

Pre-print of published article: Nielsen, M. W., Andersen, J. P., Schiebinger, L., & Schneider, J. W. (2017). One and a half million medical papers reveal a link between author gender and attention to gender and sex analysis. *Nature Human Behaviour*, 1, 791-796. DOI: 10.1038/s41562-017-0235-x

# One and a half million medical papers reveal a link between author gender and attention to gender and sex analysis

Mathias Wullum Nielsen<sup>1\*</sup>, Jens Peter Andersen<sup>2</sup>, Londa Schiebinger<sup>1</sup>, and Jesper W. Schneider<sup>2</sup>

<sup>1</sup>History of Science, Stanford University, 450 Serra Mall, Stanford, CA 94305, United States

<sup>2</sup>Danish Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University, Bartholins Alle 7, Aarhus 8000, Denmark

\*To whom correspondence should be addressed. Email: [mwn@ps.au.dk](mailto:mwn@ps.au.dk)

**Keywords:** gender and sex analysis, medicine, gender diversity, women in science, social studies of science

**Gender and sex analysis is increasingly recognized as a key factor in creating better medical research and healthcare<sup>1-7</sup>. Using a sample of more than 1.5 million medical research papers, our study examined the potential link between women’s participation in medical science and attention to gender- and sex-related factors in disease-specific research. Adjusting for variations across countries, disease topics and medical research areas, we compared the participation of women authors in studies that do and do not involve gender and sex analysis. Overall, our results show a robust positive correlation between women’s authorship and a study’s likelihood of engaging gender and sex analysis. These findings corroborate discussions of how women’s participation in medical science links to research outcomes, and illustrate the mutual benefits of promoting both women’s scientific advancement and the integration of gender and sex analysis into medical research.**

Despite a burgeoning scholarship, gender and sex differences remain unaddressed in large parts of medical research. Consider, for instance, the paradigmatic example of biological sex differences in cardiovascular disease (CVD) – the no. 1 ‘killer’ among adult populations in the Western world. Since the 1980s, the annual number of CVD-related deaths has been higher for women than for men; but women continue to be underrepresented as participants in clinical trials<sup>8-10</sup>. A similar pattern is found in cancer research, where male research subjects dominate the trials<sup>11</sup>, or in studies of cell tissue and laboratory animals, where the sex of the subject often goes unrecorded<sup>12,13</sup>. Even in research involving both sexes, data is often not analysed by sex, and results may therefore not be accurate for either male or female subjects<sup>14</sup>. Both women and men face disadvantages under these circumstances. Around one third of osteoporosis-related hip fractures occur in elderly men, but osteoporosis research tends to focus on women, resulting in osteoporosis in men being underdiagnosed, undertreated, and underreported<sup>15</sup>.

Not only biological sex but also gender (i.e., the social attitudes and behaviours associated with being a woman or a man) is known to be a critical determinant of human health<sup>8,9</sup>. Women and men are exposed to different occupational hazards (e.g., ergonomic demands and psychosocial stressors), and differ on a wide array of health-related lifestyle behaviours (e.g., exercise, alcohol and tobacco use)<sup>16,17</sup>. Further, gendered psychosocial factors have been found to predict recurrent outcomes in patients with acute coronary syndrome and to moderate patient-perceptions of pain<sup>18,19</sup>. Despite robust evidence, associations between gender, biological sex and health outcomes remain largely neglected in the literature<sup>20</sup>, with potentially life-threatening and costly consequences. Of the ten drugs withdrawn from the U.S. market between 1997 and 2000, eight involved health risks for women that may have been avoided if more attention had been devoted to gender- and sex-related factors<sup>21</sup>.

These examples demonstrate how gender and sex analysis (henceforth GSA) has the potential to improve medical diagnosis and treatment. Here we define GSA as scientific approaches aimed at understanding how social and behavioural differences between women, men and gender-diverse people (gender analysis) and biological differences between female and male research subjects (sex analysis) relate to health outcomes.

A growing number of medical scholars and journal editors already acknowledge the importance of GSA in fostering excellence in medical science and healthcare<sup>1-7</sup>. Science agencies, too, subscribe to this idea (for an overview of national policy efforts see ref.<sup>22</sup>). Through the Horizon 2020 program, the European Commission (EC) has committed itself to “integrating the gender dimension into the research and innovation content (...) in order to improve the quality of research and stimulate innovation”<sup>23</sup>. Similarly, the US National Institutes of Health (NIH) recognizes that both gender and sex “play a role in how health and disease processes differ among individuals,” and have implemented guidelines

recommending that “sex as a biological variable will be factored into research designs, analyses, and reporting in vertebrate animal and human studies”<sup>24</sup>.

Both the EC and NIH prioritize policies: 1) to advance the careers of women scientists, and 2) to encourage gender and sex analysis in research design. Yet, we know little about how these policy objectives may be linked. Are women and men equally likely to integrate gender and sex analysis into their research designs?

While recognizing the crucial role played by funding agencies and scholarly journals in putting gender and sex on the scholarly agenda, our study is unique in analysing potential gender differences in medical scholars’ involvement in GSA. If gender variations are detected in our data, we are not proposing that they stem from innate differences in women’s and men’s scientific styles or preferences. Instead, we follow tenets in the literature on cultural diversity, and see gender as a cultural category shaping cognition, experience and perspective in the workplace, with implications for research interests and focus<sup>25-27</sup>.

Sociological research illustrates how societal norms and expectations operate to cultivate gender-differentiated career aspirations in higher education and the labour market<sup>27</sup>. Existing research, for instance, documents clear gender differences in students’ selection of fields and medical specialties<sup>28,29</sup>; but gender norms and expectations may also spur variations in researchers’ choice of different forms of research within a given field or specialty<sup>27</sup>. Indeed, this idea already finds some support in the literature. Using meta-data from over eight million articles in JSTOR, Jevin West and colleagues demonstrate notable differences in women’s and men’s primary areas of specialization in a range of scholarly disciplines spanning the natural sciences, social sciences and humanities<sup>30</sup>. Using data from *Sociological Abstracts*, Ryan Light finds that women sociologists are overrepresented in research areas such as gender, race, family and medicine, while men dominate in political, comparative and economic sociology<sup>31</sup>. Juan Dolado and colleagues demonstrate notable

variations in women's and men's primary subfields in economic research: health, education, welfare and labour economics tend to be popular topics among women, while agricultural economics, fluctuations/business cycles, general equilibrium, comparative systems, and corporate finance are popular among men<sup>32</sup>. Finally, recent studies document a clear overrepresentation of women authors in social science journals on gender, feminism and sexuality<sup>33,34</sup>.

Given the traditional conception of gender- and sex-related research as a woman-dominated domain in the academy, we expect women's involvement in GSA to be more congruent with prevailing gender norms and expectations than men's. Broader cultural influences about appropriate gender-typed work may, in other words, draw a disproportionate number of women towards (and men away from) this form of research. A pioneering study finds women medical investigators more likely to address gender and sex in successful research proposals for the Canadian National Institutes of Health<sup>35</sup>. However, the results reported in this study are descriptive, and do not adjust for potential spurious associations resulting from differences in women's and men's participation across diseases and medical subfields.

Our study has been designed to accommodate these potential biases. Using a global sample of more than 1.5 million medical research papers and adjusting for variations in the gender-composition of author teams across disease topics, countries and medical research areas, we compared women's general participation and share of first and last authorships in studies that do and do not involve GSA. Our prior conjectures were that:

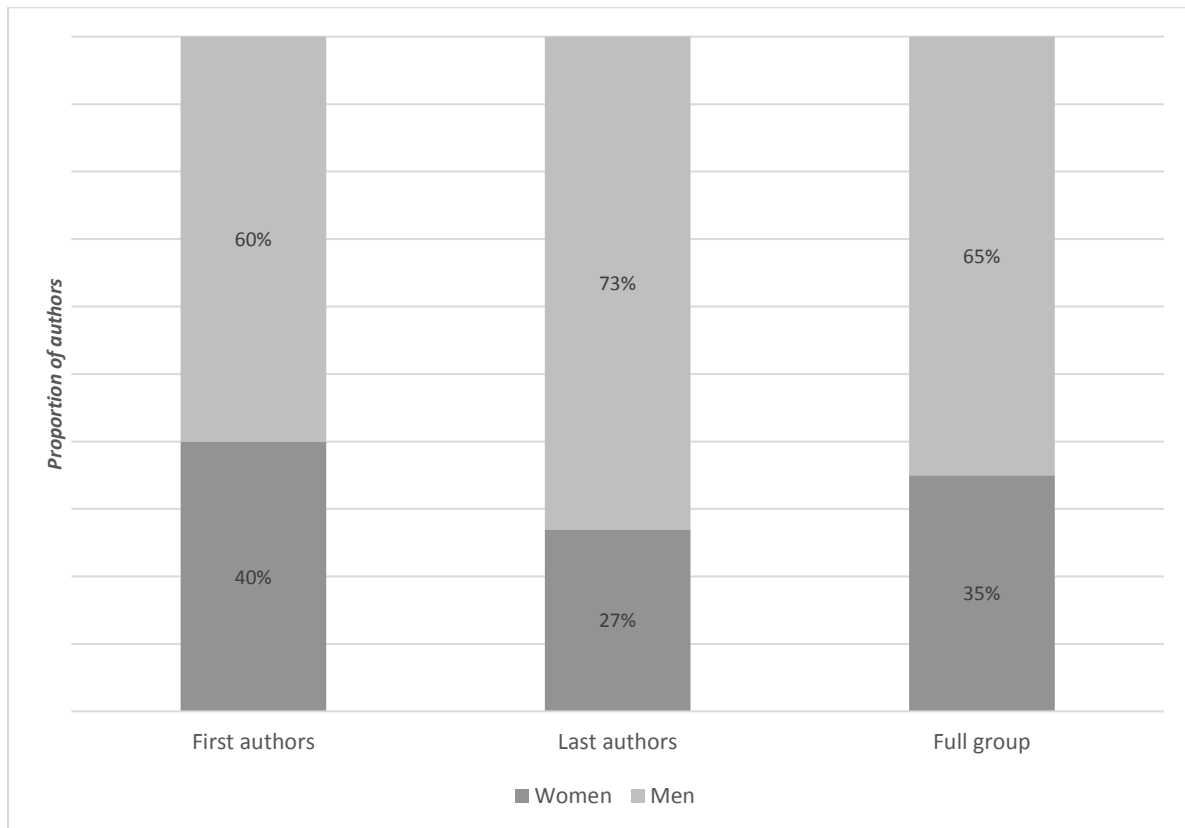
- 1) A study's likelihood of involving GSA increases with the proportion of women among its authors.
- 2) A study's likelihood of involving GSA increases if the first author is a woman.

- 3) A study's likelihood of involving GSA increases if the last author is a woman.

