions.

232 233

### 234 DISCUSSION

235 Available bibliometric data can reveal novel information on the equity and diversity of 236 scientific research. Longitudinal data permits studying temporal trends, which can be 237 crucial for monitoring purposes. Here, we presented publication disparity in five leading 238 medical journals between 2010-2019. The study analyses can be easily replicated for 239 any journals and any time period using data available at the Web of Science's database 240 and codes published with this study.

241

242 Gender underrepresentation in scientific publishing has been documented earlier. 243 Women tend to obtain university-level degrees more commonly than male in OECD countries (47% vs. 32%)<sup>12</sup>. However, female doctoral students submit and publish less 244 than their male colleagues<sup>2,3</sup>. The difference is largest in natural/biological sciences 245 and engineering and has been hypothesized to result of unevenly distributed resources 246 247 and support<sup>3</sup>. Gender disparities in publications have been shown to correlate with future academic rank signifying sustained impact on professional careers<sup>13</sup>, salary and 248 249 job satisfaction<sup>14</sup>.

250

Our findings in leading medical journals are in line with previous observations. For the 251 252 first time to our knowledge, we report that the inequity is country-specific. While the 253 distinction is visible across first, second, second last and last author positions, the 254 gender gap is most pronounced at the senior author level. In particular, publications

with a corresponding author in Germany, Austria, Japan, South Korea, and Singapore 255 256 had the fewest female authors. The finding implicate that the bias likely arises from 257 national and cultural factors rather than editorial or peer-review processes, which had previously indicated mixed results<sup>15,16</sup>. Moreover, the gender gap in first and last 258 authorships has steadily declined during the last decade in most countries, but the 259 progress has been less evident or even contrary in some countries such as China and 260 Israel. The gender disparity has decreased more rapidly in last authorships possibly 261 262 reflecting a response to the more pronounced gender underrepresentation compared 263 to first authors or dynamics in generation transition.

264

265 Gender underrepresentation was observable in all top medical journals. However, 266 female first and second authors were least frequent in articles published in NEJM or Lancet, whereas second last and last authors were least common in NatMed. The 267 268 findings are in line with a previous study examining gender representation as first authors using 4 out of 5 similar journals during 1994-2014<sup>17</sup>. The trend was replicated 269 270 also in an article examining author disparity in leading medical journals during the 271 COVID-19 pandemic<sup>11</sup>. In that study, no difference was found when examining first and last author gender of COVID vs. non-COVID-related research<sup>11</sup>. Confirming our 272 longitudinal findings, the proportion of female first (36.2% vs. 33.6%) and especially 273 274 last (29.5% vs. 23.4%) authors has increased in 2020 compared to our data covering 275 2010-2019. Similar findings were not evident between our and the earlier study 276 covering 1994-2014 emphasizing that the gender representation has started to 277 improve only recently.

278

279 By comparing research keywords, first and especially last male authors tended to publish clinical trials and oncology-associated studies, which accumulated highest 280 281 citation counts. The difference in research focus has not been demonstrated before 282 and likely accounts for some of the journal-specific gender disparities. However, our 283 data also indicated that publications first or last-authored by men tend to accumulate 284 more citations implying that differences in research fields explain only limited 285 variability.

286

In line with results of our study, articles with a woman researcher as first or last author 287 have been shown to accumulate fewer citations<sup>18</sup>. While medical research was not 288 289 included, a previous study using 1.5 million interdisciplinary papers in 1779-2011 has 290 indicated that male first authors tend to self-cite 56% themselves more commonly than 291 their female colleagues, and even more during the last decades<sup>19</sup>. According to a 292 recent preprint, males were more commonly quoted in Nature science journalism, 293 which could skew the recognition and future citation probability of publications by 294 gender<sup>20</sup>.

295

Between 2010-2019 articles the number of authors increased by nine reflecting a 296 297 fundamental change in research towards larger collaboration and building consortia. 298 This correlated with higher citation frequency per article. While the general citation rate 299 increased in the top medical journals, the rate was over 2.5 times faster in articles with 300 more collaborators. We found that publications with male researchers had more authors in concordance with a previous study<sup>21</sup>. In our study, both author gender and 301 302 number of collaborators were independent but weak predictors of future citation rate. 303

304 Computational gender prediction could be a source of error as some names can be 305 used by both women and men, especially in different countries. The name-based medRxiv preprint doi: https://doi.org/10.1101/2022.12.21.22283801; this version posted December 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

306 analysis may also misclassify authors with non-binary genders. However, previously reported gender-related bibliometric observations were in line with our findings 307 indicating that the gender prediction based on authors' first names provided realistic 308 309 results. Inclusion of gender ethnicity and career stage were unavailable but could be 310 important factors to further study gender representation.

