In these, people with chronic diseases are brought together to support each other to better manage their conditions.

Since 2009, we have been involved in the Southampton Initiative for Health, which uses an empowering, problem-solving approach to improve the diets and physical activity levels of Southampton's most disadvantaged young women and their children. The programme has involved training the staff of the city's Sure Start Children's Centres — providers of services such as baby clinics, breastfeeding and weaning support, dentistry, parenting and cookery classes — in having conversations that encourage women to identify problems and generate solutions to change behaviour.

Although the data suggest that attending centres staffed by workers using this approach enhances women's sense of empowerment, to improve their nutritional status we need both to help women to feel more in control of their food choices and to make it easier for them to make better choices.

On a small scale, such a multilevel approach has proved effective. Trials in Canada, Australia and the United States demonstrate that the diets of small-town residents can be improved when efforts to enhance people's sense of empowerment in relation to healthy eating are pursued alongside local media campaigns to promote the benefits of eating well, together with programmes that help people to gain better access to fruits and vegetables and skills in food preparation. The challenge is to scale up such efforts to the wider public-health arena, because this means engaging political and commercial interests, including those of powerful food companies.

We believe that the methods used by people working in public health to engage politicians and food companies need to undergo a similar transformation to those being used to engage individuals. So far, public-health advocates have called for regulation and legislation as a means to improve diets — an increased tax on fatty and sugary foods, for instance. Yet this is unlikely to happen because raising the tax on soft drinks, say, is not in the interests of industry, or of politicians, who are sensitive to industry pressures and to a public desire for cheap soft drinks.

"We need to help women to feel more in control of their food choices."

Instead of wagging fingers, we need to generate consensus. Empowering consumers to call for better access to better food will put pressure on politicians to respond to voters, and on the food industry to please their customers.

More than 20 years ago, one of us (D.B.) wrote[10] in this journal that "if more was known about the processes by which the environment in early life influences adult health … the rise in incidence of 'Western' disease [might be] minimized." Today, we have the knowledge to readily prevent chronic diseases, had we but the will to do so. ■

David Barker *died on 27 August 2013, after writing the first draft of this Comment (see* Nature ***502,*** *304; 2013). Before his death, he was professor of clinical epidemiology at the Medical Research Council (MRC) Lifecourse Epidemiology Unit, University of Southampton, UK; professor of cardiovascular medicine at the Heart Research Center, Oregon Health and Science University in Portland; and visiting professor at the Center for the Study of Human Health, Emory University, Atlanta, Georgia.* **Mary Barker** *is senior lecturer in psychology at the MRC Lifecourse Epidemiology Unit, University of Southampton, UK.* **Tom Fleming** *is professor of developmental biology at the Centre for Biological Sciences, University of Southampton, UK.* **Michelle Lampl** *is director of the Center for the Study of Human Health and Professor of Anthropology at Emory University, Atlanta, Georgia.*
e-mail: meb@mrc.soton.ac.uk

1. Chen, L., Magliano, D. J. & Zimmet, P. Z. *Nature Rev. Endocrinol.* **8,** 228–236 (2012).
2. Laslett, L. J. *et al. J. Am. Coll. Cardiol.* **60,** S1–S49 (2012).
3. Barker, D. J. P. *Brit. Med. Bull.* **53,** 96–108 (1997).
4. Barker, D. J. P., Eriksson, J. G., Forsén, T. & Osmond, C. *Int. J. Epidemiol.* **31,** 1235–1239 (2002).
5. Robker, R. L. *et al. J. Clin. Endocrinol. Metab.* **94,** 1533–1540 (2009).
6. Barker, D. J. P. & Thornburg, K. L. *Placenta* **34,** 841–845 (2013).
7. Barker, M. *et al. Pub. Health Nutr.* **11,** 1229–1237 (2008).
8. Manandhar, D. S. *et al. Lancet* **364,** 970–979 (2004).
9. Bodenheimer, T., Lorig, K., Holman, H. & Grumbach, K. *J. Am. Med. Assoc.* **288,** 2469–2475 (2002).
10. Barker, D. J. P. *Nature* **338,** 371–372 (1989).

Global gender disparities in science

Cassidy R. Sugimoto and colleagues present a bibliometric analysis confirming that gender imbalances persist in research output worldwide.

