nature human behaviour

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41562-017-0235-x

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[1]*, Jens Peter Andersen[2], Londa Schiebinger[1] and Jesper W. Schneider[2]

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

SUPPLEMENTARY INFORMATION

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[1]. 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 Levenshstein 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.[2], 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 >0. 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.[3-5]).

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[4]: [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 ⅓ 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.[6]). The ranked variable capturing

Domestic R&D Health expenditure as percentage of GDP (***Health_Expd***) was computed using data from WHO[7].

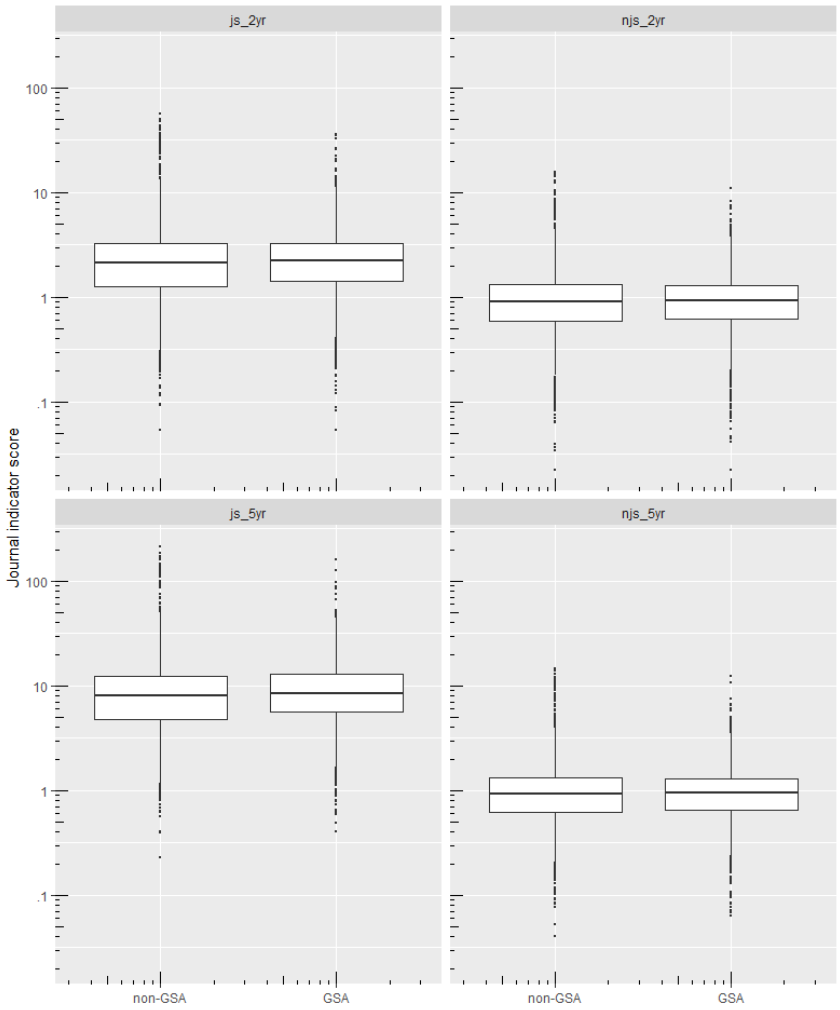