311

312 In summary, this computational audit indicated that gender disparity in medical 313 research is country-specific, partly related to distinct research focus and more evident 314 at the senior researcher level. The findings might reflect available financial resources and research support. The analysis also highlighted that the gender gap is decreasing 315

316 with country-dependent variability. medRxiv preprint doi: https://doi.org/10.1101/2022.12.21.22283801; this version posted December 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

## 317 ADDITIONAL INFORMATION

318

#### 319 ACKNOWLEDGMENTS

320 The author wishes to thank Susanna Lallukka-Brück for her patience, understanding, 321 and insightful comments. The author is grateful for Olli Dufva and to the members of 322 the Hematoscope Lab for discussion and comments. Certain data included herein are

323 derived from Clarivate Web of Science. © Copyright Clarivate 2022. All rights reserved.

324

#### 325 **AUTHORS' CONTRIBUTIONS**

- Conception and design; Collection and assembly of data; Data analysis; Manuscript 326
- 327 writing; Manuscript editing; Data interpretation; Final approval of manuscript: O.B. 328

### 329 ETHICS APPROVAL AND CONSENT TO PARTICIPATE

- 330 The study complied with the Declaration of Helsinki.
- 331

# CODE AND DATA AVAILABILITY 332

- 333 Codes and a 100-row data example are available at
- 334 https://github.com/obruck/International-Research-Impact. Raw data can be
- downloaded from Clarivate Web of Science, with instructions provided in the Github 335
- 336 repository.

# 337

### COMPETING INTERESTS 338

- 339 O.B. declares no Competing Non-Financial Interests but the following Competing 340 Financial Interests: consultancy fees from Novartis, Sanofi, and Amgen, outside the
- 341 submitted work. 342

### 343 **FUNDING INFORMATION**

- 344 This study was supported by research grants from the Helsinki University Hospital.
- 345

perpetuity. It is made available under a CC-BY-NC 4.0 International license .

# REFERENCES 346

- Association of American Medical Collegues. https://www.aamc.org/data-347 1 348 reports/students-residents/interactive-data/2022-facts-enrollment-graduates-349 and-md-phd-data. 2022.
- Nguyen M, Chaudhry SI, Asabor E, et al. Variation in Research Experiences 350 2 351 and Publications During Medical School by Sex and Race and Ethnicity. JAMA 352 Netw Open 2022; 5: e2238520-e2238520.
- 353 3 Lubienski ST, Miller EK, Saclarides ES. Sex Differences in Doctoral Student 354 Publication Rates. Educ Res 2017; 47: 76-81.
- Lerback J, Hanson B. Journals invite too few women to referee. Nature 2017; 355 4 356 **541**: 455–7.
- Wing DA, Benner RS, Petersen R, Newcomb R, Scott JR. Differences in 357 5 Editorial Board Reviewer Behavior Based on Gender. J Women's Heal 2010; 358 359 **19**: 1919–23.
- Rochon PA, Davidoff F, Levinson W. Women in Academic Medicine 360 6 361 Leadership: Has Anything Changed in 25 Years? Acad Med 2016; 91. https://journals.lww.com/academicmedicine/Fulltext/2016/08000/Women\_in\_Ac 362
- 363 ademic Medicine Leadership Has.14.aspx.
- Association of American Medical Collegues. https://www.aamc.org/data-364 7 reports/data/2018-2019-state-women-academic-medicine-exploring-pathways-365 366 equity. 2019.
- National Institutes of Health (NIH) Gender Inequality Task Force. 367 8 368 ADDRESSING GENDER INEQUALITY IN THE NIH INTRAMURAL 369 RESEARCH PROGRAM ACTION TASK FORCE REPORT AND 370 **RECOMMENDATIONS. 2016**
- https://diversity.nih.gov/sites/coswd/files/images/2018-05/Gender Inequality 371 372 Task Force final AD may 30 2018.pdf.
- 373 European Commission Directorate-General for Research and Innovation. 9 Gender equality: achievements in Horizon 2020 and recommendations on the 374 375 way forward. Publications Office, 2020 DOI:doi/10.2777/009204.
- Brück O. The Geographical Gap in Leading Medical Journals a Computational 376 10 377 Audit. medRxiv 2022; : 2022.12.12.22283368.
- Misra V, Safi F, Brewerton KA, et al. Gender disparity between authors in 378 11 379 leading medical journals during the COVID-19 pandemic: a cross-sectional 380 review. BMJ Open 2021; 11: e051224.
- 381 12 OECD. Organisation for Economic Co-operation and Development. Education 382 at a Glance 2012, 2012.
- Carr PL, Raj A, Kaplan SE, Terrin N, Breeze JL, Freund KM. Gender 383 13 384 Differences in Academic Medicine: Retention, Rank, and Leadership 385 Comparisons From the National Faculty Survey. Acad Med 2018; 93. 386 https://journals.lww.com/academicmedicine/Fulltext/2018/11000/Gender\_Differ 387 ences in Academic Medicine .32.aspx.
- Holden C. General Contentment Masks Gender Gap in First AAAS Salary and 388 14 389 Job Survey. Science (80-) 2001; 294: 396-411.
- 390 15 Kern-Goldberger AR, James R, Berghella V, Miller ES. The impact of double-391 blind peer review on gender bias in scientific publishing: a systematic review. 392 Am J Obstet Gynecol 2022; 227: 43-50.e4.
- Fox CW, Burns CS, Muncy AD, Meyer JA. Gender differences in patterns of 393 16 394 authorship do not affect peer review outcomes at an ecology journal. Funct 395 Ecol 2016; 30: 126-39.
- 396 17 Filardo G, da Graca B, Sass DM, Pollock BD, Smith EB, Martinez MA-M.