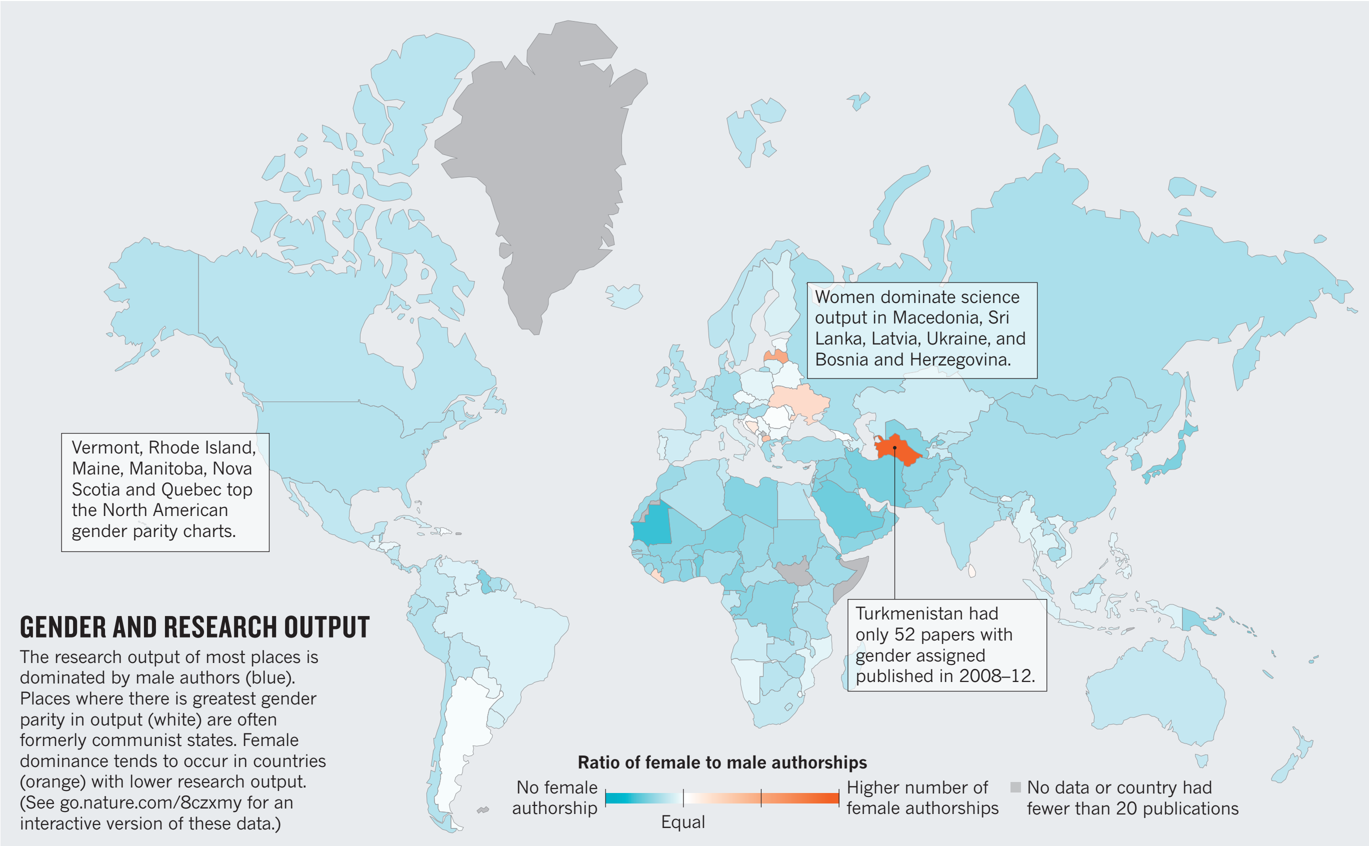
Despite many good intentions and initiatives, gender inequality is still rife in science. Although there are more female than male undergraduate and graduate students in many countries[1], there are relatively few female full professors, and gender inequalities in hiring[2], earnings[3], funding[4], satisfaction[5] and patenting[6] persist.

One focus of previous research has been the 'productivity puzzle'. Men publish more papers, on average, than women[7], although the gap differs between fields and subfields. Women publish significantly fewer papers in areas in which research is expensive[8], such as high-energy physics, possibly as a result of policies and procedures relating to funding allocations[4]. Women are less likely to participate in collaborations that lead to publication and are much less likely to be listed as either first or last author on a paper[7]. There is no consensus on the reasons for these gender differences in research output and collaboration — whether it is down to bias, childbearing and rearing[9], or other variables.

➲ NATURE.COM
See *Nature*'s special issue on women in science:
nature.com/women

It has been suggested that what women lack in research output they make up for in citations, particularly in fields with 'greater career risk'[8] — that is, fields with long lags between doctoral education and securing a faculty position, such as ecology. But again, there is no consensus on the relative impact of women's work compared to men's.