We focused on disease-specific medical research using peer-reviewed scholarly articles published in the period 2008-2015 as our basic unit of analysis. To identify studies using gender and sex analysis we used the GenderMed Database<sup>36,37</sup>, which is an extensive bibliographic archive of the existing medical literature analysing gender and sex differences. To determine the gender of authors, we used the name-to-gender assignment algorithm Gender API<sup>38</sup>. The analysis took place in two steps. First, we used descriptive statistics to document global gender disparities in authorship participation in disease-specific medical research. Second, we employed three logistic regression analyses to estimate how women's overall representation in the author group and as first and last authors influenced a study's likelihood of systematically using GSA. All models were estimated in a Bayesian framework with weakly informed default prior distributions<sup>39</sup> (see *Materials and Methods* for more details).

Figure 1 provides descriptive information on women's representation per author group and share of first and last authorships in studies published from 2008 through 2015. Women comprise 40% of first authors, 27% of last authors and, on average, 35% of authors per author group in the full sample. Importantly women's participation as first and last authors varies across geographical groupings (see Supplementary Table 1). Gender disparities are largest in East Asia (women first authors: 24%, women last authors: 16%) and smallest in Latin America (women first authors: 52%, women last authors: 40%). North American and Western European figures lie close to the global average. This latter is not surprising, given that 67% of first authors and 69% of last authors in our sample are affiliated with institutions in North America or Western Europe.

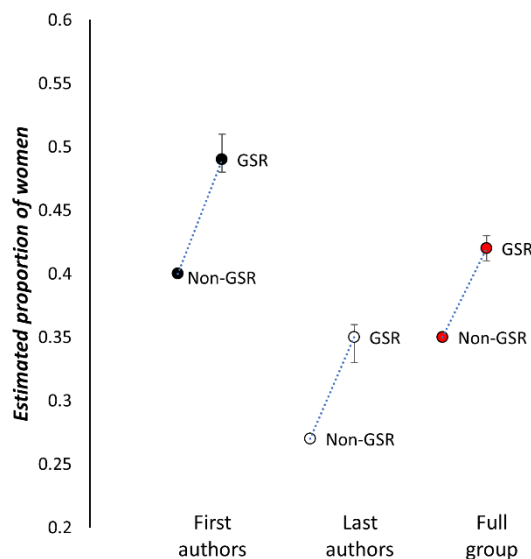
**Figure 1. Women’s global share of first authors, last authors and full-group participation in disease-specific medical research. Women comprise 40% of first authors (men 60%), 27% of last authors (men 73%) and 35% of authors overall per paper (men 65%). N: 1,542,690.**



As a second step in the analysis, we illuminated differences in women’s (and men’s) full group participation and share of first and last authorships in studies that do and do not involve GSA. Table 1 displays odds ratios, standard deviations and 95% credible intervals (CI) for the logistic regressions predicting a study’s likelihood of using GSA (outcome variable: Non-GSA=0, GSA=1) (See *Materials and Methods* for variable specifications). Model 1 and 2 specifies the effects attributable to female first and last authors, and model 3 determines the impact of women’s overall representation in the author group. Conditioned on our model specifications and prior assumptions, women’s participation positively correlates with a study’s likelihood of using GSA for all three groups. This is in accordance with our initial conjectures. Odds ratios are 1.66 (CI: 1.53-1.79) for first authors ( $f\_first$ ), 1.56 (CI: 1.44-1.68) for last authors ( $f\_last$ ) and 3.14 (CI: 2.74-3.59) for full group representation ( $fw$ ).

Figure 2 displays estimated marginal means (EM-means) and credible intervals for the main predictors. EM-means enable us to specify mean-based differences in women's (and men's) estimated participation across GSA and non-GSA, while adjusting for co-variation attributable to all other variables in the models. As illustrated in the figure, women comprise 40% of first authors in non-GSA studies and 49% in GSA studies. Differences in women's participation for last authors and full author groups are 27% vs. 35% and 35% vs. 42%, respectively. The relative difference is largest for last authors (30%) and smallest for full author groups (20%). This indicates that the effect attributable to women's participation is strongest when women serve as leaders of the author group.

**Figure 2. Plot of estimated marginal means.** The figure specifies the estimated marginal means for *f\_first*, *f\_last* and *fw* in Model 1, 2 and 3. Error bars represent 95% credible intervals (for estimate specifications, see Supplementary Table 5). The plots visualize women's participation (relative to men's) as first authors, last authors and overall representation in the byline, for studies that do and do not involve GSA. The figure shows that women's estimated share of authorships is higher in GSA studies than in non-GSA studies for all three author variables.



In comparison, logistic regression models excluding all variables but the main predictors (*f\_first*, *f\_last* and *fw*) have odds ratios of 1.91 (CI: 1.77-2.06) for first authors, 1.76 (CI: 1.63-1.89) for last authors, and 3.87 (CI: 3.43-4.36) for full author groups (for



model specifications see Supplementary Tables 6, 7 and 8). Further, using arithmetic means, differences in women's average participation in non-GSA and GSA studies are 40% vs. 53% for first authors, 27% vs. 37% for last authors, and 35% vs. 45% for full author groups.

**Table 1. Binary Logistic Regression Model Predicting GSA**

Parameter	Model 1 (first author)			Model 2 (last author)			Model 3 (full group)		
	Odds Ratio	SD	95% CI	Odds Ratio	SD	95% CI	Odds Ratio	SD	95% CI
<i>f_first</i>	1.66	0.07	[1.53:1.79]						
<i>f_first country</i>	2.75	0.80	[1.50:4.62]						
<i>f_first MeSH</i>	6.13	1.67	[3.55:10.00]						
<i>f_first SC</i>	2.11	0.47	[1.36:3.19]						
<i>f_last</i>				1.56	0.06	[1.44:1.68]			
<i>f_last country</i>				4.03	1.31	[2.02:7.13]			
<i>f_last MeSH</i>				4.76	1.55	[2.48:8.49]			
<i>f_last SC</i>				4.32	1.09	[2.60:6.84]			
<i>fw</i>							3.14	0.22	[2.74:3.59]
<i>fw country</i>							2.19	0.85	[0.99:4.26]
<i>fw MeSH</i>							4.86	1.58	[2.55:8.58]
<i>fw SC</i>							1.98	0.52	[1.16:3.18]
Arab States	2.31	0.46	[1.55:3.34]	2.41	0.48	[1.62:3.48]	2.26	0.45	[1.50:3.26]
East Asia	1.79	0.22	[1.40:2.26]	1.96	0.25	[1.54:2.49]	1.90	0.25	[1.47:2.44]
Latin America	1.19	0.19	[0.85:1.60]	1.31	0.20	[0.95:1.76]	1.20	0.19	[0.86:1.62]
Oceania	1.25	0.18	[0.93:1.66]	1.49	0.22	[1.11:1.97]	1.33	0.20	[1.00:1.75]
South & West Asia	1.27	0.21	[0.91:1.73]	1.33	0.23	[0.96:1.83]	1.29	0.22	[0.99:1.77]
So.-Cen. & Eastern Europe	1.36	0.20	[1.00:1.80]	1.50	0.21	[1.10:1.97]	1.35	0.20	[0.92:1.77]
Sub-Saharan Africa	3.03	0.58	[2.01:4.28]	3.26	0.62	[2.19:4.61]	3.17	0.60	[2.13:4.48]
North America	2.08	0.22	[1.69:2.56]	2.37	0.25	[1.93:2.90]	2.12	0.22	[1.72:2.60]
Western Europe	2.00	0.22	[1.62:2.46]	2.53	0.28	[2.05:3.12]	2.16	0.23	[1.75:2.65]

N: 1,513,638

Note: Posterior summaries of odds ratios with 95% credible intervals for Bayesian logistic regression models with Cauchy informative priors predicting GSA. Commonwealth Independent States is the reference group for the geographical variables. For more model specifications, see Supplementary Tables 2, 3 and 4.

Two control variables in model 1, 2 and 3 also deserve attention. As displayed in Table 1, GSA studies are more likely to be carried out in disease-specific research areas with a high general representation of women. Odds ratios for this factor are 6.13 (CI: 3.55-10.00) for first authors (*f\_first MeSH*), 4.76 (CI: 2.48-8.49) for last authors (*f\_last MeSH*) and 4.86 (CI: 2.55-8.58) for full author groups (*fw MeSH*). Women's representation in the broader research areas (or specialties) circumscribing a given disease topic also has a positive effect. Odds ratios for this factor are 2.11 (CI: 1.36-3.19) for first authors (*f\_first SC*), 4.32 (CI:

2.60-6.84) for last authors ( $f\_last$  SC) and 1.98 (CI: 1.16-3.18) for full author groups ( $fw$  SC). This indicates that not only the gender composition of author groups but also the general gender composition of the context in which they operate play an important role in predicting a study's focus on GSA. We do not know, however, whether this finding results from the fact that women more often specialize in topics, where disease-specific gender- and sex-related variations are likely to exist.

In summary, our results provide global evidence linking a study's likelihood of involving GSA to the presence of women in the author group, especially in leading positions as first and last authors.

Canadian researchers have already documented gender differences in funding applicants' attention to GSA; but their analysis did not adjust for potential spurious associations resulting from differences in women's and men's participation across diseases and medical subfields<sup>35</sup>. We show that adjusting for such factors (and geographical variations) leads to more robust comparisons. Specifically, the odds ratios for the main predictors ( $f\_first$ ,  $f\_last$  and  $fw$ ) are reduced from 1.91 to 1.66 for first authors, 1.76 to 1.56 for last authors and 3.87 to 3.14 for full author groups, when covariates adjusting for women's participation across geographical groupings, disease-topics and medical sub-fields are factored into the logistic regression models.

Further, our analysis documents notable global gender disparities in authorship participation. Women comprise 40% of first authors, 27% of last authors and 35% of the authors per author-group in disease-specific research. Given the findings of this paper, the modest share of women last authors is particularly troublesome, since last authors typically take the lead in identifying, planning and developing research questions in the health-related fields.

Our study adds to the existing literature in several important ways. First, it empirically links science policy-makers' efforts to increase the numbers of women in academic medicine and to promote GSA. By doing so, it provides a strong argument for both policy objectives. In this regard, it is important to underline that GSA constitutes a small subset of the total number of articles published in the disease-specific medical literature. The majority of women and men are not engaged in this form of research.

Second, our findings highlight the importance of devoting more systematic attention to the link between gender diversity and research outcomes in the academy. In a recent opinion piece in *PNAS*, NIH Director of Scientific Workforce Diversity Hannah Valentine and NIH Director Francis Collins encouraged the scientific community to develop approaches specifically designed to document the impact of diversity on the quality and outputs of academic medicine<sup>40</sup>. Our study pushes the research agenda forward by empirically demonstrating how gender diversity can expand health solutions by diversifying research methods to include gender and sex analysis. That is to say, expanding gender equality may have broader implications for knowledge and health outcomes than previously suspected.

Future research might investigate whether similar relationships between women's participation in research and the likelihood of deploying GSA can be detected in other fields. Evidence from the social sciences, where women are relatively well-represented, suggests that this may be the case<sup>31,33,34</sup>, but in the STEM-related areas, such as engineering and computer science, where gender and sex analysis is less prevalent and men dominate research teams, this question remains unexplored. There is, however, no reason to believe that the situation in STEM would be any different from the medical sciences.