medRxiv preprint doi: https://doi.org/10.1101/2022.12.21.22283801; this version posted December 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

397		Trends and comparison of female first authorship in high impact medical
398		journals: observational study (1994-2014). <i>BMJ</i> 2016; <b>352</b> : i847.
399	18	Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto CR. Bibliometrics: Global
400		gender disparities in science. Nature 2013; 504: 211–3.
401	19	King MM, Bergstrom CT, Correll SJ, Jacquet J, West JD. Men Set Their Own
402		Cites High: Gender and Self-citation across Fields and over Time. Socius 2017;
403		<b>3</b> : 2378023117738903.
404	20	Davidson NR, Greene CS. Analysis of science journalism reveals gender and
405		regional disparities in coverage. <i>bioRxiv</i> 2022; : 2021.06.21.449261.
406	21	Aksnes DW, Piro FN, Rørstad K. Gender gaps in international research
407		collaboration: a bibliometric approach. Scientometrics 2019; <b>120</b> : 747–74.
408		

medRxiv preprint doi: https://doi.org/10.1101/2022.12.21.22283801; this version posted December 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

#### TABLES 410

- Table 1. Association of gender authorship with publication metrics. Median and 25-411
- 75% interquartile ranges are reported. 412

Authorship	Variable	Female	Male	P-value
First	# Publications	3311	6554	
First	# Citations	22.0 [10.9-46.4]	29.0 [13.6-60.9]	***
First	# Authors	10.0 [6.0-17.0]	11.0 [6.0-18.0]	***
Last	# Publications	2256	7366	
Last	# Citations	23.5 [11.0-52.4]	28.2 [13.4-58.3]	***
Last	# Authors	10.0 [5.0-17.0]	11.0 [6.0-18.0]	***

413

medRxiv preprint doi: https://doi.org/10.1101/2022.12.21.22283801; this version posted December 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

#### FIGURE LEGENDS 414

Figure 1. Gender gap in medical publishing. (A) Study design. (B) Bar plot on the 415 416 association between first and last author gender. (C-D) Line plot and fitted linear 417 regression for the number of publications and (E-F) number of yearly-averaged 418 citations per publication by their publishing year and author gender. (G) Balloon plot 419 on the association of the first, second, second last and last author gender distribution 420 by the affiliation nationality of the corresponding author. The balloon color reflects the 421 gender bias. To emphasize differences white color defines the median, i.e. men than 422 women authors (red = more women, orange = balanced, blue = more men). The 423 balloon size reflects the absolute deviation from a balanced gender distribution. 424 425 Figure 2. Scope of collaboration by genders. (A) Line plot and fitted linear regression for the number of publications and (B) yearly-averaged citations per 426 427 publication by their publishing year and number of authors. (C) Balloon plot on the association of the first and last author gender distribution by the affiliation nationality of 428

429 the corresponding author. The balloon color reflects the number of authors by country

- 430 (column) compared to all other countries and the balloon size the adjusted p value of 431 that comparison.
- 432

433 Figure 3. Gender gap by subfields and journals. (A) Left side spiral plots illustrate 434 the keywords occurring ≥20 times and arranged by their citation impact starting from 435 the outside layer (most cited) towards the inner core (least cited). Right side plots show 436 the proportion (size of the bar) of the most and least cited keywords. The color of the 437 bars in all spiral plots illustrate the gender balance. (B) Bar plot illustrating the 438 proportion of publications in medical journals originating from different countries.





FIG 3