The present state of quantitative knowledge of gender disparities in science has been shaped primarily by anecdotal reports and studies that are highly localized, ▶



GENDER AND RESEARCH OUTPUT

The research output of most places is dominated by male authors (blue). Places where there is greatest gender parity in output (white) are often formerly communist states. Female dominance tends to occur in countries (orange) with lower research output. (See go.nature.com/8czxmy for an interactive version of these data.)

monodisciplinary and dated. Furthermore, these studies take little account of the rise in collaborative research and other changes in scholarly practices. Effective policy cannot be built on such foundations.

Therefore, we present here a global and cross-disciplinary bibliometric analysis of: first, the relationship between gender and research output (for which our proxy was authorship on published papers); second, the extent of collaboration (for which our proxy was co-authorships); and third, scientific impact of all articles published between 2008 and 2012 and indexed in the Thomson Reuters Web of Science databases (for which our proxy was citations). We analysed 5,483,841 research papers and review articles with 27,329,915 authorships. We assigned gender using data from the US Social Security database, among other sources (see Supplementary Information; go.nature.com/j3otjz).

We find that in the most productive countries, all articles with women in dominant author positions receive fewer citations than those with men in the same positions. And this citation disadvantage is accentuated by the fact that women's publication portfolios are more domestic than their male colleagues — they profit less from the extra citations that international collaborations accrue. Given that citations now play a central part in the evaluation of researchers, this situation can only worsen gender disparities.

In our view, the scale of this study provides much-needed empirical evidence of the inequality that is still all too pervasive in science. It should serve as a call to action for the development of higher education and science policy.

BIAS BY NUMBERS

Men dominate scientific production in nearly every country; to what extent varies by region (see 'Gender and research output'). We probed the proportion of each gender's output by comparing the proportion of identified authorships for each gender on any given paper. For example, for a paper with eight authorships, of which six were assigned a gender, each of the authorships would be granted one-sixth of a paper. These gendered fractions were then aggregated at the levels of countries and disciplines. It should be stressed that these are authorships, not individuals, therefore no author name disambiguation was necessary (see Supplementary Information).

Globally, women account for fewer than 30% of fractionalized authorships, whereas men represent slightly more than 70%. Women are similarly underrepresented when it comes to first authorships. For every article with a female first author, there are nearly two (1.93) articles first-authored by men.

South American and Eastern European countries demonstrate greater gender parity. Eastern Europe may support the idea that communist and formerly communist states may have greater gender balance than other countries. Only nine countries had female dominance in terms of proportion of authorships, and five of these (Macedonia, Sri Lanka, Latvia, Ukraine, and Bosnia and Herzegovina) had more than 1,000 articles in our analysis. In other words, female authorship is more prevalent in countries with lower scientific output.

Countries with more than 1,000 papers and high degrees of male dominance include, unsurprisingly (in order of output): Saudi Arabia, Iran, Japan, Jordan, the United Arab Emirates, Cameroon, Qatar and Uzbekistan. US states with more than 1,000 articles with a gender assigned and high male dominance include New Mexico, Mississippi and Wyoming. The US states and Canadian provinces that are closest to achieving gender parity (and have more than 1,000 articles) include Vermont, Rhode Island, Maine, Manitoba, Nova Scotia and Quebec. Again, some of these states and provinces are among the lowest ranking in terms of scientific output.

Our disciplinary results confirmed previous findings and anecdotal knowledge about fields associated with 'care'. Specialties dominated by women include nursing; midwifery; speech, language and hearing; education; social work and librarianship. Male-dominated disciplines include military sciences, engineering, robotics, aeronautics and astronautics, high-energy physics, mathematics, computer science, philosophy and economics. Although disciplines from the social sciences show a larger proportion of female authors, the humanities are still heavily dominated by men.

Next we looked at collaboration. We analysed the proportion of papers by gender that are the result of national collaboration, compared with those that result from international collaborations. For the 50 most productive countries in our analysis (accounting for 97% of the total publications), female collaborations are more domestically oriented than are the collaborations of males from the same country.

And what of impact? We analysed prominent author positions — sole authorship, first-authorship and last-authorship. We discovered that when a woman was in any of these roles, a paper attracted fewer citations than in cases in which a man was in one of these roles (see 'Lead-author gender and citation'). The gender disparity holds for national and international collaborations.

AGE-OLD STORY

There are several limitations to the conclusions that can be drawn from our findings. Foremost among them is that age indisputably has a role — perhaps even the major role — in explaining gender differences in scientific output, collaboration and impact. As is well known, the academic pipeline from junior to senior faculty leaks female scientists, and the senior ranks of science bear the imprint of previous generations' barriers to the progression of women. Thus it is likely that many of the trends we observed can be explained by the under-representation of women among the elders of science. After all, seniority, authorship position, collaboration and citation are all highly interlinked variables.