Like any other study, ours is characterized by certain caveats and limitations. First, the search algorithm employed to establish the corpus of GSA studies may, despite a

systematic and thorough screening strategy, not capture the full gamut of potentially relevant publications<sup>36,37</sup>. Consequently, an unspecified number of “false negatives” (i.e., eligible GSA studies not specified as such) may exist in our baseline group of non-GSA studies. However, since GSA-related research constitutes a small subset of the total number of articles published in the disease-specific medical literature, we do not expect this to influence our results in any noteworthy way. And if so, it would most likely imply even stronger support for our conjectures, given the higher level of female participation in GSA studies.

Second, status-related factors, not addressed in this paper, may play a part in explaining our findings. We know from social science research that when women enter male-dominated fields, subtle forms of gender segregation tend to persist. Women may, for instance, self-select or be ‘ghettoized’ into less competitive and prestigious work areas<sup>41,42</sup>. To address such issues, future studies could investigate whether the relative status of GSA as well as funding opportunities (or lack hereof) are pushing a disproportionate number of women towards (and men away from) this form of research. In a supplementary analysis (see Supplementary Methods), we take a first step towards illuminating potential status differentials. Specifically, we compare the distribution of GSA and non-GSA publications according to journal impact. Despite a vast array of problems with such measures, metrics of journal impact are widely used as proxies of achievement and status in the medical sciences<sup>43,44</sup>, and most medical scholars acknowledge the existence of a journal hierarchy in their disciplines. As demonstrated in Supplementary Table 9 and Supplementary Fig. 1, our data provides no indications of any systematic status variations between the journals where GSA and non-GSA studies are published, but more research is needed to illuminate the influence of other potential status- and resource-related differences.

Finally, engaging GSA requires know how. Anyone, irrespective of gender, can be trained to do this effectively. Indeed, granting agencies and universities have begun to initiate

workshops and incentive programs to integrate GSA into research design, and future studies could investigate the impact of such programs in enhancing the external validity and applicability of scientific research with human outcomes.

Our study establishes an empirical link between gender diversity in the scientific workforce and research outcomes. Our findings show a symbiotic relationship between increasing the numbers of women in academic medicine and enhancing excellence in research by incorporating gender and sex analysis. Hence, our study provides empirical evidence for science policy makers to promote both women's scientific careers and gender and sex analysis in research design. Taken together these objectives support the twin goals of diversity and excellence in science.

## **Materials and Methods**

**Data.** Data for this study were harvested from PubMed's Medline. Medline is one of the most exhaustive databases of the medical journal literature and offers a systematic hierarchy of Medical Subject Headings (MeSH) for indexing scholarly articles; e.g., on the basis of disease topics. We gathered bibliographic metadata for all Medline articles with disease-specific MeSH terms for the period 2008 through 2015, resulting in 2,512,371 records (see Supplementary Methods). To obtain information on author first names and country of institutional affiliation, 2,124,998 of these were matched to records in the citation database Web of Science (WoS) (see Supplementary Methods). This information was utilized to determine the gender of authors using the name-to-gender assignment algorithm Gender API<sup>38</sup> (see Supplementary Methods). For each first name and nation pair, Gender API provided an estimate specifying the certainty of the given name-to-gender assignment. To validate its accuracy, we conducted a manual quality control of 500 randomly selected authors from the dataset (see Supplementary Methods and Supplementary Table 10). Papers

lacking full first-name information for one or more authors were excluded from the analysis. Supplementary Fig. 2 details the data inclusion and exclusion steps, including the assignment of gender to author first names, leading to the final sample of 1,542,690 documents (61.4% of the total population). Supplementary Fig. 3 displays the distribution of  $f_w$ -scores as a function of the number of authors per paper in the dataset.

To identify studies using gender and sex analysis, we matched the WoS sample with the GenderMed Database<sup>36,37</sup>. GenderMed is an extensive bibliographic archive of the existing medical literature analysing gender and sex differences. Based on a screening of more than 13 million MEDLINE abstracts and 40,000 full-text manuscripts, approximately 13,000 studies, extending back to 1975, have been made available (see Supplementary Methods). GenderMed includes 4,830 studies with MeSH terms subordinate to a disease category for the period 2008-2015, of which 3,394 (70.3% of the population) matched with our WoS sample (see Supplementary Methods). A manual quality check of 500 GenderMed articles was carried out to verify the accuracy of the database in identifying research involving GSA (see Supplementary Methods).

The GenderMed database limits its scope to selected diseases that field experts have deemed epidemiologically relevant for GSA<sup>37</sup>. Thus, we excluded all studies in our final sample (N: 1,542,690) that did not overlap with studies in the GenderMed subsample (N: 3,394) with respect to disease-specific MeSH terms. This exclusion resulted in a reduced sample of 1,513,638 unique disease-specific papers, which was used in the logistic regression analyses (see Supplementary Methods).

**Outcome variable.** *GSA* is the binary outcome variable in models 1, 2 and 3. The variable is used to predict a study's likelihood of systematically using gender and sex analysis. Specifically, it dissociates the GenderMED sample from the remaining articles in the WoS-database (Non-GSA=0, GSA=1).

**Main predictors and covariates.** For each article in the full data-set (N: 1,542,690), we computed the weighted indicator  $f_w$  to specify the general participation of women authors in a given study.  $f_w$  values range from 0 to 1, with values closer to 1 indicating a higher share of women in the author group (see Supplementary Methods). To account for within-group variations in women's (and men's) representation across countries, disease topics and medical subject areas, we also computed average  $f_w$  scores for the following variables: MeSH disease-terms ( $f_w$  MeSH), WoS subject categories classified based on journal information ( $f_w$  SC) and last-author country ( $f_w$  country).  $f_w$  MeSH was included to adjust for differences in the average gender composition of authors across disease-specific research topics, while  $f_w$  SC was constructed to account for covariation attributable to women's varying participation across medical research areas and specialties (See Supplementary Methods). The latter control variable is important, since a given disease-term may be addressed by researchers in different research areas and specialties. A complementary method using MeSH disease-terms to capture medical research specialties was examined but did not yield any improvements compared to the WoS subject-category based approach<sup>45</sup>.

Based on Gender API estimates, we also calculated gender indices for first and last authors ( $f_{\text{first}}$  and  $f_{\text{last}}$ ). Further, we computed average  $f_{\text{first}}$  and  $f_{\text{last}}$  scores for MeSH disease-terms ( $f_{\text{first}}$  MeSH and  $f_{\text{last}}$  MeSH), WoS subject-categories ( $f_{\text{first}}$  SC and  $f_{\text{last}}$  SC) and author country ( $f_{\text{first}}$  country and  $f_{\text{last}}$  country) (see Supplementary Methods). A supplementary analysis was carried out to estimate the prevalence of non-traditional author listings based on alphabetical order, as these may bias the odds-ratio estimations for the key predictors  $f_{\text{first}}$  and  $f_{\text{last}}$  in the analysis. We found the intentional use of alphabetized author listings to be very rare and unlikely to influence our results in any notable way (see Supplementary Methods).

Additional control variables were constructed to help prevent potential misspecifications in the logistic regression models. To capture the influence of cultural and socio-economic factors across area-specific contexts, we included 10 categorical variables based on geographical groupings (for specifications on country groupings see Supplementary Table 13). Further, author-country information was used to compute ranked variables capturing covariation attributable to national differences in overall level of gender equality and annual R&D-related health expenditure, but these variables were excluded from the final models due to trivial variance (see Supplementary Methods) (variable specifications are available in Supplementary Table 14).

**Statistical models.** All models were estimated in a Bayesian framework with weakly informed default prior distributions<sup>39</sup>. A Bayesian framework is transparent and enables flexible estimation of parameters based on prior information and data. Proper uncertainties in estimates are calculated providing intuitive interpretations and direct inferences of hypotheses conditioned on the data. A simple Bayesian logistic modelling approach to estimate posterior distributions of the coefficients was applied. A basic generic binominal model with a logit link function was used for all specifications (Equation 1).

(1)

$$y_t \sim \text{Bin}(n, p_t), \quad p_t = \text{logit}^{-1}(\pi_t),$$

$$\pi_t = \beta_0 + \sum_{i=1}^k \beta_i$$

Following Gelman et al.<sup>39</sup>, we assigned default independent Cauchy prior distributions with location parameter 0 ( $\mu$ ) and scale parameter 2.5 ( $\sigma$ ), to all coefficients  $\beta_0, \beta_i, i = 1, \dots, k$ .

This prior is proposed as a default choice because it is weakly informative, but allows inferences to be made even in the presence of complete separation. Given the size of the data



sets used in models 1, 2 and 3, the expected influence of the weakly informed priors on the estimated posterior parameters will be less pronounced. The posterior distributions were estimated empirically by Markov Chain Monte Carlo (MCMC) techniques using the R package “MCMCpack”<sup>46,47</sup>. All models used 5,000 burn-in and 50,000 Metropolis iterations for the sampler with a tuning parameter of 0.25. To summarize and visualize findings from the three models, we estimated marginal means for the main predictors using the “lsmeans” package in R<sup>48</sup>. The descriptive analysis of gender disparities in authorship was based on the full sample of 1,542,690 papers, while the reduced sample of 1,513,638 was employed in the regression models.

**Data availability:** All data that support our main findings are publicly available at:

**<https://osf.io/r6hd9/>**

**Code availability:** The code-script for the logistic regression analysis and estimated marginal means is available at: <https://osf.io/r6hd9/>

**Correspondence and requests for materials** should be addressed to M.W.N.

**Competing interests:** The authors declare no competing interests.

**Acknowledgements:** We kindly thank Sabine Oertelt-Prigione and the Institute of Gender in Medicine, Charité-Universitätsmedizin Berlin, for data acquisition from the GenderMed database.

**Author contributions:** M.W.N. designed research; J.P.A and M.W.N. constructed the database; M.W.N., J.W.S. and J.P.A. analysed data; M.W.N., J.W.S., L.S. and J.P.A. wrote the paper.

