Institute for Scientific Information

Global Research Report
Multi-authorship and research analytics

Jonathan Adams, David Pendlebury, Ross Potter and Martin Szomszor

Web of Science Group
A Clarivate Analytics company

Author biographies

Jonathan Adams is Director of the Institute for Scientific Information, at the Web of Science Group. He is also a Visiting Professor at King's College London, Policy Institute, and was awarded an Honorary D.Sc. in 2017 by the University of Exeter, for his work in higher education and research policy.

Ross Potter is a Data Scientist at the Institute for Scientific Information. He has extensive research experience within academia, including NASA-related postdoctoral positions at the Lunar and Planetary Institute, Houston, Texas, and Brown University, Providence, Rhode Island.

Martin Szomszor is Head of Research Analytics at the Institute for Scientific Information. He joined from Digital Science where, as Chief Data Scientist he applied his extensive knowledge of machine learning, data integration and visualization techniques to found the Global Research Identifier Database. He was named a 2015 top-50 UK Information Age data leader for his work in creating the REF2014 impact case studies database for the Higher Education Funding Council for England (HEFCE).

David Pendlebury is Head of Research Analysis at the Institute for Scientific Information, at the Web of Science Group. Since 1983 he has used *Web of Science* data to study the structure and dynamics of research. He worked for many years with ISI founder Eugene Garfield. With Henry Small, David developed ISI's *Essential Science Indicators*.

Foundational past, visionary future

The Institute for Scientific Information

ISI builds on the work of Dr. Eugene Garfield – the original founder and a pioneer of information science. Named after the company he founded – the forerunner of the Web of Science Group – ISI was re-established in 2018 and serves as a home for analytic expertise, guided by his legacy and adapted to respond to technological advancements.

Our global team of industry-recognized experts focus on the development of existing and new bibliometric and analytical approaches, whilst fostering collaborations with partners and academic colleagues across the global research community.

Today, as the 'university' of the Web of Science Group, ISI both:

- Maintains the foundational knowledge and editorial rigor upon which the *Web of Science* index and its related products and services are built. Our robust evaluation and curation have been informed by research use and objective analysis for almost half a century. Selective, structured and complete data in the *Web of Science* provide rich insights into the contribution and value of the world's most impactful scientific and research journals. These expert insights enable researchers, publishers, editors, librarians and funders to explore the key drivers of a journal's value for diverse audiences, making better use of the wide body of data and metrics available.
- Carries out research to sustain, extend and improve the knowledge base and disseminates that knowledge to our colleagues, partners and all those who deal with research in academia, corporations, funders, publishers and governments via our reports and publications and at events and conferences.

ISBN 978-1-9160868-6-9

Executive summary

The *Web of Science* has identified a growing number of research articles with 1,000 or more unique authors across more than 100 different countries. The combination of many authors/ many countries creates a complex authorship pattern that differs from more typical academic papers and drives elevated citation rates.

In this report we describe two patterns linking complex authorship with effects that increase citation rates: a general increase associated with multi-authorship (more than 10 authors and more than five countries); and more perturbing outcomes of hyper-authorship (more than 100 authors spread across more than 30 countries).

Across the *Web of Science*, the most **frequent number of authors** on an article is three and 95% of global output has 10 or fewer authors (Table 1). The most **frequent number of countries** on an article is one and 99% of global output has authors from five or fewer countries. (Figure 1, Table 2)

Complex authorship (many authors, many countries) has continued to rise in the last five years. The largest relative increases are associated with a marked rise in hyper-authorship. (Figure 2)

One additional country on an article has a greater benefit than one additional author: complex authorship is correlated with indicators of research performance (Category Normalized Citation Impact - CNCI, Figure 3); author count is linked to a slight but continuous impact rise (Figure 4); country count is linked to a steeper and more erratic impact rise (Figure 5).

Author and impact patterns vary between disciplines. In Biology, rising author and country counts are coherently linked to rising citation impact, but in Clinical Medicine the effect is more erratic for higher counts with higher CNCI up to 100 times the world average. In Chemistry there is no strong link between author count and citations (Figure 6); Particle Physics has erratically high impact values at high country counts (Figure 7).

The effect of **multi and hyper-authorship can be observed at country level.** The effect depends on the size of a country's domestic research base. For all countries, citation impact increases with rising authorship, but gains at higher counts are more evident and variable for smaller countries (Figure 8).

Every country gains citation impact through its share of the 5% of global multi-author 10 or more) articles. In small and growing research economies the average CNCI of these articles is five or more times higher than typical articles (Table 3).

We recommend…

That the presence, in any sample, of **articles with more than 10 authors** should be acknowledged and separately described because it will influence interpretation. Although **multi-authorship** leads to higher impact, this link is coherent, progressive and regular for most (but not all) discipline categories and for some fields there is little or no effect. No change needs to be made to data management or analytics in this regard.

That articles with **hyper-authorship**, **beyond 100 authors and/or 30 countries**, be treated differently. These articles are, to put it simply, different: they have unpredictable, incoherent effects that can sometimes be very large. There is a strong argument for removing such data from all associated analyses at national as well as at institutional level. Hyper-authorship produces particularly different and erratic patterns across Clinical Medicine and Particle Physics. The effects do not fit into a broader pattern, are not repeated across all disciplines and are far from consistent. The presence of such articles may be significant, even distorting, at institutional level.

Introduction

The Web of Science Group has long monitored the growth of author and address counts on research publications.

In 2012, writing in ISI's Science Watch, Chris King noted that the numbers of publications indexed on *Web of Science* that had more than 50 authors rose from around 400 to more than 1,000 between 1998 and 2011, while the number with more than 100 authors doubled to 600 over the same period (King, 2012).

Prior to 2000, the maximum number of authors on a single article rarely exceeded 500. In 2004 the 1,000 cap was broken with a paper of 2,500 authors. The abundance of such articles continues to increase and the record is now held by a 2015 article by the ATLAS team on the Higgs Boson, with 5,153 authors at more than 500 institutional addresses (Aad et al., 2015; Mallapaty, 2018).