Another key limitation is that authorship of papers is only one of many indicators of research activity. Our analysis includes only journal articles, not books, conference proceedings, database construction or code, for example. Also problematic is the lack of universal norms associated with authorship attribution and position. For example, it is possible that some women do not appear as authors despite their contribution to research activities, and there are fields in which authors are listed alphabetically. There is also a concern that gender-assignment techniques can introduce errors (see Supplementary Information). We have tried to mitigate this with validation exercises, but there is always room for improvement.

> *"Programmes fostering international collaboration for female researchers might help to level the playing field."*

Future research should drill into questions raised by this analysis. What distinguishes pockets of anomalously high parity? Are there characteristics of the work itself that contribute to disparities in output and citation? Are there other, perhaps less quantitative, aspects of scholarship that reveal a different story regarding gender balance in science? Furthermore, is there anything intrinsic to certain disciplines or cultures that make them more or less appealing to scientists of a particular gender?

LEVELLING THE PLAYING FIELD

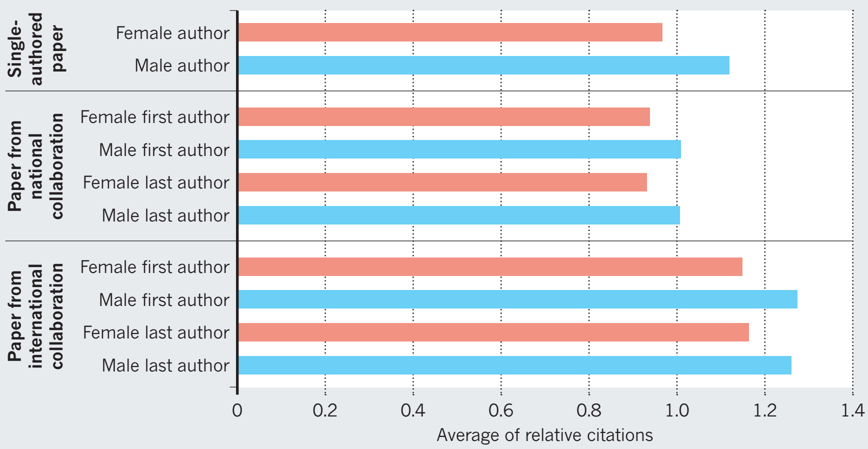
Those of a misogynistic bent might read this study as confirming their view that women's research is weaker than men's and there is less of it. Such a simplistic interpretation dismisses the vast implications embedded in these data. Our study lends solid quantitative support to what is intuitively known: barriers to women in science remain widespread worldwide, despite more than a decade of policies aimed at levelling the playing field. UNESCO data show[10] that in 17% of countries an equal number of men and women are scientists. Yet we found a grimmer picture: fewer than 6% of countries represented in the Web of Science come close to achieving gender parity in terms of papers published.

For a country to be scientifically competitive, it needs to maximize its human intellectual capital. Our data suggest that, because collaboration is one of the main drivers of research output and scientific impact, programmes fostering international collaboration for female researchers might help to level the playing field.

That said, if there were a simple solution or programme that could improve matters, this issue would already be solved. Unfortunately, behind this global imbalance lie local and historical forces that subtly contribute to the systemic inequalities that hinder women's access to and progress in science. Any realistic policy to enhance women's participation in the scientific workforce must take into account the variety of social, cultural, economic and political contexts in which students learn science and scientific work is performed. Each country should carefully identify the micro-mechanisms that contribute to reproducing the past order. No country can afford to neglect the intellectual contributions of half its population. ■

LEAD-AUTHOR GENDER AND CITATION

Papers with female authors in key positions are cited less than those with male authors in key positions, be they papers with one author, or those resulting from national or international collaborations.



➔ FOR AN INTERACTIVE VERSION OF THIS CHART SEE: go.nature.com/j3otjz

Vincent Larivière *is chaired assistant professor of transformations of scholarly communication at the University of Montreal, Canada.* **Chaoqun Ni** *is in the School of Informatics and Computing at Indiana University Bloomington.* **Yves Gingras** *is chaired professor of history and sociology of science at the University of Quebec at Montreal, Canada.* **Blaise Cronin** *is chaired professor of information science at Indiana University Bloomington.* **Cassidy R. Sugimoto** *is assistant professor in the School of Informatics and Computing, Indiana University Bloomington.*
e-mail: sugimoto@indiana.edu.