**Conflict of interest:** The authors declare no conflict of interest.

## References

1. Arnold A.P. Promoting the understanding of sex differences to enhance equity and excellence in biomedical science. *Biol. Sex Differ.* **1**(1), 1 (2010).
2. Schiebinger, L., Leopold, S. & Miller, V.M. Editorial policies for sex and gender analysis. *Lancet* **388**(10062), 2841-2842 (2016).
3. International Committee of Medical Journal Editors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals [http://www.icmje.org/news-and-editorials/icmje-recommendations\\_annotated\\_dec16.pdf](http://www.icmje.org/news-and-editorials/icmje-recommendations_annotated_dec16.pdf) (2016).
4. Heidari, S., Babor, T.F., De Castro, P., Tort, S. & Curno, M. Sex and Gender Equity in Research: Rationale for the SAGER guidelines and recommended use. *Res. Integr. Peer Rev.* **1**(1), 2 (2016).
5. Miller V.M. In pursuit of scientific excellence: sex matters. *Physiol. Genomics* **44**(9), 485-486 (2012).
6. Nieuwenhoven, L. & Klinge, I. Scientific excellence in applying sex-and gender-sensitive methods in biomedical and health research. *J. Womens Health* **19**(2), 313-321 (2010).
7. Johnson, J.L., Greaves, L. & Repta, R. Better science with sex and gender: facilitating the use of a sex and gender-based analysis in health research. *Int. J. Equity Health* **8**(1), 1 (2009).
8. Oertelt-Prigione, S. & Regitz-Zagrosek, V. (eds.). *Sex and Gender Aspects in Clinical Medicine* (Springer, 2012).
9. Kim, E. S. H. & Menon, V. Status of women in cardiovascular clinical trials. *Arterioscler. Thromb. Vasc. Biol.* **29**, 279–83 (2009).
10. Mosca, L., Hammond, G., Mochari-Greenberger, H., Towfighi, A. & Albert, M.A. Fifteen-year trends in awareness of heart disease in women results of a 2012 American Heart Association National Survey. *Circulation* **127**(11), 1254-1263 (2013).
11. Kwiatkowski, K., Coe, K., Bailar, J.C. & Swanson, G.M. Inclusion of minorities and women in cancer clinical trials, a decade later: Have we improved? *Cancer* **119**(16), 2956-2963 (2013).
12. Beery A. K. & Zucker, I. Sex bias in neuroscience and biomedical research. *Neurosci. Biobehav. Rev.* **35**(3), 565-572 (2011).
13. Shah, K., McCormack, C. E. & Bradbury N. A. Do you know the sex of your cells? *Am. J. Physiol., Cell Physiol.* **306**(1), C3-C18 (2014).
14. Klein, S. L., Marks, M. A., Li, W., Glass, G. E., Fang, L. Q., Ma, J. Q., & Cao, W. C. Sex differences in the incidence and case fatality rates from hemorrhagic fever with renal syndrome in China, 2004–2008. *Clin. Infect. Dis.* **52**(12), 1414-1421 (2011).
15. Adler, R. A. Osteoporosis in men: a review. *Bone res.* **2**, 14001 (2014).
16. Smith, P.M. & Koehoorn, M. Measuring gender when you don't have a gender measure: constructing a gender index using survey data. *Int. J. Equity Health* **15**(1), 1-9 (2016).
17. Courtenay, W. H. Behavioral factors associated with disease, injury, and death among men: Evidence and implications for prevention. *J. Mens Stud.* **9**(1), 81-142 (2000).
18. Alabas, O.A., Tashani, O.A., Tabasam, G. & Johnson, M.I. Gender role affects experimental pain responses: a systematic review with meta-analysis. *Eur. J. Pain* **16**(9), 1211-1223 (2012).
19. Pelletier, R., Khan, N.A., Cox, J., Daskalopoulou, S.S., Eisenberg, M.J., Bacon, S.L., Lavoie, K.L., Daskupta, K., Rabi, D., Humphries, K.H., Norris, C.M., Thanassoulis, G., Behloul, H. & Pilote, L. Sex versus

- gender-related characteristics: which predicts outcome after acute coronary syndrome in the young? *J. Am. Coll. Cardiol.* **67**(2), 127-135 (2016).
20. Schiebinger, L. & Stefanick, M.L. Gender matters in biological research and medical practice. *J. Am. Coll. Cardiol.* **67**(2), 136-138 (2016).
21. U.S. General Accounting Office *Drug Safety: Most Drugs withdrawn in Recent Years had Greater Health Risks for Women* (Government Publishing Office, 2001).
22. Sex and gender analysis policies of major granting agencies, in Schiebinger, L., Klinge, I., Paik, H. Y., Sánchez de Madariaga, I., Schraudner, M., and Stefanick, M. (eds) (2011-2017). Gendered Innovations in Science, Health & Medicine, Engineering, and Environment <http://genderedinnovations.stanford.edu/sex-and-genderanalysis-policies-major-granting-agencies.html> (2017)
23. European Commission. Gender Equality in Horizon 2020, Version 1 [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/gender/h2020-hi-guide-gender\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/gender/h2020-hi-guide-gender_en.pdf) (2014).
24. National Institutes of Health. Consideration of sex as a biological variable in NIH-funded research. *National Institutes of Health* <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-102.html> (2015).
25. Ely, R.J. & Thomas, D.A. Cultural diversity at work: The effects of diversity perspectives on work group processes and outcomes. *Adm. Sci. Q.* **46**(2), 229–273 (2001).
26. Page, S.E. *The difference: How the power of diversity creates better groups, firms, schools, and societies*. (Princeton University Press, 2008).
27. Nielsen, M.W., Alegria, S., Börjeson, L., Etzkowitz, H., Falk-Krzesinski, H.J., Joshi, A., Leahey, E., Smith-Doerr, L., Wolley, A.W. & Schiebinger, L. Opinion: Gender diversity leads to better science. *Proc. Natl. Acad. Sci. U.S.A.* **114**(8), 1740-1742 (2017).
28. Charles, M. & Bradley, K. Indulging our gendered selves? Sex segregation by field of study in 44 countries. *Am. J. Sociol.* **114**(4), 924-976 (2009).
29. Alers, M., Leerdam, L.V., Dielissen, P., Lagro-Janssen, A. Gendered specialities during medical education: a literature review. *Perspect. Med. Educ.* **3**(3), 163-178 (2014).
30. West, J. D., Jacquet, J., King, M. M., Correll, S. J. & Bergstrom, C. T. The role of gender in scholarly authorship. *PloS one* **8**(7), e66212 (2013).
31. Light, R. in *Networks, Work, and Inequality* (Emerald Group Publishing, 2013).
32. Dolado, J. J., Felgueroso, F., & Almunia, M. Are men and women-economists evenly distributed across research fields? Some new empirical evidence. *SERIEs* **3**(3), 1-27 (2012).
33. Kretschmer, H., Kundra, R., Beaver, D.D. & Kretschmer, T. Gender bias in journals of gender studies. *Scientometrics* **93**(1), 135-150 (2012).
34. Söderlund, T. & Madison, G. Characteristics of gender studies publications: a bibliometric analysis based on a Swedish population database. *Scientometrics* **105**(3), 1347-1387 (2015).
35. Johnson, J., Sharman, Z., Vissandjee, B. & Stewart, D.E. Does a change in health research funding policy related to the integration of sex and gender have an impact? *PloS one* **9**(6), e99900 (2014).
36. Oertelt-Prigione, S., Parol, R., Krohn, S., Preißner, R. & Regitz-Zagrosek, V. Analysis of sex and gender-specific research reveals a common increase in publications and marked differences between disciplines. *BMC Med.* **8**(1), 70 (2010).
37. Oertelt-Prigione, S., Gohlke, B.O., Dunkel, M., Preissner, R. & Regitz-Zagrosek, V. GenderMedDB: an interactive database of sex and gender-specific medical literature. *Biol. Sex Differ.* **5**(1), 1 (2014).

38. Gender API. Gender API - Determines the gender of a first name <https://gender-api.com/> (2016).
39. Gelman, A., Jakulin, A., Pittau, M.G. & Su, Y.S. A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2**(4), 1360-1383 (2008).
40. Valantine, H.A. & Collins, F.S. National Institutes of Health addresses the science of diversity *Proc. Natl. Acad. Sci. U.S.A.* **112**(40), 12240-12242 (2015).
41. Gneezy, U., Niederle, M. & Rustichini, A. Performance in competitive environments: Gender differences. *Q. J. Econ.* **118**(3), 1049-1074 (2003).
42. Reskin, B.F. & Roos, P.A. *Job queues, gender queues: Explaining women's inroads into male occupations* (Temple University Press, 2009).
43. Patel, V.M., Ashrafian, H., Ahmed, K., Arora, S., Jiwan, S., Nicholson, J.K., Darzi, A. & Athanasiou, T. How has healthcare research performance been assessed? A systematic review. *J. R. Soc. Med.* **104**(6), 251-61 (2011).
44. Young, N.S., Ioannidis, J.P.A. & Al-Ubaydli, O. Why current publication practices may distort science. *PLoS Med.* **5**(10), e201 (2008).
45. Darmoni, S.J., Névél, A., Renard, J.M., Gehanno, J.F., Soualmia, L.F., Dahamna, B. & Thirion, B. A MEDLINE categorization algorithm. *BMC Med. Inform. Decis. Mak.* **6**(1), 7 (2006).
46. Martin, A.D., Quinn, K.M. & Hee, P.J. MCMCpack: Markov Chain Monte Carlo in R. *J. Stat. Softw.* **42**(9), 1-21 (2011).
47. Martin, A.D., Quinn, K.M. & Hee, P.J. Package: "MCMC-pack", version 1.3-9 <https://cran.r-project.org/web/packages/MCMCpack/MCMCpack.pdf> (2017)
48. Lenth, R. Package 'lsmeans', version 2.2. <https://cran.r-project.org/web/packages/lsmeans/lsmeans.pdf> (2016).

In the format provided by the authors and unedited.

# One and a half million medical papers reveal a link between author gender and attention to gender and sex analysis

Mathias Wullum **Nielsen**<sup>1\*</sup>, Jens Peter **Andersen**<sup>2</sup>, Londa **Schiebinger**<sup>1</sup> and Jesper W. **Schneider**<sup>2</sup>

<sup>1</sup>History of Science, Stanford University, Stanford, CA, USA. <sup>2</sup>Danish Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University, Aarhus, Denmark. \*e-mail: [mwn@ps.au.dk](mailto:mwn@ps.au.dk)

## **SUPPLEMENTARY INFORMATION**

<b>1. Supplementary Methods</b>	<b>p. 3</b>
<b>2. Supplementary Figures</b>	<b>p. 7</b>
<b>3. Supplementary Tables</b>	<b>p. 12</b>
<b>4. Supplementary References</b>	<b>p. 21</b>

## 1. Supplementary Methods

**Analysis concerning journal status.** We examined the distribution of GSA and non-GSA publications according to journal status. Journal status is operationalized as a journal citation indicator, where we measure the citation impact for the annual volume of research and review articles for a particular journal in the year it published the GSA and/or non-GSA study. We use two journal citation indicators, one unnormalised indicator (i.e. *js* = journal score) where citations are not adjusted to variations in citation rates across fields; and one normalised indicator (i.e. *njs* = normalised journal score). Both indicators are calculated with a two and a five-year citation window (i.e. *js\_2yr*, *js\_5yr*, *njs\_2yr*, *njs\_5yr*). Notice, the unnormalised journal score essentially corresponds to Clarivate Analytics' (formerly Thomson Reuters) Journal Impact Factor. We calculate medians and means for the GSA and non-GSA publications as an indicator of their general journal publication status. Notice, as strong outliers are present in both cases we focus on median status. As a comparison to the GSA-sample, we randomly selected first 4,000 and then 50,000 non-GSA publications and estimated their central tendencies. Supplementary Table 9 provides descriptive statistics, and Supplementary Fig. 1 presents boxplots of the random sample of 4,000 non-GSA publications (results are robust so we do not show a similar plot for the 50,000 sample).

**MeSH.** The MeSH thesaurus is a controlled, hierarchical indexing system curated by the National Institutes of Health. The vocabulary is freely accessible in a machine-readable format using the Entrez e-utils<sup>1</sup>. We used this option to harvest article metadata for all studies indexed with MeSH-terms subordinate to the "Diseases Category".