The count of countries listed among author affiliations has similarly increased. International collaboration was relatively scarce in the 1980s but has grown rapidly: more than half of the articles attributable to any one country now have a co-author from another (Adams, 2013). The incentives for such collaboration are diverse; data, however, can be a key driver and the growth of multinational authorship is seen in both health research and environmental science. An article on human weight trends has the most unique country authors at 108 (more than half the members of the UN) and more than 1,000 institutional addresses (NCD Risk Factor Collaboration, 2017).

The reasons for, and benefits of, collaboration have been widely studied (e.g. Katz and Martin, 1997; Bozeman et al., 2013). *The Economist* (2016) reported on the rise of co-authorship and suggested that rising article counts per author did not represent increasing productivity. It pointed to 'guest authorship' and reciprocal offers of free author tickets to colleagues as drivers that have led to an actual drop in global productivity of articles per author.

Rising author counts were noted long ago by Derek De Solla Price (Price, 1963). Some may be driven by a cultural habit of adding senior members to research group outputs (Croll, 1984) which is said to be more common in some countries with relatively hierarchical structures. The increase in the number of biomedical papers that include a departmental chair as co-author has been noted (Drenth, 1998); however, the increase in multi-authorship is too widespread a phenomenon for this to be a general explanation. There are discipline analyses of authorship cultures and patterns in specific areas, including social sciences (Endersby, 1996), economics (Hudson, 1996) and medically related research where single authorship in the *New England Journal of Medicine* fell from 98% to 5% during the 20th century (Constantian, 1999).

This rise in authorship counts caused some concern (Cronin, 2001) about what a name in an author list meant in contemporary terms: is an author any longer synonymous with a writer? There are also questions of where accountability and collective responsibility lie as the authorship of a paper rises (Croll, 1984): is a paper with 100 authors the same kind of communication as a paper with one or a few authors? King (2012) reported the continuing rise in counts of papers with more than 1,000 authors (termed hyper-authorship: Cronin, 2001). These were concentrated in physics, including work at international space observatories and at CERN.

Cronin (2001) concluded that hyper-authorship signifies a change in research nature. The major challenges of research – population studies, epidemiology, climate change, particle and space sciences – require investment in equipment, data collection, longitudinal studies and analytical processing associated with large teams. The lone researcher is now a less viable model for major innovations.

Is a paper with 100 authors the same kind of communication as a paper with one or a few authors?

Authorship poses three questions for research analysts and policy makers. The first is to document and understand authorship changes and to determine whether they are discipline peculiarities or globally pervasive. The second is whether there is any relationship between the academic impact of publications and an increasingly large author count, and whether the hyper-authored publications are in a category of their own for analytical exercises. The third, though not addressed here, is the attribution of credit for a paper that has many authors (see Waltman and van Eck, 2015).

How many authors? How many countries?

To establish the general recent trend and current spread of authorship, we looked at 15.7 million documents identified specifically as articles (not reviews or other document types) in journals indexed in *Web of Science* over the period 2009-2018.

The earlier five-year period (2009-2013) contained 6.9 million articles; the latter five years (2014-2018), 8.8 million.

As the academic literature has already established, authorship is skewed; most articles have a small number of authors with a few having an exceptionally high number.
There are an additional 1,414 articles with 'zero' authors; these are group-authored with Agricultural Sciences accounting for 60%.

The most common author count over the period was three (Figure 1 – left). More than 10 million articles (over two-thirds of the total) have five or fewer authors and 14.9 million (almost 95% of the total) have 10 or fewer authors. This is an important background statistic; while there is no doubt that multi-authorship has increased and that hyper-authorship is now relatively frequent, the articles produced by small author groups still dominate the research literature. The massively authored article remains relatively scarce.

The distribution by countries (Figure 1 - right) also provides an important reference point - the most common group, by far, is that of articles on which all authors are from the same country. This may seem surprising given the headlines regarding the rise of international collaboration.

However, while international networks are common in Europe, our recent G20 Scorecard showed that about two-thirds of output from the USA and more than three-quarters of China's output is domestic (Adams et al., 2019) with collaboration networks still developing in much of Asia and Latin America.

Figure 1 provides an important reality check for multi-authorship: we need to be careful about what we assume is typical. There may be changes in the pattern of authorship but there is no evidence of disruptive shifts in traditional patterns.

The next step is to look at the evidence for recent change by comparing the spread of authorship for the most recent five years (2014-2018) compared to the five years before that (2009-2013) (Figure 2). Volume has increased between the two periods.
For articles grouped by author count, that seems to result in a similar order of magnitude increase across most groups, but there are two points of interest. First, the increase in count for articles with 1-5 authors is small compared to other groups suggesting a broad increase in multi-authorship. Second, there is evidence of a shift among the high (> 100) author counts with similar numbers of articles in the 100-500 and 500-1,000 author groups but a much greater increase in articles with over 1,000 authors.

For article authorship analyzed by country affiliation, the growth in frequency of authorship across national boundaries is much clearer. Counts in the groups up to 30 author-countries have increased more, relative to the 1-5 country group, whereas the group with 31-40 countries has become less frequent. Above 40 author-countries, however, has grown dramatically.
In the period up to 2013, there were only three articles with more than 50 author-countries; from 2014, however, these groups have significant numbers of articles, with some containing more than 100 country affiliations.

> There has been a great increase in articles with over 1,000 authors

Figure 1.
The frequency of articles by count of authors (left) and by count of countries identified in author affiliations (right), for 10 years (2009-2018) of *Web of Science* publication records. These results compare favourably to those of Waltman and van Eck (2015).