1. Organisation for Economic Co-operation and Development. *Education at a Glance 2012* (OECD, 2012).
2. Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J. & Handelsman, J. *Proc. Natl Acad. Sci. USA* **109,** 16474–16479 (2012).
3. Shen, H. *Nature* **495,** 22–24 (2013).
4. Ley, T. J. & Hamilton B. H. *Science* **322,** 1472–1474 (2008).
5. Holden, C. *Science* **294,** 396–411 (2001).
6. Ding, W. W. *Science* **313,** 665–667 (2006).
7. West, J. D., Jacquet, J., King, M., Correll, S. J. & Bergstrom, C. T. *PLoS ONE* e66212 (2013).
8. Duch, J. *et al. PLoS ONE* **7,** e51332 (2012).
9. Ceci, S. J. & Williams, W. M. *Proc. Natl Acad. Sci. USA* **108,** 3157–3162 (2011).
10. United Nations Educational, Scientific and Cultural Organization. *Science, Technology and Gender: An International Report* (UNESCO, 2007).

Supplementary Information to:
Global gender disparities in science (Comment in *Nature* 504, 211–213; 2013)

Vincent Larivière, Chaoqun Ni, Yves Gingras, Blaise Cronin & Cassidy R. Sugimoto

DATA DESCRIPTION

WEB OF SCIENCE

Data for this project are drawn from Thomson Reuters' Web of Science database, covering the Science Citation Index Expanded, the Social Sciences Citation Index and the Arts and Humanities Citation Index. All articles from 2008 to 2012 were included in the analysis. The raw data were transformed into a relational database on an SQL server, hosted at Observatoire des sciences et des technologies (OST) at the University of Quebec at Montreal, Canada, in order to perform the various analyses. Since 2008, the Web of Science (WoS) includes the full first name of authors, which allows for gender classification of authors (see next sections). Thomson Reuters also indexes institutional address (institution, country, city, etc.) of each author, which allows for precise geographical assignation of articles by gender.

Indicators presented in this research are based on the number of articles and review articles published by authors of each gender. Other types of documents — such as editorials, letters to the editor, and book reviews — are excluded from the analysis because they are generally not peer-reviewed, nor considered as original contributions to scholarly knowledge[1]. These numbers are based on fractional counting of papers: that is, each author is given $1/x$ count of the authorship where x represents the number of authors for which a gender could be assigned on the given paper.

Citation measures account for all citations received by a given paper, from its publication year to the end of 2012. To compare data between different specialties, each article's number of citations is divided by the average number of citations received by articles in the same discipline published that year[2,3]. When the average of relative citations (ARC) is above 1, a given article is cited above the world average for the same field. Conversely, an ARC below 1 means that the number of citations received is below the world average. Of course, the well-known limitations of bibliometrics apply to this analysis, as the Web of Science does not index all the world's scholarly literature. This is more problematic for the social sciences and the humanities, where (a) there is virtually no coverage of research output in media other than journal articles[4] and (b) there is very limited coverage of research output in the form of articles written in languages other than English[5].

NAME GENDER ASSIGNMENT

GENDER NAME INFORMATION LISTS

Gender information of WoS authors was determined by matching names with universal and country-specific name lists. Universal lists were applied to the entire set of WoS authors, and country-specific lists were applied to subsets of WoS authors associated with the corresponding countries. The following Table S1 displays the lists utilized to categorize authors' gender.

Table S1. Gender name information lists

List	List Source
US Census	https://www.census.gov/genealogy/www/data/1990surnames/names_files.html
WikiName	http://wiki.name.com/en/Baby_Names
Wikipedia	http://en.wikipedia.org/wiki/Category:Given_names_by_gender
French	http://en.wikipedia.org/wiki/French_name http://en.wikipedia.org/wiki/Category:French_feminine_given_names http://en.wikipedia.org/wiki/Category:French_masculine_given_names
Quebec Census	http://www.rrq.gouv.qc.ca/en/enfants/Pages/banque_prenoms.aspx
Korea	http://en.wikipedia.org/wiki/List_of_Korean_given_names http://en.wikipedia.org/wiki/Category:Korean_given_names
Lithuania	http://en.wikipedia.org/wiki/Lithuanian_name
Persian / Iran	*http://www.top-100-baby-names-search.com/baby-names-persian.html*
Romania	http://en.wikipedia.org/wiki/Romanian_name http://en.wikipedia.org/wiki/Category:Romanian_given_names
Brazil/Portugal	http://en.wikipedia.org/wiki/Brazilian_name#Brazilian_names
Serbia	http://en.wikipedia.org/wiki/Serbian_name http://en.wikipedia.org/wiki/Slavic_names
Ukraine	http://en.wikipedia.org/wiki/Ukrainian_names http://en.wikipedia.org/wiki/Slavic_names http://www.top-100-baby-names-search.com/ukrainian-baby-names.html
Thailand	http://www.top-100-baby-names-search.com/thai-first-names.html
India	http://en.wikipedia.org/wiki/Category:Indian_given_names http://www.studentsoftheworld.info/penpals/stats.php3?Pays=IND www.pkp.in/info/downloads/India%20Baby%20Names.xls
Japan	http://en.wikipedia.org/wiki/Category:Japanese_given_names http://en.wikipedia.org/wiki/Japanese_name