**Web of Science.** In contrast to PubMed, WoS includes full first-name information for the vast majority of articles registered since 2008. We used a modified version of the WoS database maintained by the CWTS (Center for Science and Technology Studies) at Leiden University to retrieve WoS records for our PubMed dataset. More specifically, WoS and PubMed records were matched using first DOI, then ISSN, journal names, pagination, volume and fuzzy title matches relying on relative Levenshtein distance. Due to query limitations, a matching approach using PMIDs in the WoS interface was not viable. Tests also indicated that merely 2.5% additional records would have been matched using this approach. Supplementary Fig. 2 provides an overview of the data inclusion and exclusion steps.

**Gender API.** Gender API determines the gender of a given individual based on first name and country of origin (Laurence, for instance, is a female name in France but a typical male name in the U.S.). Accordingly, Gender API supports gender assignment for 1,871,879 names from 178 countries. The underlying coding-script and determination process remains unspecified by the API developers, but a randomized test sample confirms its accuracy (more on this below). Due to data-limitations, *de facto* information on country of origin is not available to us. Thus, for the purpose of country specification, we used WoS information for the country of a given author's institutional affiliation. Full first-names were not available for all WoS documents (all papers with authors using first-name initials were excluded from the final data set), and the Gender API algorithm



failed to assign gender to rare first names, leaving a total of 1,542,690 papers for further analysis (See Supplementary Fig. 2).

Specifically, Gender API provides an accuracy score estimating a given first name's probability of belonging to a man or a woman, accounting for country. This prediction ranges from 50 to 100. We converted the Gender API accuracy scores for women and men into a single indicator specifying the probability of a name belonging to a woman, denoted  $f$ .  $f$ -scores range from 0 to 1 with values closer to 1 indicating a higher likelihood of the author being female.

### **Validation of Gender API Assignments.**

Following the approach of Larivière et al.<sup>2</sup>, we used WoS author information to search the web for biographical information, resumes or photos and other information that could confirm its estimations. Supplementary Table 10 provides an overview of the outcomes of this validation. The second column represents the outcomes of the Gender API algorithm. The three columns to the right represent the outcomes of the manual validation. Numbers marked in bold represent the share of “false positives”, i.e. female authors with a “male”  $f$ -score of  $<.41$ , or male authors with a “female”  $f$ -score of  $>.60$ . Thirteen author names (i.e. 0.3% of the total sample) were unknown to the algorithm and 27 (i.e. 0.5%) fell into the unisex category – here defined as author names with  $f$ -scores ranging from  $.41$ -. $60$ . Specifications on the countries and journals represented in the Gender-API validation sample are displayed in Supplementary Fig. 4 and 5.

**GenderMed database.** GenderMed combines systematic searches in PubMed with an Apache Lucene based text-mining algorithm to regularly screen MEDLINE for new studies adopting gender- and sex-based approaches. Manual quality control is continuously performed to ensure the accuracy of the tool (for further specifications on the database structure and methodology, see ref.<sup>3-5</sup>).

**Identification of studies involving gender and sex analysis.** In total, the GenMed database includes 5,185 gender- and sex-focused articles for the period 2008-2015. Of these, 4,830 are indexed with MeSH “Disease-categories”. The initial matching with the WoS records resulted in a data-set of 3,093 gender- and sex-based studies with author-gender specifications. Yet, after a manual hand-coding of articles with only first-name initials and unreadable author-names due to formatting issues, this number increased to 3,394 (70.3% coverage).

**Validation of GenderMed data.** We randomly selected 500 of the 3,394 GenderMed studies included in the final sample. We carefully read the abstract (and in cases of doubt also the full text) of each study, to determine its eligibility for inclusion in GenderMed. We used the original inclusion and exclusion criteria laid out by Oertelt and colleagues<sup>4</sup>: [inclusion  $a$ ] The study includes “a description of sex/gender-specific differences in the analysed species (human, mouse, rat and so on)”. [inclusion  $b$ ] The study presents an “analysis of data with respect to sex/gender-specific differences”. [exclusion  $a$ ] The study does not include sex/gender-specific description and analysis of results. [exclusion  $b$ ] The study presents “generalized statements without descriptions

of performance analysis, for example ‘no gender differences were found’’. [exclusion *c*] The study refers “to the analysed condition (for example, “hypertension”) only as co-morbidity, confounder, or anamnestic finding”. 488 papers (i.e. 97.6% of the test sample) were deemed eligible for inclusion in GenderMed. The remaining 12 studies did not present an analysis of data with respect to sex/gender differences. Yet many of them presented themes closely related to GSA, such as: pharmacokinetic parameters for pregnant women; gender-specific HIV prevention interventions targeting African American men; condom use and self-perceived HIV risk among female sex workers; and gender-specific health promotion targeting African American men.

**Excluding disease-specific topics not covered by the GenderMed database.** We identified all disease-specific MeSH terms (and their related lower-level terms in the MeSH tree) addressed by at least one of the 3,394 studies in the GenderMed subsample. The full sample (N: 1,542,690) covered 4,709 unique disease-specific terms of which 4,341 were assigned to studies in the GenderMed subsample. We excluded all studies that did not have at least one disease-specific MeSH term (or a related lower-level term) in common with a study in the GenderMed subsample.

**Predictors in the logistic regressions.** Gender API provides an estimate of its accuracy in determining the gender of any name and country pair. We convert this estimate into a probability of a name belonging to a female or a male researcher and denote this probability  $f$ .  $f$  is used to compute the weighted indicator,  $fw$ , which is the mean  $f$ -value for the author group as a whole. Due to the uncertainty associated with the Gender API classification, the use of this indicator is only meaningful at the aggregate level. A paper with  $fw = 0.8$  could, for instance, be authored by two women, while another all-male paper might have  $fw = 0.2$ . We also calculate specific  $f$ -scores for first and last authors, denoted  $f_{first}$  and  $f_{last}$ . MeSH disease-terms derived from MEDLINE’S thesaurus are used as a proxy for disease-specific topics. WoS subject categories are used to capture the broader research areas or medical specialties in which a given study is carried out. Specifically, WoS classifies papers into subject categories (e.g. oncology, ophthalmology and orthopaedics) based on journal information. Medical studies are typically indexed under more than one MeSH disease-term and WoS subject-category. Thus, our calculations of  $fw$  MeSH and  $fw$  SC are based on weighted averages. Specifically, we calculated the average  $fw$ -score for all papers with a given MeSH disease-term or SC-category assigned to them. For a given study with three MeSH-terms, we weighted each term by  $\frac{1}{3}$  in the calculation of  $fw$  MeSH. The same procedure was used for  $fw$  SC.  $fw$  country was calculated as the average  $fw$ -score per last-author country.  $f_{first}$  MeSH,  $f_{last}$  MeSH,  $f_{first}$  SC,  $f_{last}$  SC,  $f_{first}$  country and  $f_{last}$  country were computed using the same approach.

**Covariates in the logistic regressions.** Since last authors typically take the lead in identifying topics and developing research questions, the coding of the 10 geographical variables (**Arab states**, **East Asia**, **Oceania** etc.), was based on the last-authors’ country affiliation. We constructed the gender equality indicator, **GE Index**, based on the United Nations’ ranking of countries in the Gender Development Index (for specifications on rank see ref.<sup>6</sup>). The ranked variable capturing

Domestic R&D Health expenditure as percentage of GDP (*Health\_Expd*) was computed using data from WHO<sup>7</sup>.

**Listing of authors.** A custom-made algorithm was built to examine the prevalence of author listings based on alphabetical order. The algorithm compared the actual listing of authors in each document in the dataset to an alphabetical order. Only papers with at least four authors were included in the analysis, as the chance of randomly occurring alphabetization is too high for smaller numbers of authors. A pseudo-code representation of the algorithm is shown below.

```

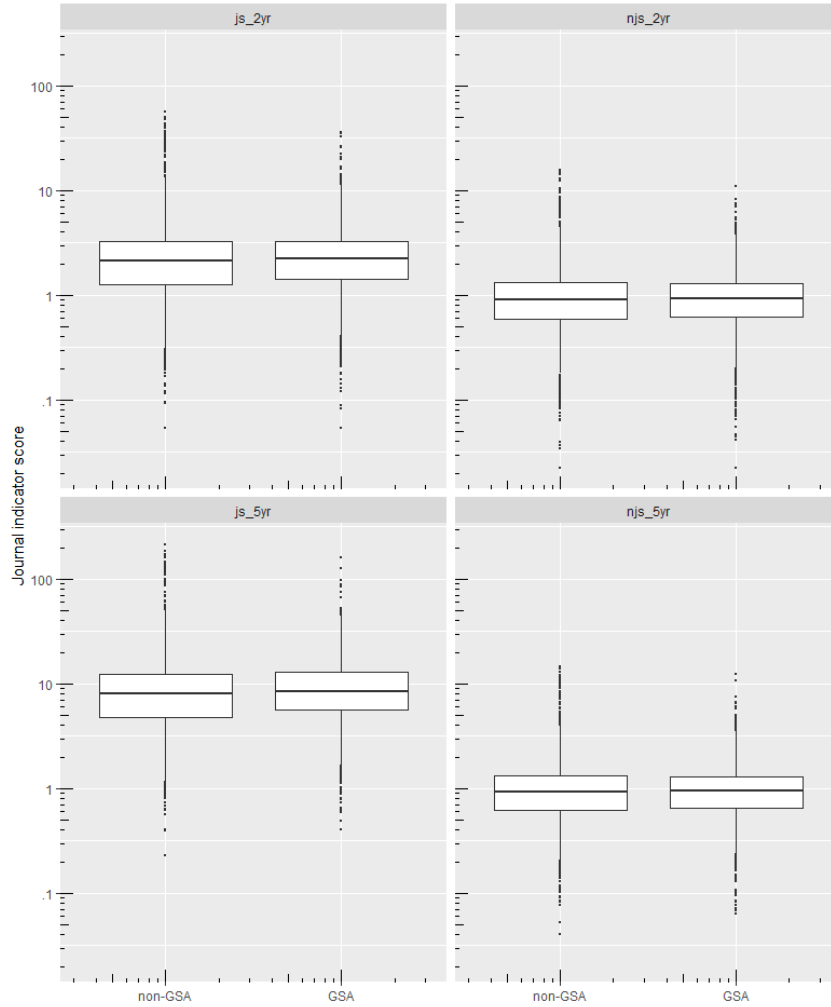
    alphabetization = true;
    x = 0;
    select author_order_nums from document, order by author_last_name ascending, first name
    ascending;
    for author_order_num in author_order_nums
        if author_order_num < x
            alphabetization = false;
            break
        else
            x = author_order_num
    return alphabetization