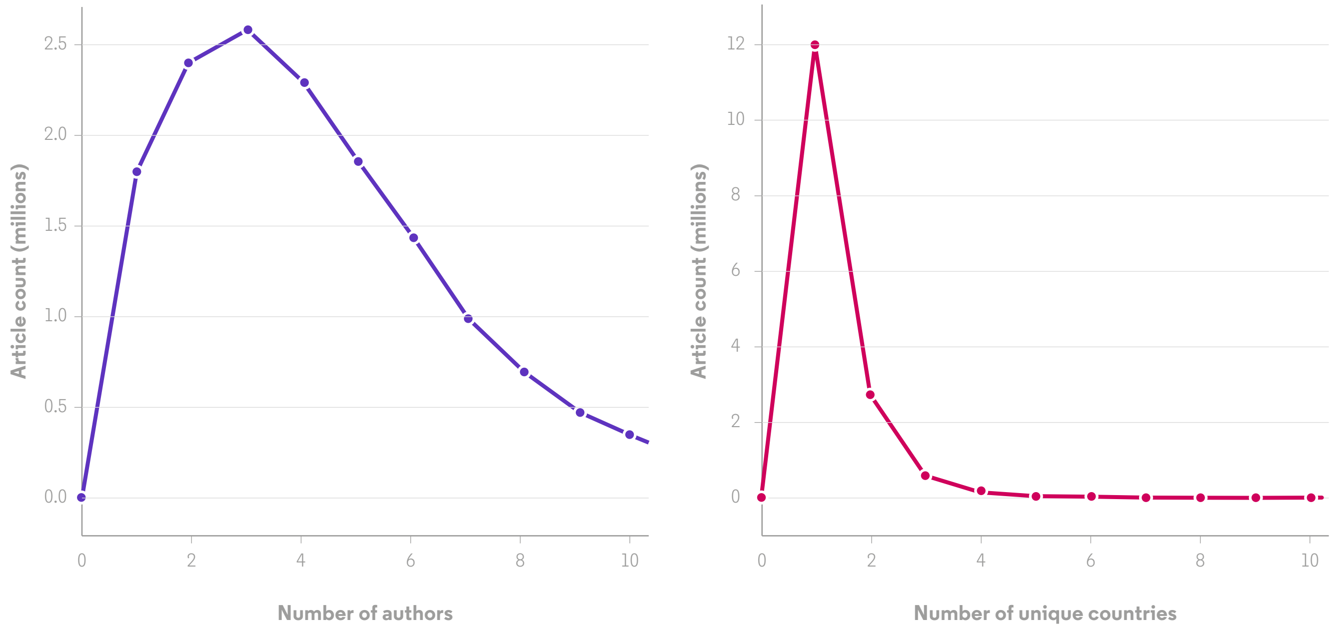
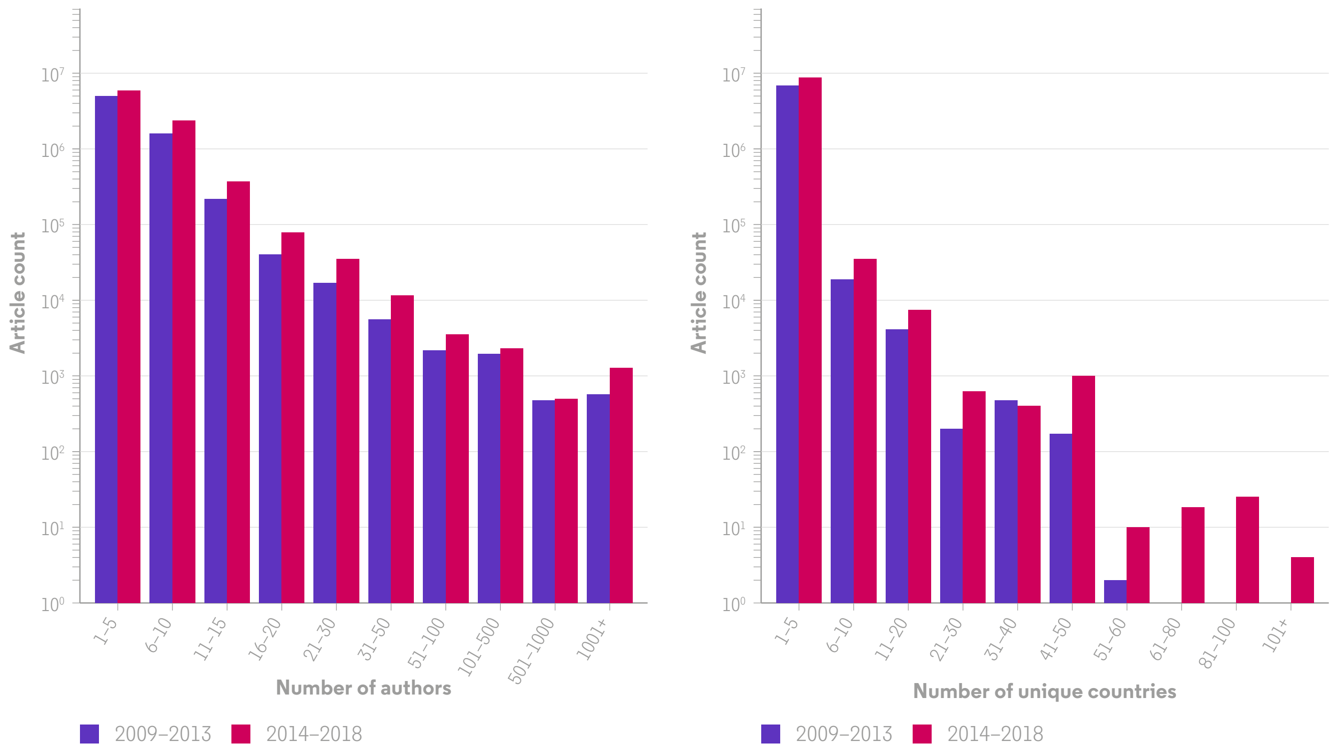


Figure 2.
Changing profiles of authorship by unique individuals and by unique country addresses. Data for articles indexed in *Web of Science* for 2009-2013 and 2014-2018.



Do all fields have multi-author publications?

The overall trend is clear: a relative growth in the numbers of multi-authored articles and a substantial increase in the numbers of articles with authors from many countries. Does this pattern extend across all fields or is it limited to specific research areas?

To explore this we assigned our *Web of Science* data to the major discipline categories in *Web of Science's Essential Science Indicators* (ESI). There are 21 categories covering broad areas like Chemistry and Physics plus a Multidisciplinary category, covering journals like *Nature* and *Science* (Table 1).

There are marked differences between categories. The most common authorship count remains 1-5 in most fields, but in the biomedical categories – some of which have more than one million articles published over the 10-year period – this has shifted towards higher counts.

For example, in Clinical Medicine, Microbiology and Molecular Biology more than 10% of articles have more than 10 authors and in Immunology this is almost 20%. That contrasts with related areas such as Plant & Animal Sciences where ~ 3% of articles have authorship of 10 or more.

Physics and Space Science stand out because they are categories where low authorship counts remain common (respectively 74% and 66% of articles with five or fewer authors) but where those articles with hyper-authorship are more abundant compared to other categories (0.6% and 1.5% with more than 50 authors, respectively).