US CENSUS

The US Census provides lists of given names and the percentage of the population with a specific given name and associated gender. Therefore, with the given names of authors obtained from WoS data, each author was coded for possible gender using these lists. In cases where a name was used for both genders, it was only attributed to a specific gender when it was used at least ten times more frequently for one gender than the other. Otherwise it was categorized as a "unisex" name. The US Census data were utilized as the primary source in this project to categorize authors by gender. Other universal lists were only used for names that could not be categorized using the US Census list.

WIKINAME

This list provided 8,155 female and male names (non-exclusively). This was used to categorize names not matched by the US Census. As with the previous procedure, names appearing in both lists were categorized as unisex.

WIKIPEDIA

Wikipedia's given-name list provides names associated with more than 60 countries. This list was used to categorize authors that were not successfully categorized using the US Census data and WikiName.

QUEBEC AND FRENCH

This is a list of Canadian university professors' given names by gender, and a list of Quebec's newborns by gender. All non-English EU characters in this list were converted to corresponding basic English characters in order to match with the WoS author set (WoS provides English names).

KOREA

Korean names not matched in the universal lists were matched using a series of rules. For example, names ending with -jae are typically male names, while names with -mi- are typically associated with female names.

LITHUANIA

A rule-based approach to the remaining Lithuanian names was also applied: female names usually end with: -a, -e, or -ia; and ale names usually end with: -s, -as, -is, -ys, -us, and ius.

JAPAN

Rules were also used for classifying remaining Japanese names. Female names usually end with: -a, -chi, -e, -ho, -i, -ka, -ki, -ko, -mi, -na, -no, -o, -ri, -sa, -ya, and -yo. Male names usually end with: -aki, -fumi, -go, -haru, -hei, -hiko, -hisa, -hide, -hiro, -ji, -kazu, -ki, -ma, -masa, -michi, -mitsu, -nari, -nobu, -nori, -o, -rou, -shi, -shige, -suke, -ta, -taka, -to, -toshi, -tomo, -ya, and -zou.

RUSSIA AND RELATED COUNTRIES

Previous assignments were based on first names. For Russian names, however, last names were also used. Men's family names typically end in -ov, -ev or -in. Women's typically end in -ova, -eva or -ina. These 'suffixes' were thus applied to Russian authors, as well as to other countries where 95% or more of the women or men's names already assigned ended in one of the abovementioned suffixes (Czech Republic, Bulgaria, Latvia, Kazakhstan, Uzbekistan, Lithuania and Luxembourg).

PERSIAN / IRAN, BRAZIL, ROMANIA, PORTUGAL, SERBIA, UKRAINE, THAILAND AND INDIA

For Iran, Brazil, Romania, Portugal, Serbia, Ukraine, Thailand and India, we compiled specific lists of names and gender for each county based on information obtained online. Please refer to Table 1 for the sources used in compiling country-specific lists and naming rules.

CHINA

There were 84,462 unique author names associated with affiliations located in China, corresponding with 1,841,748 authorships. The distribution of number of authorships over unique author names follows a power law distribution. That is, majority of the author names were associated with a small number of papers, while a minority of author names were associated with a large number of papers. Specifically, there were 12,828 author names (15.17% of total) associated with 20 or more papers, accounting for about 84.25% of the total authorships in China. Therefore, we selected author names associated with at least 20 papers, and assigned the gender of each name manually. Two native speakers from China manually coded these names. They coded the gender of each name based on their knowledge of Chinese language and Chinese names. Web searches were also used in ambiguous cases to identify, for example, the predominant gender that arose in Google Images and were associated with various Facebook accounts.

TAIWAN

There is no unified pinyin system for translating Chinese names into English — Taiwanese choose from one of four different pinyin systems. Therefore, our assignment involved: 1) looking up the pinyin system used to translate the name into English; and 2) comparing it with zhuyin fuaho (http://www.boca.gov.tw/content?CuItem=5609&mp=1) to ascertain the correct punctuation. If the name was not in a pinyin system, it is labelled as unknown. If it is in the system, the pronunciation was used to determine a gender (evaluated by a native speaker). Any names considered ambiguous were marked as unknown.