Listing of authors. A custom-made algorithm was built to examine the prevalence of author listings based on alphabetical order. The algorithm compared the actual listing of authors in each document in the dataset to an alphabetical order. Only papers with at least four authors were included in the analysis, 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
```

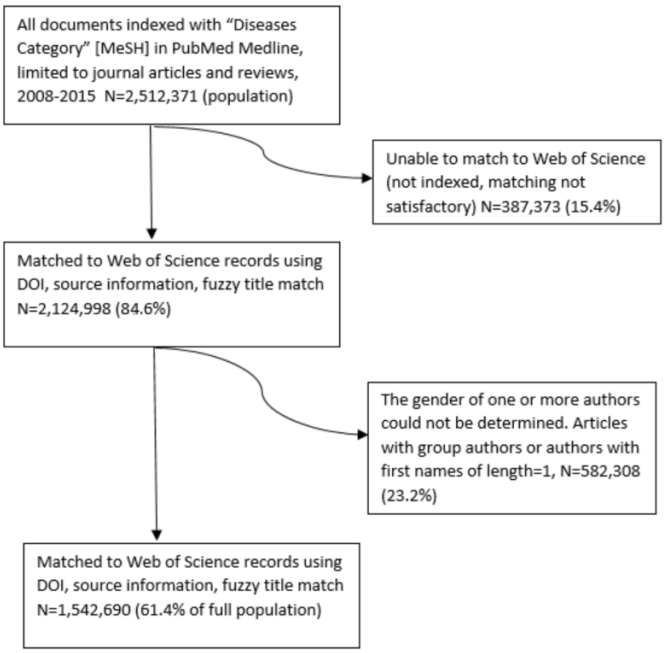
Supplementary Table 11 presents the results of the analysis. **n_authors** refers to the number of authors per paper**, p_total** specifies the total number of papers per n_authors value, **p_alpha** gives the number of papers using alphabetical order per n_authors value and **pp_alpha** shows the percentage 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.

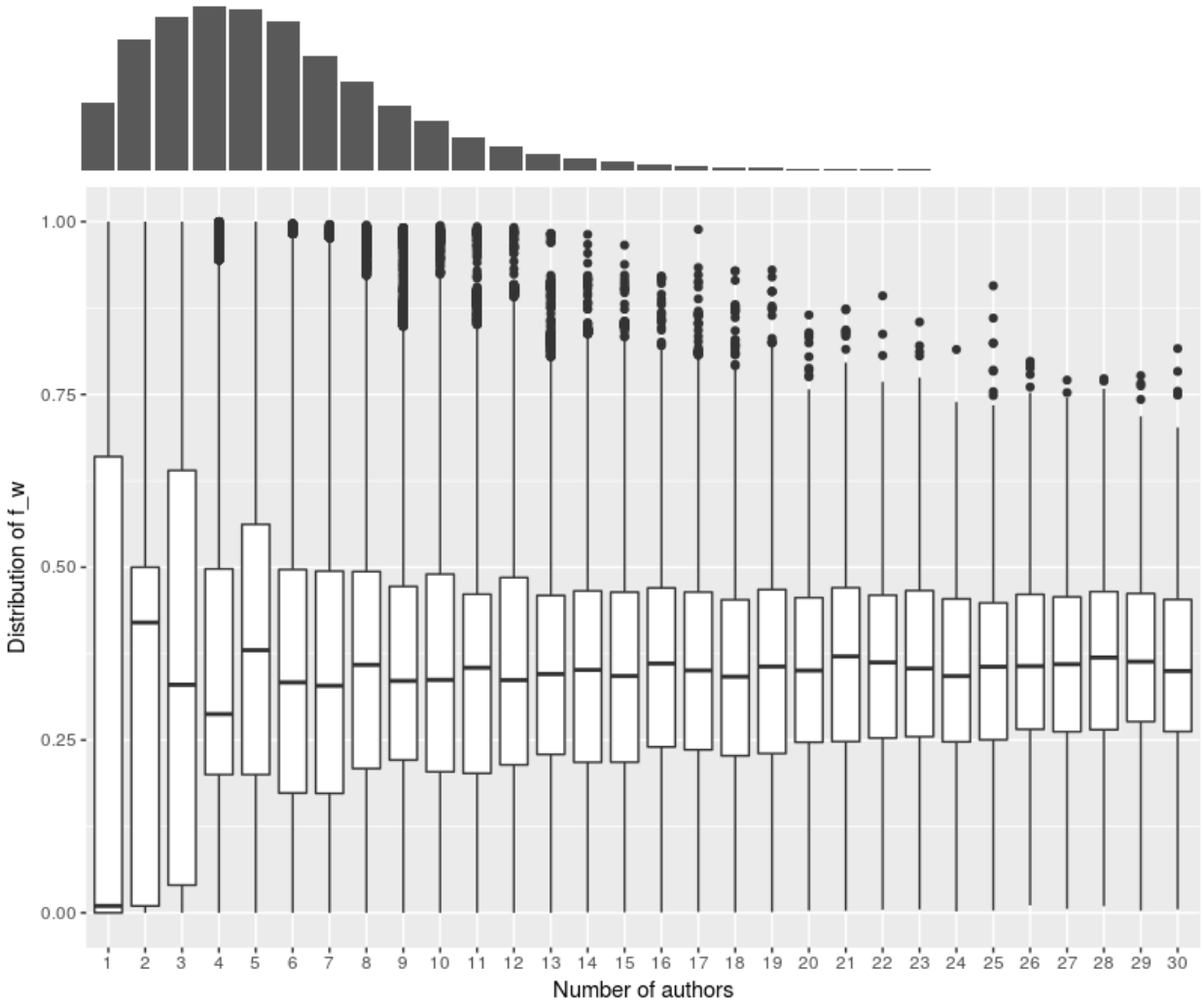
2. Supplementary Figures



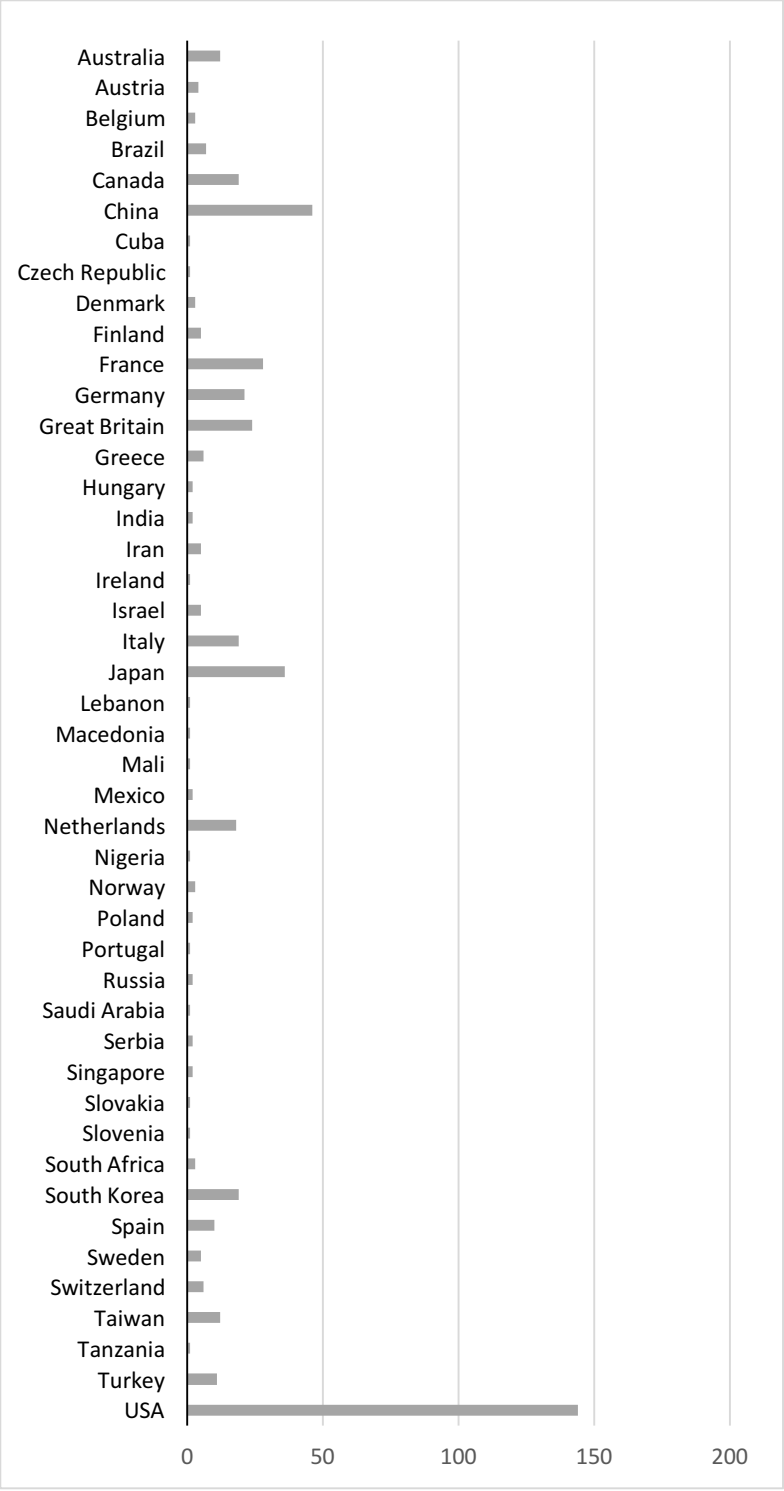
Supplementary Figure 1. Boxplot of aggregate journal-publication status for GSA and non-GSA publications. Journal status is approximated with journal citation indicators, two unnormalised (*js*) and two normalised indicators (*njs*). Both sets of journal indicators are calculated with two and five year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