These patterns suggest different research cultures between fields. The biomedical sciences appear to now commonly require large teams whereas Physics is still rooted in smaller, traditional teams while also supporting work dependent on massive 'hyper'-collaboration.

The Social Sciences largely publish with smaller authorship counts, but so do Mathematics and Engineering. In fact, the appearance of more than 1% of Social Sciences articles with authorship greater than 10 is perhaps something to watch as a marker of a shifting culture in an area where the individual and the small group have been the paradigm.

Table 1.
The relative frequency of author counts (shown as percentages within category) for articles indexed in *Web of Science* (2009-2018), grouped by *Essential Science Indicators* categories.

Count of articles 2009-18	ESI category	Count of authors									
		1–5	6–10	11–15	16–20	21–30	31–50	51–100	101–500	501–1000	1001–6000
184,499	Mathematics	99.2	0.8	0.0	0.0	0.0	0.0	0.00	0.000	0.000	0.000
316,589	Economics and Business	98.8	1.1	0.1	0.0	0.0	0.0	0.00	0.000	0.000	0.000
387,710	Computer Science	90.8	8.7	0.4	0.1	0.0	0.1	0.00	0.001	0.000	0.000
140,433	Social Sciences, general	90.0	8.7	1.0	0.2	0.1	0.3	0.01	0.004	0.000	0.000
466,600	Engineering	87.6	11.6	0.6	0.1	0.0	0.1	0.00	0.000	0.000	0.000
394,584	Psychiatry/Psychology	79.6	17.5	2.2	0.4	0.2	0.5	0.02	0.005	0.000	0.000
440,682	Geosciences	74.9	21.7	2.4	0.5	0.3	0.12	0.03	0.005	0.000	0.000
1,049,588	Physics	74.0	21.5	2.8	0.6	0.3	0.18	0.13	0.221	0.084	0.084
1,271,457	Environmental/Ecology	71.6	25.2	2.4	0.4	0.2	0.09	0.03	0.007	0.000	0.000
682,099	Plant and Animal Science	69.5	27.4	2.6	0.3	0.1	0.03	0.01	0.003	0.000	0.000
1,597,180	Chemistry	67.5	29.8	2.2	0.3	0.1	0.04	0.01	0.007	0.001	0.001
400,356	Agricultural Sciences	66.9	30.4	2.0	0.4	0.3	0.03	0.01	0.001	0.000	0.000
1,096,214	Space Science	66.3	20.4	5.7	2.5	2.3	1.37	0.80	0.650	0.028	0.028
442,270	Materials Science	65.6	31.5	2.6	0.2	0.0	0.00	0.00	0.000	0.000	0.000
661,887	Biology and Biochemistry	53.9	38.2	6.4	1.0	0.3	0.09	0.02	0.007	0.000	0.000
442,808	Neuroscience and Behavior	50.9	39.0	7.8	1.5	0.6	0.19	0.05	0.010	0.000	0.000
342,601	Pharmacology and Toxicology	50.4	41.6	6.8	0.9	0.3	0.05	0.01	0.004	0.000	0.000
2,541,166	Clinical Medicine	48.7	40.2	8.3	1.9	0.8	0.20	0.04	0.012	0.001	0.001
415,115	Microbiology	47.3	42.2	8.4	1.5	0.5	0.11	0.02	0.002	0.000	0.000
823,451	Molecular Biology and Genetics	42.1	40.7	11.7	3.1	1.5	0.59	0.20	0.092	0.001	0.001
214,950	Immunology	34.0	46.3	14.6	3.4	1.3	0.31	0.05	0.016	0.001	0.001
21,845	Multidisciplinary	62.1	28.1	7.0	1.7	0.8	0.29	0.07	0.023	0.000	0.000

Have all fields shown increasing international collaboration?

It would be expected that the most frequent count of countries of authorship would similarly have increased.

The data summarized in Table 2 may therefore be of some surprise as the country counts are dominated by the 1-5 country group, accounting for around 99% of articles in all categories except Immunology and Space Science. This reflects the extent to which the increase in the authorship count is driven by collaboration between groups rather than networks of individuals.

It is also in accord with Adams and Gurney (2018) who note that the majority of international collaborations are between two countries.

For example, just 1% of the UK's publications between 2002-2011 had co-authors from all of the USA, France and Germany (its most frequent partners). For the USA, the share of quadrilateral articles with frequent partners is even lower: around 0.1%.

That said, while the articles with multiple country counts are scarce in themselves, their presence across so many categories does confirm Cronin's (2001) view that very large global networks are now an established part of research management and publication output in many disciplines, not just in Physics and the biomedical categories.

Table 2.
Relative count of unique countries (shown as percentages within category) given as author affiliations on articles indexed in *Web of Science* (2009-2018), grouped by *Essential Science Indicators* categories.

Count of articles 2009-18	ESI category	Count of countries 1–5	6–10	11–20	21–30	31–40	41–100+
184,499	Mathematics	99.99	0.01	0.00	0.00	0.00	0.0000
442,270	Materials Science	99.95	0.05	0.00	0.00	0.00	0.0000
1,597,180	Chemistry	99.93	0.06	0.01	0.00	0.00	0.0018
466,600	Engineering	99.93	0.06	0.00	0.00	0.00	0.0000
316,589	Economics and Business	99.91	0.08	0.01	0.00	0.00	0.0000
387,710	Computer Science	99.89	0.10	0.01	0.00	0.00	0.0000
400,356	Agricultural Sciences	99.80	0.17	0.02	0.00	0.00	0.0002
661,887	Biology and Biochemistry	99.78	0.19	0.03	0.00	0.00	0.0003
140,433	Social Sciences, general	99.75	0.21	0.04	0.00	0.00	0.0012
342,601	Pharmacology and Toxicology	99.71	0.26	0.03	0.00	0.00	0.0000
394,584	Psychiatry/Psychology	99.66	0.26	0.07	0.01	0.00	0.0000
682,099	Plant and Animal Science	99.63	0.32	0.04	0.01	0.00	0.0000
415,115	Microbiology	99.56	0.39	0.05	0.01	0.00	0.0000
442,808	Neuroscience and Behavior	99.40	0.51	0.08	0.00	0.00	0.0007
440,682	Geosciences	99.32	0.61	0.07	0.00	0.00	0.0011
1,271,457	Environmental/Ecology	99.27	0.60	0.12	0.01	0.00	0.0017
2,541,166	Clinical Medicine	99.19	0.66	0.13	0.01	0.00	0.0025
1,049,588	Physics	99.18	0.39	0.24	0.01	0.07	0.1097
823,451	Molecular Biology and Genetics	99.09	0.71	0.17	0.02	0.00	0.0002
214,950	Immunology	98.65	1.10	0.22	0.02	0.01	0.0051
1,096,214	Space Science	93.77	5.12	1.02	0.08	0.01	0.0007
21,845	Multidisciplinary	99.43	0.50	0.07	0.00	0.00	0.0000

How does author and country collaboration affect citation rates?