METHODS

WoS AUTHOR NAME PRE-PROCESSING

The author-name list contains the given names of authors indexed by WoS. The given name was provided in a separate field, but not in a unified form. Some given names are initials instead of complete names, or contain special characters like "()", "-", "." or a space. In order to match with the source lists introduced above, the author-name set was preprocessed as follows:
All characters in "()" of a given name were extracted and treated as nick names;

- Identify initials:
  - Calculate the "." in the given name:
    - If no ".", calculate the space
    - If there is ".", calculate the length of whole string
      - If the length of a string is smaller than the 3 times the number of ".", then they are treated as initials.
      - If not: leave for next step
- For names that are not initials, split given name to several parts by space;
- Replace all hyphens in each part into a space: for instance, "Jean-Pierre" will be converted to "Jean Pierre".

It should be noted that we identified authorships, not individuals — that is, we were interested in identifying the gender of each authorship, but were not concerned with matching authors across papers. That is, we were interested in the gender of each author on each given paper, but not on how many papers were authored by that individual author. Our analysis is on the aggregate level — how many papers had a female or male author, not only how many papers were authored by each individual female or male author.

MATCHING WITH GENDER-NAME LISTS

As mentioned above, the author given-name set was matched with the universal and country specific lists to determine the gender of WoS authors. The match was done using the following order:

- US Census

- WikiName
- Wikipedia
- France and Quebec list
- Other country-specific lists

The US Census list was used as the basic source of gender information. Therefore, all the other lists (except for the country-specific lists) were only used to match given names that could not be matched by the US Census.

COVERAGE AT WORLD AND COUNTRY LEVELS

After these steps, we managed to assign a gender, female or male (F or M), to 56.1% of distinct given names (e.g. John, Linda), and 59.5% of distinct full authors' names (e.g. John Smith, Linda Madden) (see Table S2). A significant proportion of authors' names only provide initials (31.0% of distinct authors' names). Therefore, in terms of the percentage of authors that provided given name information beyond initial(s), gender was assigned to 57.3% of distinct given names and 83.0% of distinct full names.

Table S2. Number and percentage of full names and of given names assigned a gender.

Gender	Full names			Given names		
	N	% of all	% (All - Initials)	N	% of all	% (All - Initials)
Female	1,194,340	25.0%	35.0%	209,737	25.3%	25.8%
Male	1,642,066	34.4%	48.1%	256,166	30.8%	31.5%
Unisex	123,023	2.6%	3.6%	23,919	2.9%	2.9%
Unknown	456,020	9.6%	13.4%	323,687	39.0%	39.8%
Initials	1,354,802	28.4%	-	16,945	2.0%	-
All	4,770,251	100.0%	-	830,454	100.0%	-

At the level of distinct papers and paper-authors (e.g. the sum of each author appearing on the byline of articles), the results are similar (Table S3). 81.3% of papers had at least one of their authors assigned a gender, and 65.2% of the author paper combinations had a gender assigned. When authors with only initials are excluded, this percentage increases to 86.1%.

Table S3. Number and percentage of distinct papers and of author-papers assigned a gender.

Gender	Distinct papers		Author-paper combinations		
	N	% of all	N	% of all	% (All - Initials)
Female	2,750,850	50.2%	5,546,226	20.3%	26.8%
Male	4,116,595	75.1%	12,264,088	44.9%	59.3%
Any gender	4,458,622	81.3%	17,810,314	65.2%	86.2%
Unisex	496,825	9.1%	563,954	2.1%	2.7%
Unknown	1,542,186	28.1%	2,298,439	8.4%	11.1%
Initials	1,153,640	21.0%	6,657,208	24.4%	-
N papers	5,483,841	100.0%	27,329,915	100.0%	-

Table S4 (provided in full at http://dx.doi.org/10.1038/504211a) presents the number distinct authors and given names falling in each of the categories, along with the percentage (of all and of all minus initials) of those assigned a gender, while Table S5 (http://dx.doi.org/10.1038/504211a) presents the same measures for distinct papers and paper-author combinations. Although not identical, the coverage of different countries in terms of the proportion of authors and papers assigned is generally in the same range.

VALIDATION STUDY

To assess the accuracy of our analysis, we selected 1,000 records at random representing an individual author who had been categorized into each of the following five categories: initials, unknown, unisex, male and female. These authors were associated with a specific country, institution, and, in some cases, an email address. This information was used to locate biographical information or a photo on the web that could be used to verify the accuracy of the categorization. The percent of the random sample that could be gender identified varied by category (see Table S6) and was dependent on many variables, including the status of the author. For example, in the male category, many of the authors were technicians and staff members who lacked lengthy biographical information (which would contain pronouns) or photographs.