```

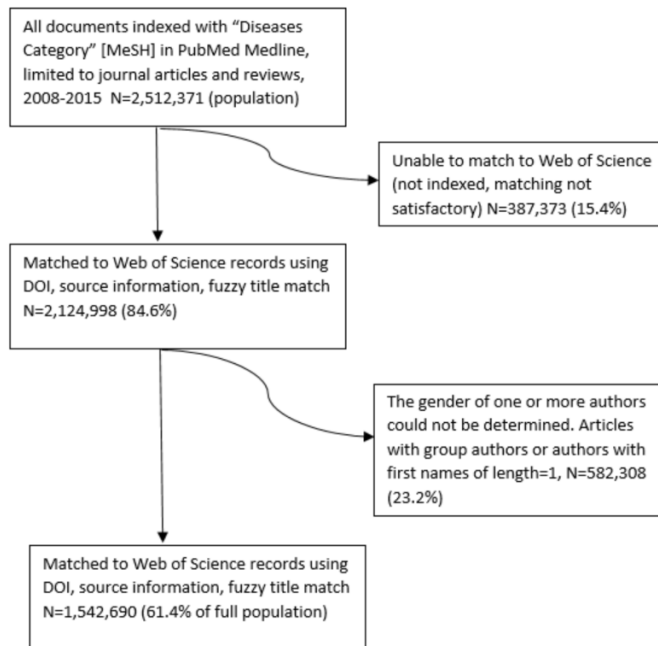
Supplementary Table 11 presents the results of the analysis. **n\_authors** refers to the number of authors per paper, **p\_total** specifies the total number of papers per n\_authors value, **p\_alpha** gives the number of papers using alphabetical order per n\_authors value and **pp\_alpha** shows the percentage share of documents using alphabetical order per n\_authors value. **chance** specifies the expected occurrence of alphabetical order by chance. Specifically, we calculated all possible permutations for each n\_author value, assuming one of these was ordered alphabetically (e.g. n\_authors: 4, possible permutations: 24, chance: 1/24= 4.17%). Since last names starting with a letter in the beginning of the alphabet may be more likely to occur in the database, this approach is not entirely accurate, but it enables us to roughly estimate the proportion of intentionally alphabetized author listings, which is presented in the column **int\_alpha**. We calculated this proportion by subtracting chance from pp\_alpha for each n\_authors value. As displayed in the int\_alpha column the prevalence of intentionally alphabetized author listings is limited to between 0.08% and .24% of the papers with 4+ authors in our sample.

Traditionally, authors in health economics have been known to follow conventions in economics and list authors based on alphabetical order. Since the typical number of authors per paper in health economics may be smaller than four, we carried out a specific analysis for this subfield. Supplementary Table 12 specifies the prevalence of alphabetized author listings for seven core-field journals in health economics with at least ten disease-specific studies included in the dataset. Two journals, *Journal of health economics* and *Health economics*, have relatively high pp\_alpha values, while values are low for the remaining journals. At the aggregate level, the intentional use of alphabetized author listings appears to be rare in disease-specific studies in health economics, and we have therefore chosen not to exclude papers published in this subfield from our analysis.

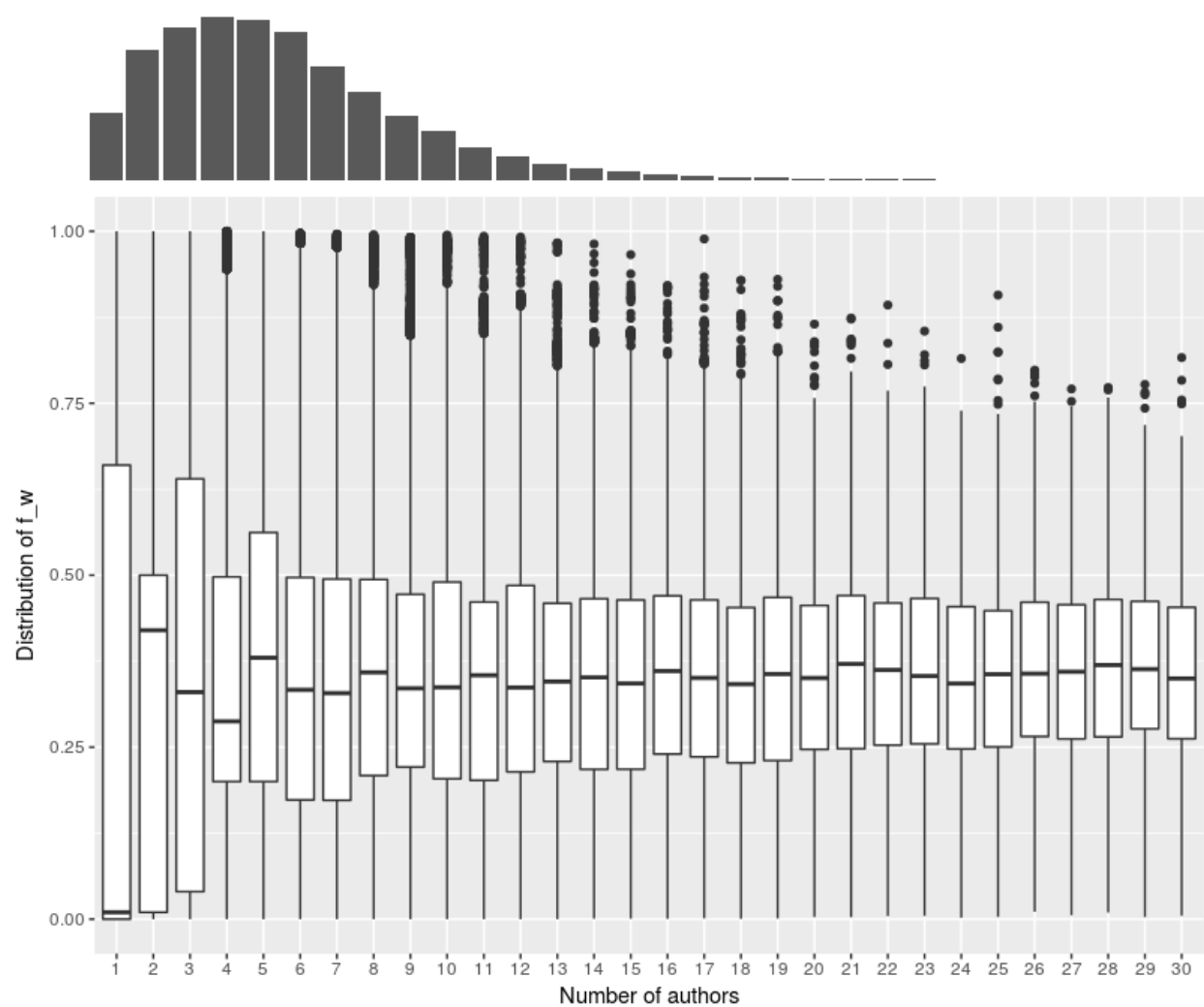
## 2. Supplementary Figures



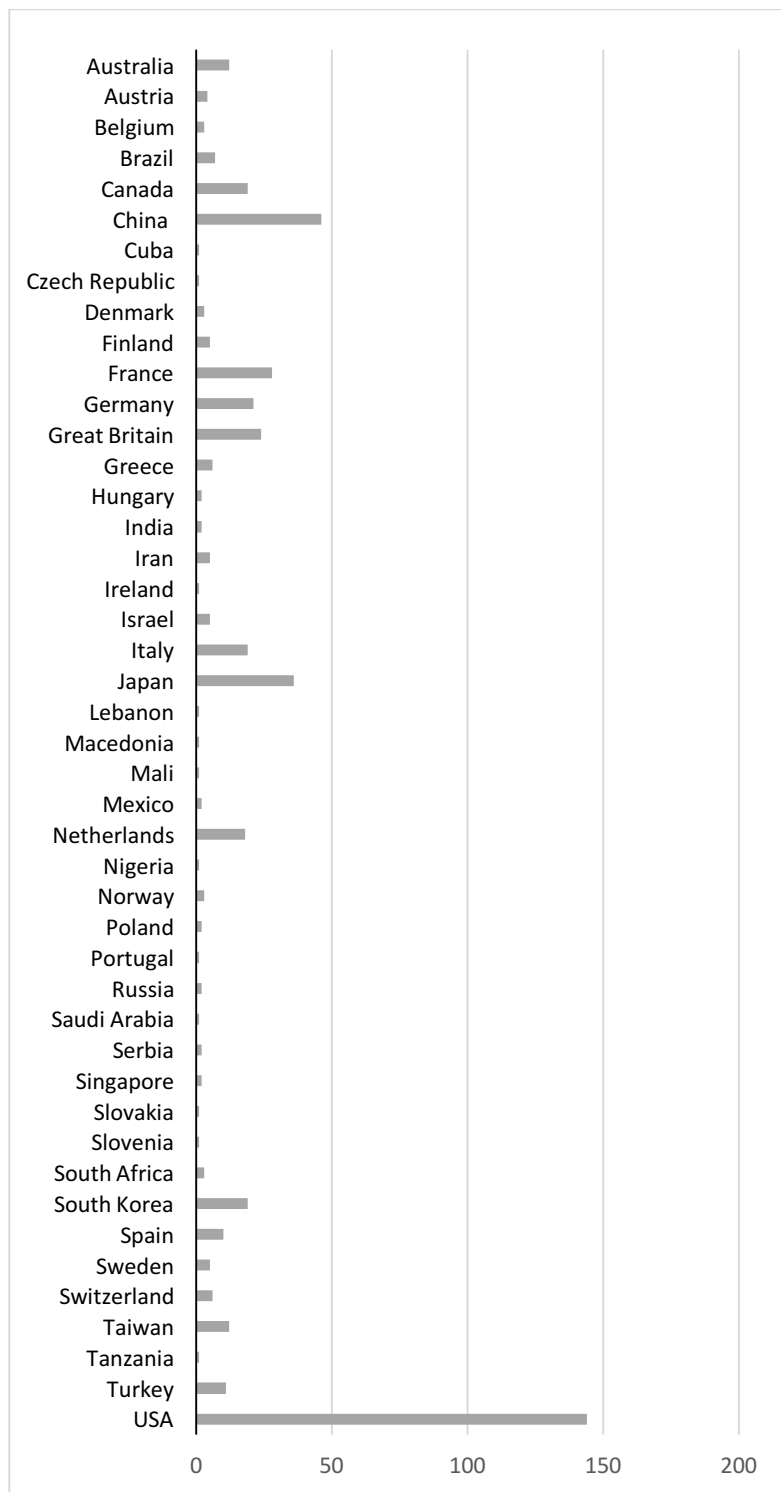
**Supplementary Figure 1.** Boxplot of aggregate journal-publication status for GSA and non-GSA publications. Journal status is approximated with journal citation indicators, two unnormalised (*js*) and two normalised indicators (*njs*). Both sets of journal indicators are calculated with two and five year citation windows. Journal indicator scores on the y-axis are shown on a log-scale to enable inclusion of outliers.