Glanzel and Schubert (2004) identified a link between international collaboration and rising impact and Waltman and van Eck (2015: Figure 2) drew attention to a rise in the average citation impact for a journal article as the authorship count increased. Our data confirm this broad pattern. Both studies show that the effect of additional countries (up to around eight, in agreement with Adams and Gurney, 2018) is much more marked than additional authors (Figure 3).

The averages displayed in Figure 3 obviously hide a distribution of actual citation impact values. Note that the values shown here are Category Normalized Citation Impact (CNCI).

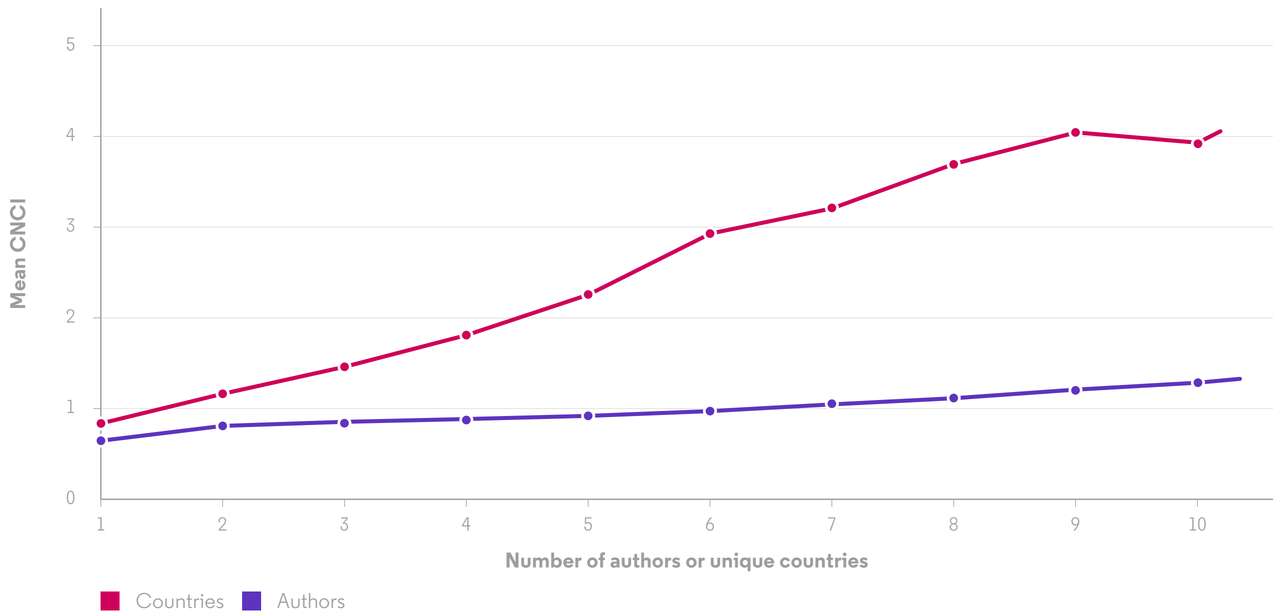
This is a conventional index used to normalize citation counts because counts rise over time at a rate that is discipline dependent (see BOX 1).

Box 1.

Citation impact. The citation of an article by later publications is an indication of that article's influence. In science and technology disciplines, there is a correlation between citation counts and peer-review indicators of research significance. Citation counts rise over time at rates that are discipline dependent. Older publications and life science publications are, on average, cited more often than recent papers in physical and social sciences.

To create a general index, the count for each article is 'normalized' against the average for that discipline category in the same publication year. This is referred to as Category Normalized Citation Impact (CNCI). The benchmark global average CNCI = 1.0

Figure 3.
The variation of Category Normalized Citation Impact (CNCI) related to the numbers of unique authors and unique countries identified in author affiliations. An additional country has a much greater effect on citation impact than an additional author.



Box 2.