Table S6. Percent male and female in each category

Category	# and % identified	# and % female (of identified)	# and % male (of identified)
Initials	839 (83.9%)	198 (23.6%)	641 (76.4%)
Unknown	890* (89.0%)	282 (31.7%)	607 (68.2%)
Unisex	540 (54.0%)	113 (20.9%)	427 (79.1%)
Male	605 (60.5%)	10 (1.7%)	595 (98.3%)
Female	830 (83.0%)	720 (86.7%)	110 (13.3%)

*The number here is not the sum of the male and female due to the fact that one author self-identified as ‘other’. They are, therefore, neither male, female, nor unidentified.

DATA ANALYSIS & VISUALIZATIONS

R was the primary data analysis and visualization tool, and ArcGIS was used to display the North America Details. Tableau software and Data-Driven Document (D3) JavaScript library were also used, mainly for the interactive versions of visualizations.

A list of 206 countries/territories was originally extracted from WoS based on the author address information provided by each publication. A list of countries with less than 20 publications in the studied time period was excluded for the analysis regarding productivity, collaboration and impact, in order to reduce possible distortion resulting from a small number of samples. The name of countries (in English) provided by the WoS database was used. While making the global map, names as provided by International Organization for Standardization (ISO)'s *3166 standard* were used instead of WoS names. For instance, the Democratic Republic of the Congo is Zaire in the WoS database, which officially refers to the state that existed between 1971 and 1997. South Korea is the name from WoS, while it should be the Republic of Korea according to the ISO *3122 Standard*. For each country, the number of publications and their corresponding citations were obtained by aggregating at the country level.

A world map was utilized as the base map to display the differences in female and male research output by country using D3 library. The counting of papers by gender presented in the world map and discipline map is based on fractionalized authorships, which are obtained by compiling, for each paper, the proportion of male and female authors. Hence, for a paper having 8 authors, of which a gender could be assigned to 6, each author — and its corresponding gender — was assigned 1/6 of a paper (authors for which no gender could be assigned were excluded from the denominator). These gendered fractions of were then aggregated at the levels of countries and disciplines and serve as a basis for the F–M ratios presented in the world map and the discipline map. Each country was colour coded on the basis of differences in female and male research output: the bluer a country is, the higher the male to female research output is in that country; the oranger a country is, the higher the female to male research output is in that country. It should be noted that there are some countries without any publication records in our WoS data set for the years 2008–12. Those countries were coloured grey in the geographic map. A similar analysis was done for US states and Canadian provinces.

The proportion of scholarly output of female and male was also examined at the level of the 554 *UCSD Map of Science* subject categories, which was approximated as a discipline/specialty in this project. Like in the world map, these ratios were compiled based on fractionalized authorship (see above). The difference between the research output of male and female, therefore, was calculated by dividing the sum of fractionalized authorships of women to that of men for each discipline. To visually display the difference by discipline, the UCSD map of science was utilized as the base map and the difference in research output of each gender in each discipline was overlaid on top of the base map, using D3 library. Each discipline (a node at the map) was coloured based on the value of difference: the bluer a node is, the more active males are in the corresponding discipline; the more orange a node is, the more active females are in the corresponding discipline.

Collaboration patterns of female and male were also investigated, at both international and national level. In this project, the international collaboration rate of female (male) authors in a country was calculated as the number of papers finished by female (male) author collaborating with others from another country divided by the number of papers in that country with at least one female (male) author in bylines. Similarly, the national collaboration rate of female (male) authors in a country was calculated as the number of papers finished by female (male) author collaborating with others from the same country divided by the number of papers in that country with at least one female (male) author in bylines. A bar chart was adopted here to show the international and national collaboration of female and male by country. (Interactive version online at http://dx.doi.org/10.1038/504211a, with countries displayed in descending order of the female national collaboration rate.)

A heatmap was constructed to display the difference in impact of publications of different authorship categories. It should be noted that citations were counted with an open citation window and were normalized by the average citation rates of the papers published in the same specialty the same year. The heatmap here is a visualization of each county's impact in different categories of publications, i.e. a matrix of countries by categories of publications, with each crossing cell showing the citation count. The citation count was colour coded using a red–white–green diverging colour palette, each colour corresponding with the minimum–median–maximum number of citations. That is, the redder the colour is, the less citations received; the greener the colour is, the more citations received.

1. Moed, H. F. *Scientometrics* **35**, 177–191 (1996).
2. Moed, H. F., De Bruin, R. E. & van Leeuwen, TH. N. *Scientometrics* **33**, 381–422 (1995).
3. Schubert, A. & Braun, T. *Scientometrics* **9**, 281–291 (1986).
4. Larivière, V., Archambault, É., Gingras, Y. & Vignola-Gagné, É. *J. Am. Soc. Inf. Sci. Tech.* **57**, 997–1004 (2006).
5. Archambault, É. Vignola-Gagné, É., Côté, G., Larivière, V. & Gingras, Y. *Scientometrics* **68**, 329–342 (2006).