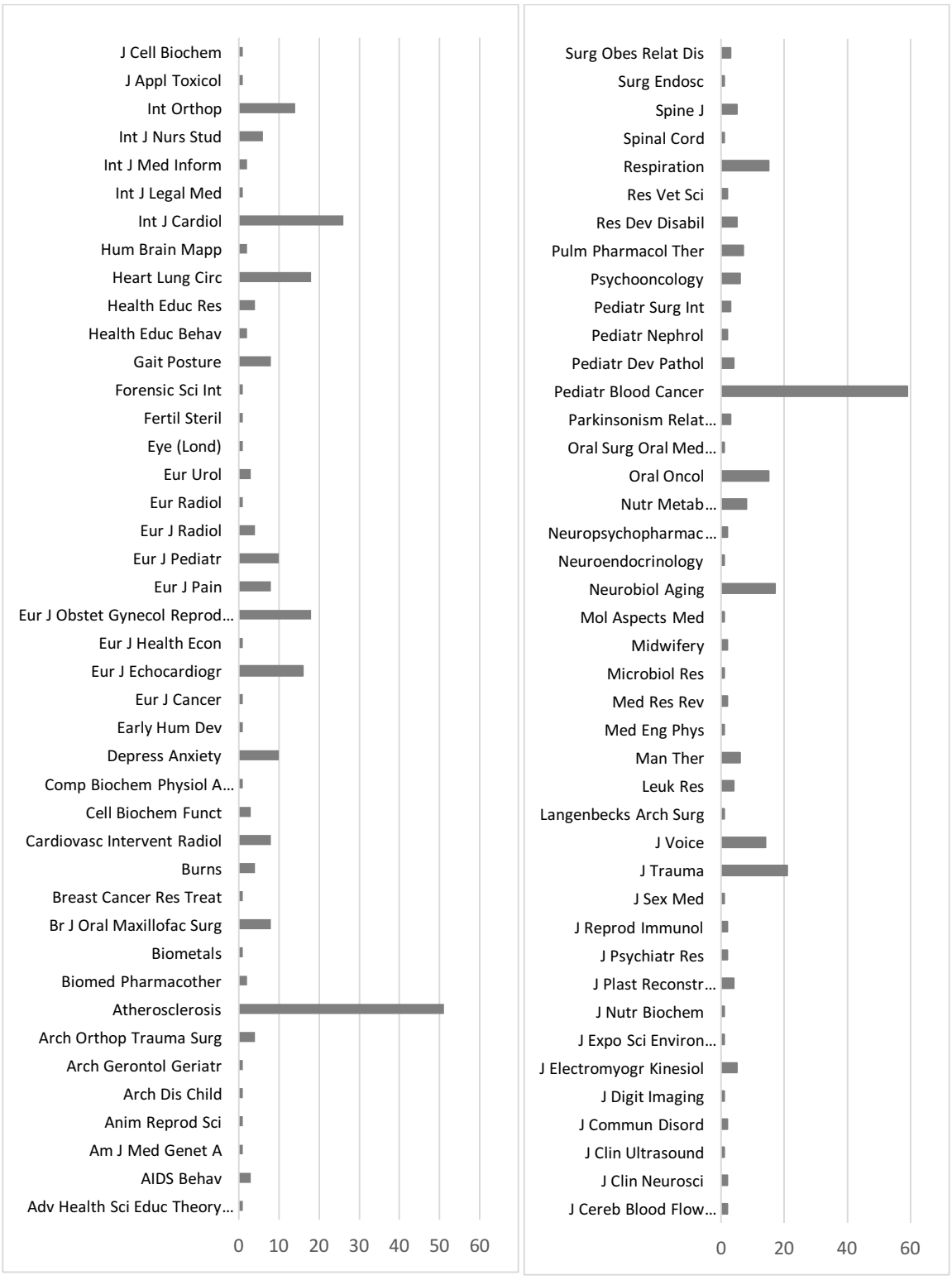
Supplementary Figure 2. Flowchart of data inclusion and exclusion.



Supplementary Figure 3. Boxplots of *fw* 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

	Women first authors	*Women last authors*
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]			
f_first	0.50	0.04	[0.43:0.58]	1.66	0.07	[1.53:1.79]
f_first country	0.97	0.29	[0.40:1.53]	2.75	0.80	[1.50:4.62]
f_first MeSH	1.78	0.26	[1.27:2.30]	6.13	1.67	[3.55:10.00]
f_first SC	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]			
f_last	0.44	0.04	[0.36:0.52]	1.56	0.06	[1.44:1.68]
f_last country	1.34	0.32	[0.71:1.96]	4.03	1.31	[2.02:7.13]
f_last MeSH	1.51	0.31	[0.96:2.19]	4.76	1.55	[2.48:8.49]
f_last SC	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]			
fw	1.14	0.07	[1.01:1.28]	3.14	0.22	[2.74:3.59]
fw country	0.71	0.37	[-0.01:1.45]	2.19	0.85	[0.99:4.26]
fw MeSH	1.53	0.31	[0.93:2.15]	4.86	1.58	[2.55:8.58]
fw SC	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
f_first	0.40	0.0003	[0.40:0.40]	0.49	0.0073	[0.48:0.51]
f_ last	0.27	0.0003	[0.27:0.27]	0.35	0.0067	[0.33:0.36]
fw	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]			
f_first	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]			
f_last	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]			
fw	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

	js_2yr	*njs_2yr*	*js_5yr*	*njs_5yr*
GSA, n = 3242				
1st 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
3rd quartile	3.3	1.3	12.9	1.3
non-GSA, n = 4000				
1st 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
3rd quartile	3.3	1.3	12.5	1.3
non-GSA, n = 50000				
1st 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
3rd 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 *f* > .60	154 (31%)	105	98 (93%)	**7 (7%)**
Unisex *f*:.41-.60	27 (.5%)	12	5 (42%)	7 (58%)
Male *f* <.41	306 (61%)	230	**5 (2%)**	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

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