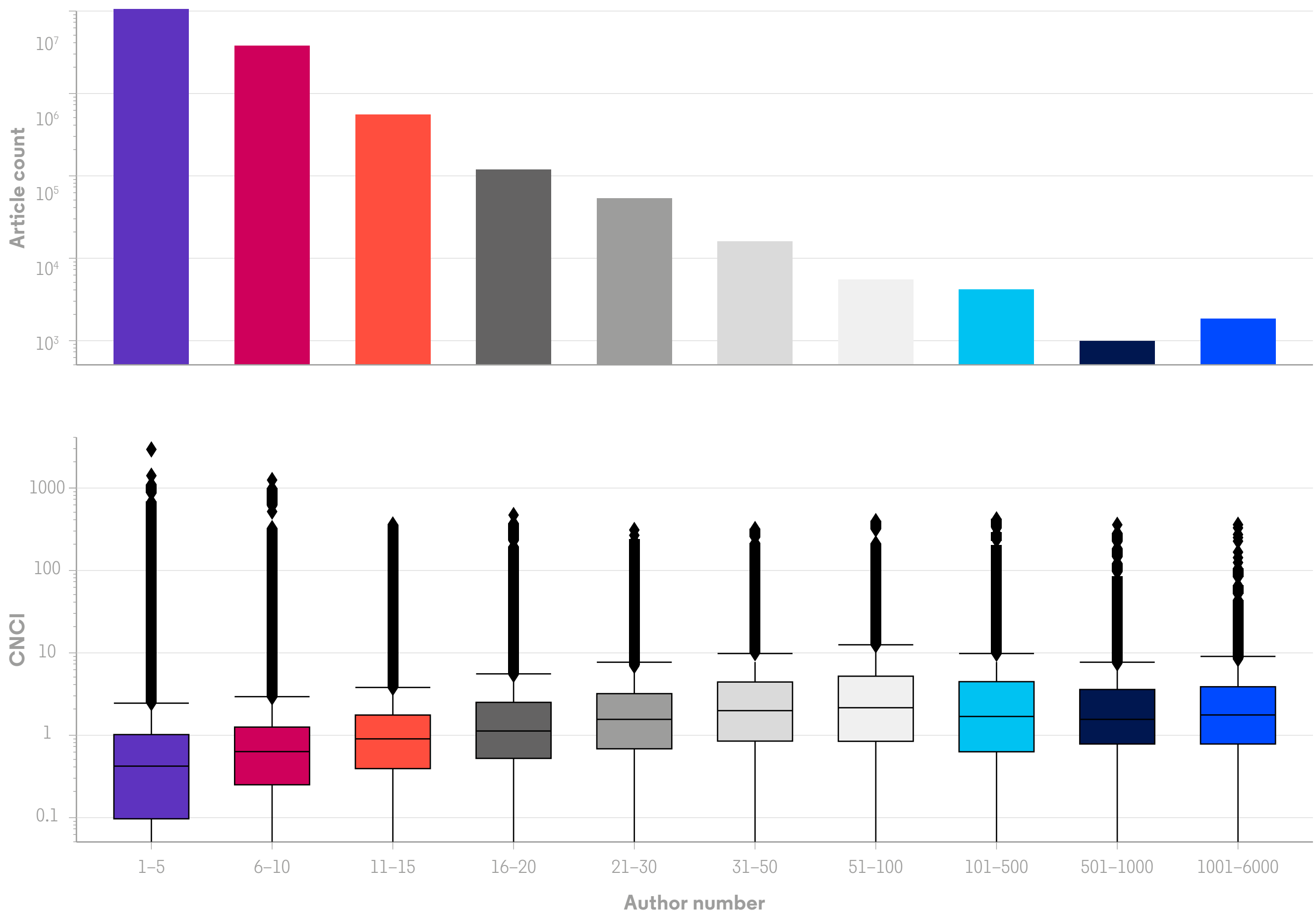
Boxplots. The analyses in Figure 4 (author counts) and Figure 5 (country counts) show the range of values for a variable while emphasising the core of the distribution. The coloured box spans the range between the upper and lower quartiles of the data distribution; the bar within it is the median for the range. The bar above the box marks 1.5 times the interquartile range (following Tukey: see McGill et al., 1978) and the values beyond that are then high outliers.

Figures 4 and 5 show the frequency distribution for the number of articles, in a bar chart, and CNCI (Category Normalized Citation Impact), in a boxplot, for different groups of author and country numbers, respectively.

Overall, the data distribution in successive authorship groups appear entirely coherent (Figure 4). There is no marked break-point between historically typical articles and any other group marked by greater authorship. We would not, on the basis of this analysis, point to any level above which the density of authorship makes the publication functionally 'different' for analytical purposes.

CNCI evidently does rise with author count. The average for each successive group slowly climbs to around 30 authors after which there is no marked average gain. However, the upper value at 1.5 times the interquartile range does continue to rise until around 50 authors. In fact, above 30 authors, the boxplot limit values and ranges are remarkably similar.

Figure 4.
The frequency distribution of grouped author counts across all categories (upper histogram) and the associated range of citation impact (CNCI) for each group (boxplot: see BOX 2)



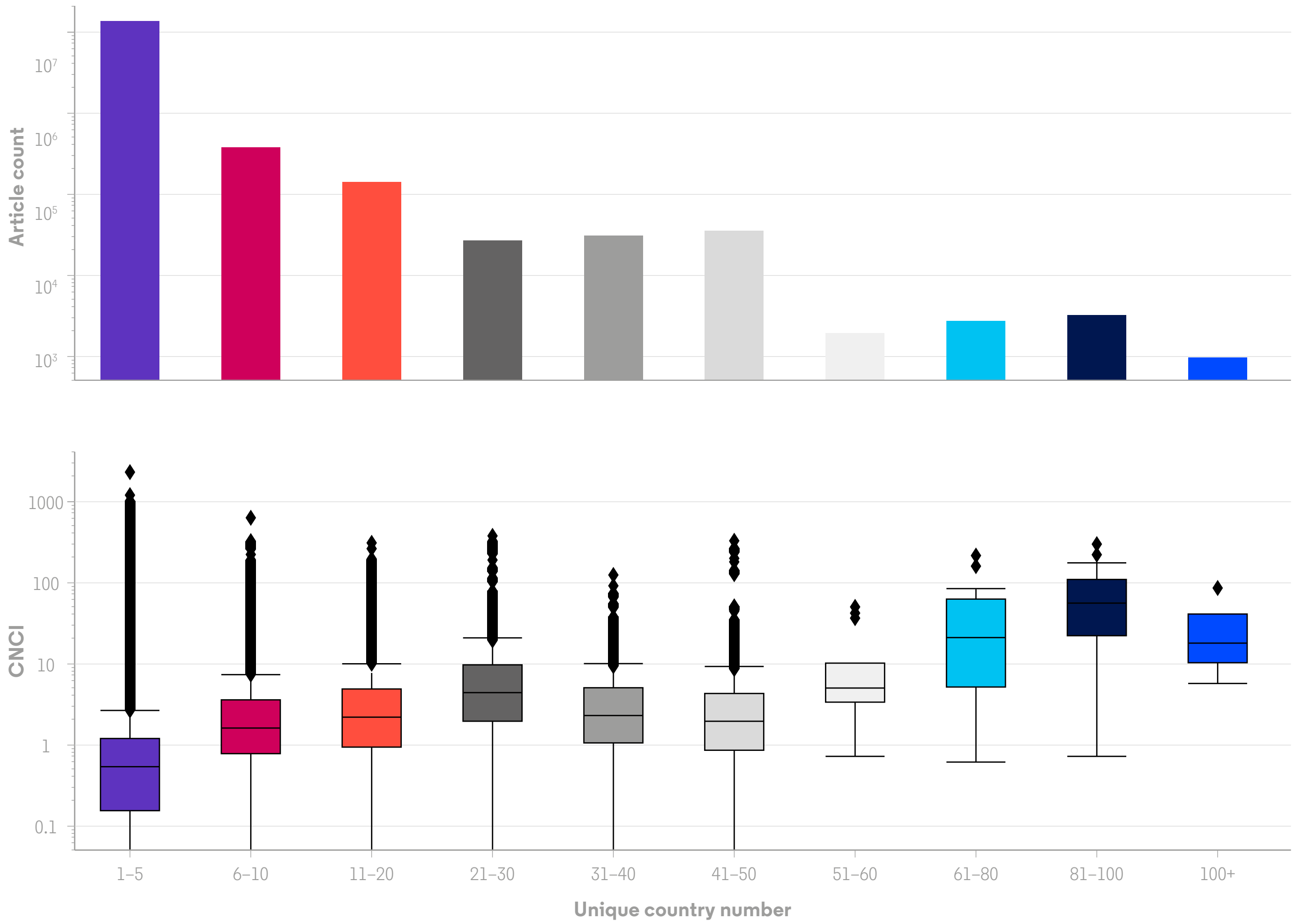
There are some notable outlier values with extraordinarily high CNCI (over 1,000 times the relevant world average) in the 1-5 author group. On preliminary inspection, the most highly-cited of these are identifiable as fundamental articles on widely-used methods (e.g. clinical practice guidelines, crystal structure refinement, image classification), but there may be cases where some other factor has come into play and this could emerge on detailed analysis.