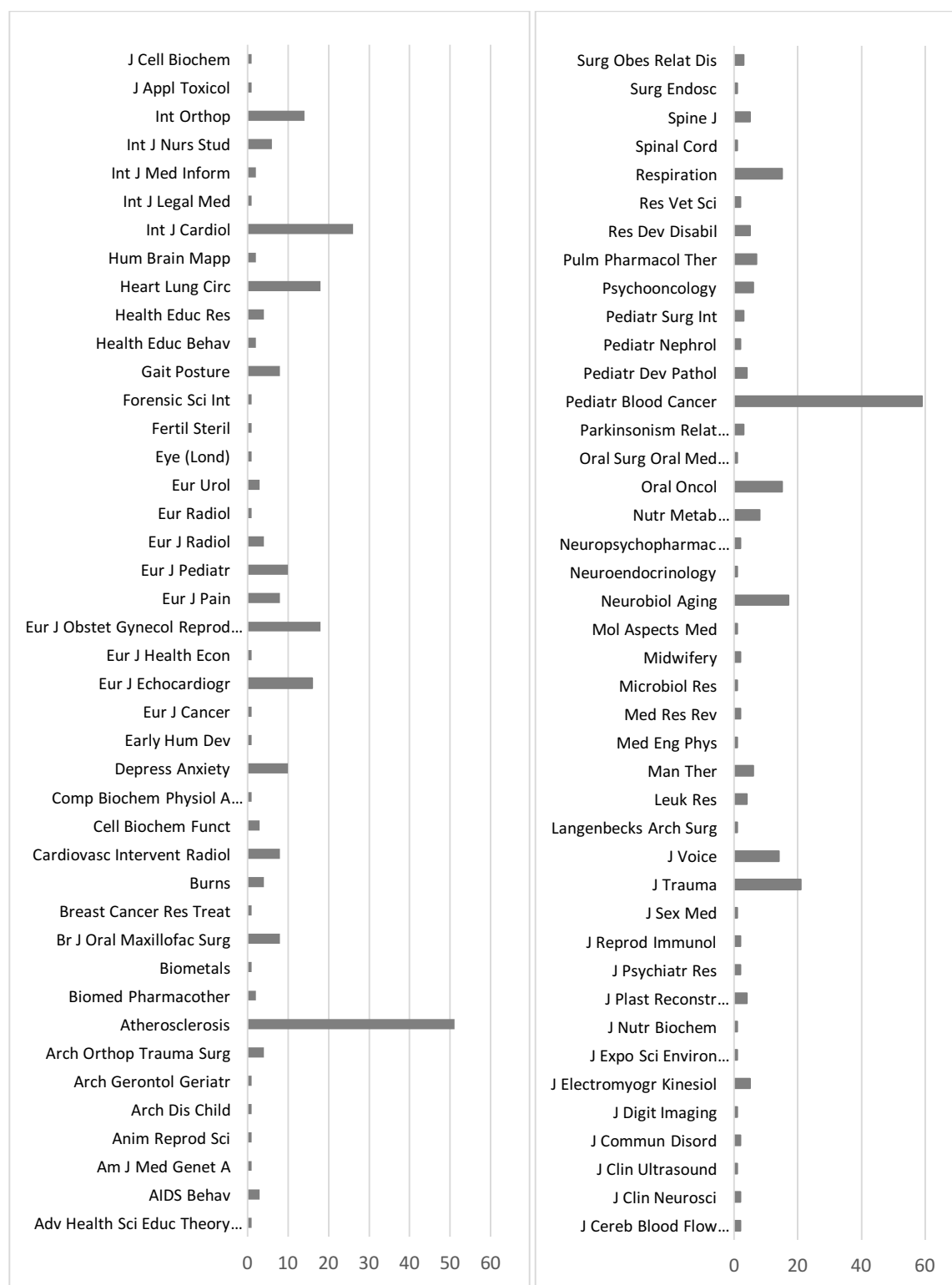
**Supplementary Figure 2.** Flowchart of data inclusion and exclusion.



**Supplementary Figure 3.** Boxplots of  $f_w$  as a function of the number of authors per paper. Each boxplot displays the median, interquartile ranges and 95% ranges plus outliers as individual dots. The horizontal histogram specifies the distribution of papers across the x-axis. The figure has been set to cut off at 30 authors per paper, but the actual data involves cases with up to 256 authors per paper.



**Supplementary Figure 4.** Number of authors per country in Gender API validation sample.



**Supplementary Figure 5.** Number of authors per journal in Gender API validation sample.



### 3. Supplementary Tables

**Supplementary Table 1.** Women's participation as authors in geographical groupings

	<i>Women first authors</i>	<i>Women last authors</i>
Arab states	.35	.30
East Asia	.24	.16
CW Independent states	.47	.34
Latin America	.52	.40
Oceania	.49	.32
South & West Asia	.33	.27
So-Centr. & East. Europe	.44	.33
Sub-Saharan Africa	.36	.30
North America	.41	.28
Western Europe	.42	.25
N: 1,542,690		

**Supplementary Table 2.** Model 1: Binary logistic regression model predicting GSA (first author)

Parameter	Mean	SD	95% Credible Intervals	Odds Ratio	SD	95% Credible Intervals
Intercept	-8.40	0.17	[-8.70:-8.04]			
<i>f_first</i>	0.50	0.04	[0.43:0.58]	1.66	0.07	[1.53:1.79]
<i>f_first country</i>	0.97	0.29	[0.40:1.53]	2.75	0.80	[1.50:4.62]
<i>f_first MeSH</i>	1.78	0.26	[1.27:2.30]	6.13	1.67	[3.55:10.00]
<i>f_first SC</i>	0.73	0.22	[0.31:1.16]	2.11	0.47	[1.36:3.19]
Arab States	0.82	0.20	[0.44:1.21]	2.31	0.46	[1.55:3.34]
East Asia	0.57	0.12	[0.33:0.82]	1.79	0.22	[1.40:2.26]
Latin America	0.16	0.16	[-0.16:0.47]	1.19	0.19	[0.85:1.60]
Oceania	0.21	0.15	[-0.07:0.50]	1.25	0.18	[0.93:1.66]
South & West Asia	0.22	0.15	[-0.09:0.55]	1.27	0.21	[0.91:1.73]
South-Central & Eastern Europe	0.30	0.15	[0.01:0.59]	1.36	0.20	[1.00:1.80]
Sub-Saharan Africa	1.10	0.19	[0.70:1.45]	3.03	0.58	[2.01:4.28]
North America	0.73	0.10	[0.53:0.94]	2.08	0.22	[1.69:2.56]
Western Europe	0.69	0.11	[0.48:0.90]	2.00	0.22	[1.62:2.46]
N	1,513,638					

Note: Posterior summaries and 95% credible intervals for Bayesian logistic regression model with Cauchy informative priors predicting GSA (0 = non-GSA, 1 = GSA). Commonwealth Independent States is the reference group for the geographical variables.

**Supplementary Table 3.** Model 2: Binary logistic regression model predicting GSA (last author)

Parameter	Mean	SD	95% CI	Odds Ratio	SD	95% CI
Intercept	-8.19	0.16	[-8.51:-7.88]			
<i>f_last</i>	0.44	0.04	[0.36:0.52]	1.56	0.06	[1.44:1.68]
<i>f_last country</i>	1.34	0.32	[0.71:1.96]	4.03	1.31	[2.02:7.13]
<i>f_last MeSH</i>	1.51	0.31	[0.96:2.19]	4.76	1.55	[2.48:8.49]
<i>f_last SC</i>	1.43	0.25	[0.94:1.92]	4.32	1.09	[2.60:6.84]
Arab States	0.86	0.20	[0.48:1.25]	2.41	0.48	[1.62:3.48]
East Asia	0.67	0.13	[0.43:0.92]	1.96	0.25	[1.54:2.49]
Latin America	0.26	0.15	[-0.05:0.56]	1.31	0.20	[0.95:1.76]
Oceania	0.39	0.14	[0.11:0.68]	1.49	0.22	[1.11:1.97]
South & West Asia	0.26	0.15	[-0.04:0.60]	1.33	0.23	[0.96:1.83]
South-Central & Eastern Europe	0.39	0.15	[0.11:0.68]	1.50	0.21	[1.10:1.97]
Sub-Saharan Africa	1.16	0.19	[0.78:1.53]	3.26	0.62	[2.19:4.61]
North America	0.86	0.10	[0.66:1.06]	2.37	0.25	[1.93:2.90]
Western Europe	0.92	0.11	[0.72:1.14]	2.53	0.28	[2.05:3.12]
N	1,513,638					

Note: Posterior summaries and 95% credible intervals for Bayesian logistic regression model with Cauchy informative priors predicting GSA (0 = non-GSA, 1 = GSA). Commonwealth Independent States is the reference group for the geographical variables.

**Supplementary Table 4.** Model 3: Binary logistic regression model predicting GSA (full author group)

Parameter	Mean	SD	95% CI	Odds Ratio	SD	95% CI
Intercept	-8.25	0.19	[-8.63:-7.89]			
<i>fw</i>	1.14	0.07	[1.01:1.28]	3.14	0.22	[2.74:3.59]
<i>fw</i> country	0.71	0.37	[-0.01:1.45]	2.19	0.85	[0.99:4.26]
<i>fw MeSH</i>	1.53	0.31	[0.93:2.15]	4.86	1.58	[2.55:8.58]
<i>fw SC</i>	0.65	0.26	[0.15:1.16]	1.98	0.52	[1.16:3.18]
Arab States	0.79	0.20	[0.40:1.18]	2.26	0.45	[1.50:3.26]
East Asia	0.63	0.13	[0.39:0.89]	1.90	0.25	[1.47:2.44]
Latin America	0.17	0.16	[-0.15:0.48]	1.20	0.19	[0.86:1.62]
Oceania	0.27	0.15	[-0.01:0.57]	1.33	0.20	[1.00:1.75]
South & West Asia	0.24	0.17	[-0.08:0.57]	1.29	0.22	[0.99:1.77]
South-Central & Eastern Europe	0.29	0.15	[0.00:0.57]	1.35	0.20	[0.92:1.77]
Sub-Saharan Africa	1.14	0.19	[0.76:1.50]	3.17	0.60	[2.13:4.48]
North America	0.75	0.10	[0.45:0.96]	2.12	0.22	[1.72:2.60]
Western Europe	0.76	0.11	[0.56:0.97]	2.16	0.23	[1.75:2.65]
N	1,513,638					

Note: Posterior summaries and 95% credible intervals for Bayesian logistic regression model with Cauchy informative priors predicting GSA (0 = non-GSA, 1 = GSA). Commonwealth Independent States is the reference group for the geographical variables.

**Supplementary Table 5.** Estimated marginal means for main predictors in logistic regression models 1 to 3

	Non-GSA			GSA		
	EM-mean	SE	95% CI	EM-mean	SE	95% CI
<i>f_first</i>	0.40	0.0003	[0.40:0.40]	0.49	0.0073	[0.48:0.51]
<i>f_last</i>	0.27	0.0003	[0.27:0.27]	0.35	0.0067	[0.33:0.36]
<i>fw</i>	0.35	0.0001	[0.35 0.35]	0.42	0.0040	[0.41:0.43]
N	1,513,638					

Note: Estimated marginal means, standard errors and 95% credible intervals for the main predictors in the logistic regression models 1-3.