The CNCI data in successive country-count groups (Figure 5) are relatively coherent up to around 30 countries (although it is again noticeable that in the 1-5 country group there are extreme outliers with very high CNCI), becoming increasingly erratic in range and threshold values above this value. The 31-40 group has lower threshold values and range while the 41-50 has several high outlier values. The larger groups, based on very small publication numbers, have a more sharply rising average CNCI up to the 100 country level. The lack of clear coherence in this series suggests that there may be an argument for excluding articles with high country counts from standard analysis. We will return to this point later.

Overall, the data, on average, confirm the point made at Figure 3: an additional country has a greater effect on CNCI than an additional author.

Impact rises with authorship but the effect of additional countries is greater than that of additional authors

Figure 5.
The frequency distribution of grouped country counts (upper histogram) and the associated range of citation impact (CNCI) for each group (boxplot: see BOX 2)

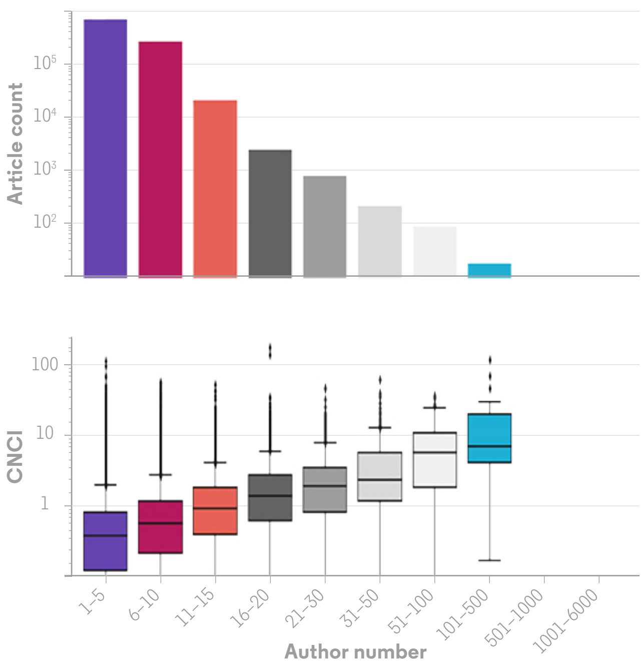


Is there a greater citation effect in some fields?

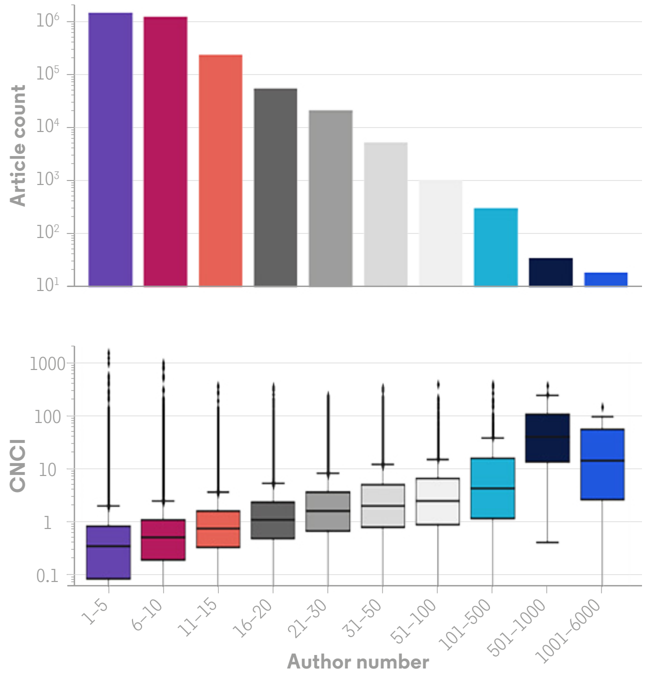
The shifts in collaborative authorship evidently vary by field. Table 1 showed that Plant & Animal Science might be taken as a mid-table reference category with a shift towards increased author counts in the 6-10 range but with about two-thirds of articles still having five or fewer authors. How does the variation in citation impact for authors and for countries compare with other categories?