**Supplementary Table 6.** Binary logistic regression model predicting GSA (first author)

Parameter	Mean	SD	95% Credible Intervals	Odds Ratio	SD	95% Credible Intervals
Intercept	-6.40	0.03	[-6.45:-6.35]			
<i>f_first</i>	0.65	0.04	[0.57:0.72]	1.91	0.07	[1.77:2.06]
N	1,513,638					

Note: Posterior summaries and 95% credible intervals for Bayesian logistic regression model with Cauchy informative priors predicting GSA (0 = non-GSA, 1 = GSA).

**Supplementary Table 7.** Binary logistic regression model predicting GSA (last author)

Parameter	Mean	SD	95% Credible Intervals	Odds Ratio	SD	95% Credible Intervals
Intercept	-6.28	0.02	[-6.32:-6.23]			
<i>f_last</i>	0.56	0.04	[0.49:0.64]	1.76	0.07	[1.63:1.89]
N	1,513,638					

Note: Posterior summaries and 95% credible intervals for Bayesian logistic regression model with Cauchy informative priors predicting GSA (0 = non-GSA, 1 = GSA).

**Supplementary Table 8.** Binary logistic regression model predicting GSA (full author group)

Parameter	Mean	SD	95% Credible Intervals	Odds Ratio	SD	95% Credible Intervals
Intercept	-6.64	0.03	[-6.70:-6.57]			
<i>f_w</i>	1.350	0.06	[1.23:1.472]	3.87	0.24	[3.43:4.36]
N	1,513,638					

Note: Posterior summaries and 95% credible intervals for Bayesian logistic regression model with Cauchy informative priors predicting GSA (0 = non-GSA, 1 = GSA).

**Supplementary Table 9.** Descriptive statistics for aggregated journal-publication status

	<i>js_2yr</i>	<i>njs_2yr</i>	<i>js_5yr</i>	<i>njs_5yr</i>
GSA, n = 3242				
1 <sup>st</sup> quartile	1.4	0.6	5.6	10.7
Median	2.3	0.9	8.5	1.0
Mean	2.8	1.1	10.8	1.1
3 <sup>rd</sup> quartile	3.3	1.3	12.9	1.3
non-GSA, n = 4000				
1 <sup>st</sup> quartile	1.3	0.6	4.8	0.6
Median	2.1	0.9	7.9	0.9
Mean	2.8	1.1	10.5	1.1
3 <sup>rd</sup> quartile	3.3	1.3	12.5	1.3
non-GSA, n = 50000				
1 <sup>st</sup> quartile	1.3	0.6	4.8	0.6
Median	2.1	0.9	7.9	0.9
Mean	2.9	1.1	10.8	1.1
3 <sup>rd</sup> quartile	3.3	1.3	12.4	1.3

**Supplementary Table 10.** Validation of Gender API estimates

Category	Gender API # and % of sample	Identified	Female (identified) # and % of identified	Male (identified) # and % of identified
Unknown	13 (.3%)	6	0 (0%)	6 (100%)
Female <i>f</i> > .60	154 (31%)	105	98 (93%)	<b>7 (7%)</b>
Unisex <i>f</i> .41-.60	27 (.5%)	12	5 (42%)	7 (58%)
Male <i>f</i> < .41	306 (61%)	230	<b>5 (2%)</b>	225 (98%)
Total	500	353	108	245

Note: False positives (i.e. wrong API estimations marked in bold).

**Supplementary Table 11.** Prevalence of intentionally alphabetized author listings in the dataset

n_authors	p_total	p_alpha	pp_alpha	chance	int_alpha
4	201074	8595	4.27%	4.17%	.10%
5	196662	2110	1.07%	.83%	.24%
6	180965	508	0.28%	.14%	.14%
7	139403	169	0.12%	.02%	.10%
8	108379	92	0.08%	.00%	.08%
9	78368	66	0.08%	.00%	.08%
10	60195	59	0.10%	.00%	.10%

**Supplementary Table 12.** Prevalence of alphabetized author listings in health economics journals (papers with 4+ authors)

Journal	p_total	p_alpha	pp_alpha
Journal of health economics	18	5	27.7%
Health economics	43	6	14.0%
PharmacoEconomics	164	5	3.0%
Value in health	459	4	.90%
The European journal of health economics	87	1	1.1%
Expert review of pharmacoeconomics & outcomes research	111	1	.90%
Journal of medical economics	25	0	0%
Total	907	22	2.4%

**Supplementary Table 13.** Geographical groups

**Arab States**

Algeria  
Egypt  
Jordan  
Kuwait  
Lebanon  
Oman  
Qatar  
Saudi Arabia  
Syria  
Tunisia  
United Arab Emirates  
Morocco

**East Asia**

Brunei  
Cambodia  
China  
Indonesia  
Japan  
Malaysia  
Myanmar  
Philippines  
South Korea  
Singapore  
Mongol Peoples Republic  
Taiwan  
Thailand  
Vietnam

**Commonwealth Independent States**

Kazakhstan  
Kyrgyzstan  
Republic of Georgia  
Russia  
Tajikistan  
Uzbekistan

**Latin America**

Argentina  
Bahamas  
Belize  
Bolivia  
Brazil  
Chile  
Columbia  
Costa Rica  
Cuba  
Dominican Republic  
Ecuador  
El Salvador  
French Guyana  
Guyana

Haiti  
Honduras  
Jamaica  
Mexico  
Nicaragua  
Panama  
Paraguay  
Peru  
St. Lucia  
Trinidad & Tobago  
Uruguay  
Venezuela

**North America**

United States of America  
Canada

**Oceania**

Australia  
Micronesia  
Fiji Islands  
New Caledonia  
New Zealand  
Solomon Islands

**South and West Asia**

Afghanistan  
Bangladesh  
Bhutan  
India  
Iran  
Nepal  
Pakistan  
Sri Lanka

**South-Central and Eastern Europe**

Albania  
Bosnia Hercegovina  
Bulgaria  
Croatia  
Czech Republic  
Estonia  
Hungary  
Latvia  
Lithuania  
Macedonia  
Poland  
Romania  
Serbia  
Slovakia  
Slovenia  
Turkey  
Ukraine

**Sub-Saharan Africa**

Benin  
Botswana  
Burkina Faso  
Burundi  
Cameroon  
Congo  
Eritrea  
Ethiopia  
Ghana  
Ivory Coast  
Kenya  
Lesotho  
Malawi  
Mali  
Mauritius  
Mozambique  
Namibia  
Niger  
Reunion  
Rwanda  
Senegal  
Seychelles  
Sierra Leone  
South Africa  
Swaziland  
Togo  
Tanzania  
Uganda  
Zambia  
Zimbabwe

**Western Europe**

Austria  
Belgium  
Cyprus  
Denmark  
Finland  
France  
Germany  
Great Britain  
Greece  
Iceland  
Ireland  
Israel  
Italy  
Luxembourg  
Malta  
Monaco  
Netherlands  
Norway  
Portugal  
Spain  
Sweden  
Switzerland



**Supplementary Table 14.** Variable specifications

<b>Outcome (y)</b>		<b>Explanation</b>	<b>Measurement type</b>
	<b>GSA</b>	Dissociates studies that do and do not involve a gender and/or sex-focused component. (Non-GSA =0, GSA=1)	Binary (0, 1)
<b>Predictors (x)</b>			
	<b>f_first</b>	The probability of a first-author name belonging to a female researcher.	Continuous (0 – 1)
	<b>f_first country</b>	Average <b>f_first</b> score for country-affiliation of first author	Continuous (0 – 1)
	<b>f_first MeSH</b>	Weighted average <b>f_first</b> score for MeSH disease-terms assigned to a study.	Continuous (0 – 1)
	<b>f_first SC</b>	Weighted average <b>f_first</b> score for WoS SC-categories assigned to a study.	Continuous (0 – 1)
	<b>f_last</b>	The probability of a last-author name belonging to a female researcher.	Continuous (0 – 1)
	<b>f_last country</b>	Average <b>f_last</b> score for country-affiliation of last author	Continuous (0 – 1)
	<b>f_last MeSH</b>	Weighted average <b>f_last</b> score for MeSH disease-terms assigned to a study.	Continuous (0 – 1)
	<b>f_last SC</b>	Weighted average <b>f_first</b> score for WoS SC-categories assigned to a study.	Continuous (0 – 1)
	<b>fw</b>	Values closer to 1 indicate a higher share of women in the author group.	Continuous (0 – 1)
	<b>fw country</b>	Average <b>fw</b> score for country-affiliation of last author	Continuous (0 – 1)
	<b>fw MeSH</b>	Weighted average <b>fw</b> score for MeSH disease-terms assigned to a study.	Continuous (0 – 1)
	<b>fw SC</b>	Weighted average <b>f_first</b> score for WoS SC-categories assigned to a study.	Continuous (0 – 1)
<b>Covariates (x)</b>			
	<b>Arab states</b>	Geographical location of last author's institutional affiliation (1=Arab states)	Binary (0, 1)
	<b>East Asia</b>	Geographical location of last author's institutional affiliation (1=East Asia)	Binary (0, 1)
	<b>Commonwealth Independent States</b>	Geographical location of last author's institutional affiliation (1=Commonwealth Independent States)	Binary (0, 1)
	<b>Latin America</b>	Geographical location of last author's institutional affiliation (1=Latin America)	Binary (0, 1)
	<b>Oceania</b>	Geographical location of last author's institutional affiliation (1=Oceania)	Binary (0, 1)
	<b>South &amp; West Asia</b>	Geographical location of last author's institutional affiliation (1=South & West Asia)	Binary (0, 1)
	<b>South-Central Eastern Europe</b>	Geographical location of last author's institutional affiliation (1=Eastern Europe)	Binary (0, 1)
	<b>Sub-Saharan Africa</b>	Geographical location of last author's institutional affiliation (1=Sub-Saharan Africa)	Binary (0, 1)
	<b>North America</b>	Geographical location of last author's institutional affiliation (1=North America)	Binary (0, 1)
	<b>Western Europe</b>	Geographical location of last author's institutional affiliation (1=Western Europe)	Binary (0, 1)

#### 4. Supplementary References

1. National Center for Biotechnology Information. Entrez programming utilities help <https://www.ncbi.nlm.nih.gov/books/NBK25501/> (2010)
2. Lariviere V., Ni, C., Gingras, Y. & Cronin, B., Sugimoto, C. R. Global gender disparities in science. *Nature* **504**(7479), 211-213 (2013).
3. GenderMed. GenderMed Database and the Pilot Project Gender in Medicine <http://gendermeddb.charite.de> (2016).
4. Oertelt-Prigione, S., Parol, R., Krohn, S., Preißner, R. & Regitz-Zagrosek, V. Analysis of sex and gender-specific research reveals a common increase in publications and marked differences between disciplines. *BMC Med.* **8**(1), 70 (2010).
5. Oertelt-Prigione, S., Gohlke, B.O., Dunkel, M., Preissner, R. & Regitz-Zagrosek, V. GenderMedDB: an interactive database of sex and gender-specific medical literature. *Biol. Sex Differ.* **5**(1), 1 (2014).
6. UNDP (United Nations Development Programme). Table 4: Gender Development Index (GDI) <http://hdr.undp.org/en/composite/GDI> (2015)
7. World Health Organization. World Health Statistics 2016: Monitoring Health for the SDGs (WHO, 2016).