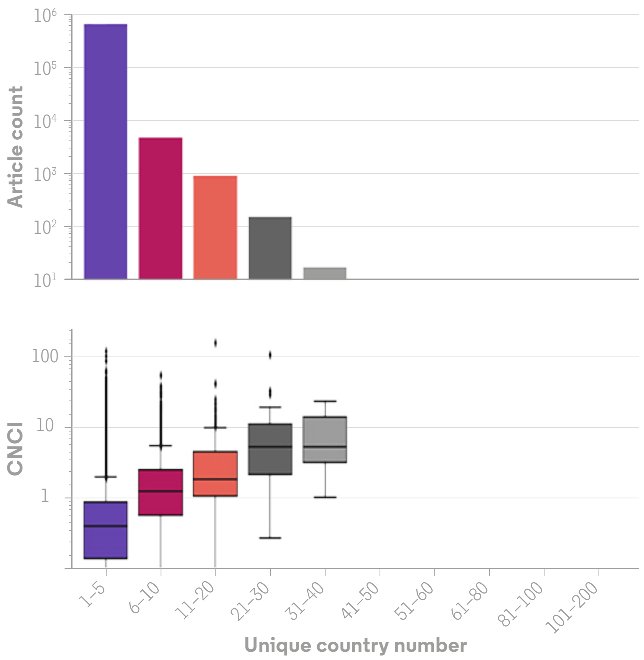
And how do these patterns change between the relatively broad categories of the *Essential Science Indicators* (ESI) and the finer level of *Web of Science* journal categories?

Plant & Animal Science



Clinical Medicine





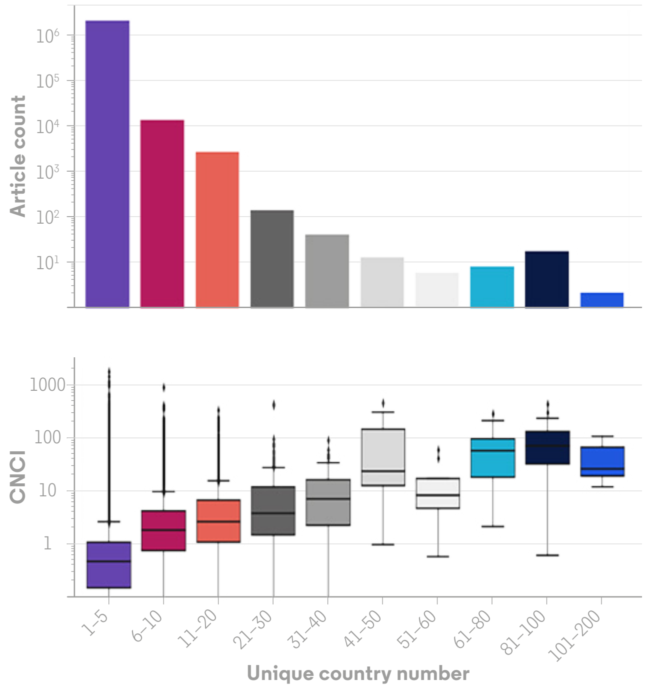
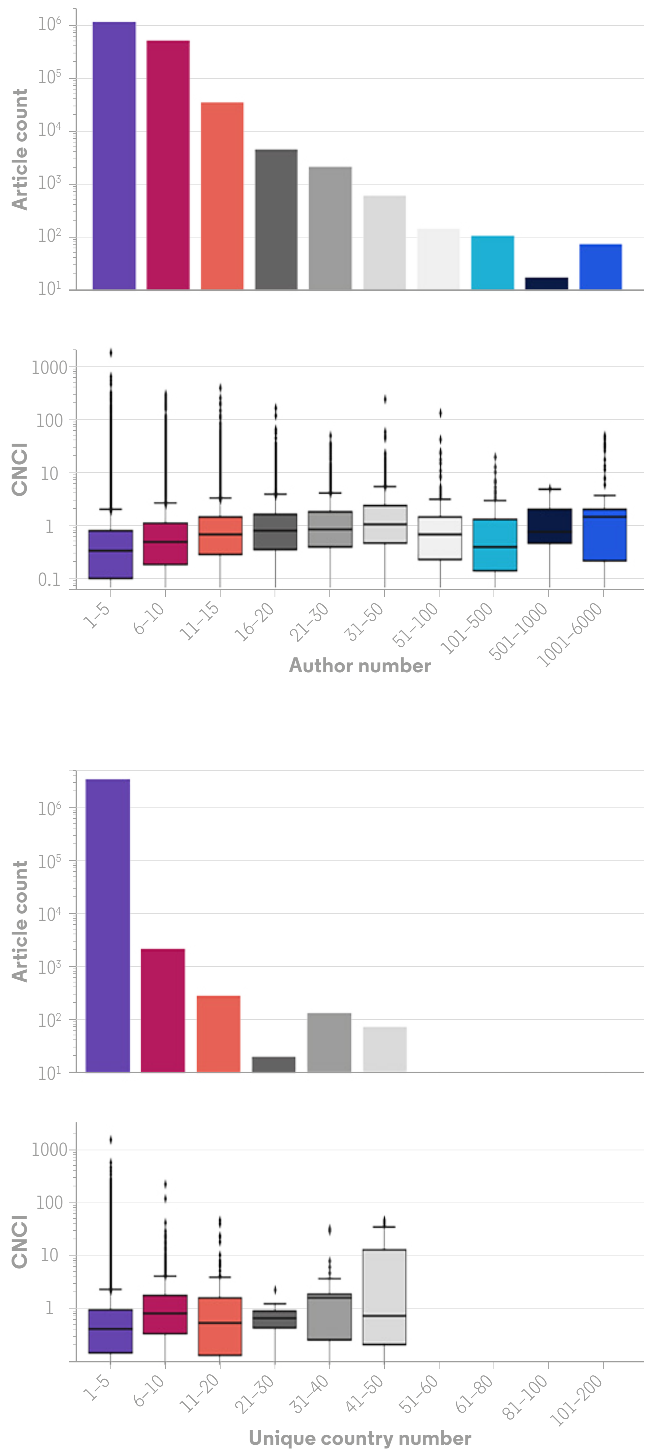


Figure 6.
The data illustrate the altering range of citation impact associated with changes in the frequency of authors and countries for a selection of disciplines at the broad level *Essential Science Indicators* journal categories.

The CNCI rise in Plant & Animal Science is coherent and progressive with increasing author and country count. Clinical Medicine clearly has a more extended scale at the upper ends of which CNCI variation becomes less coherent. Chemistry, by contrast, gains little CNCI with rising author count and almost none for additional countries whilst Mathematics has an idiosyncratic pattern associated with generally low author counts. This confirms a diversity of author typology between major disciplinary areas. The more extreme authorship trends in Physics and Medicine are not typical. This outcome is also important for analytical variations, such as fractional counting, which would have very different consequences in, for example, Medicine and Chemistry.

Chemistry



Mathematics

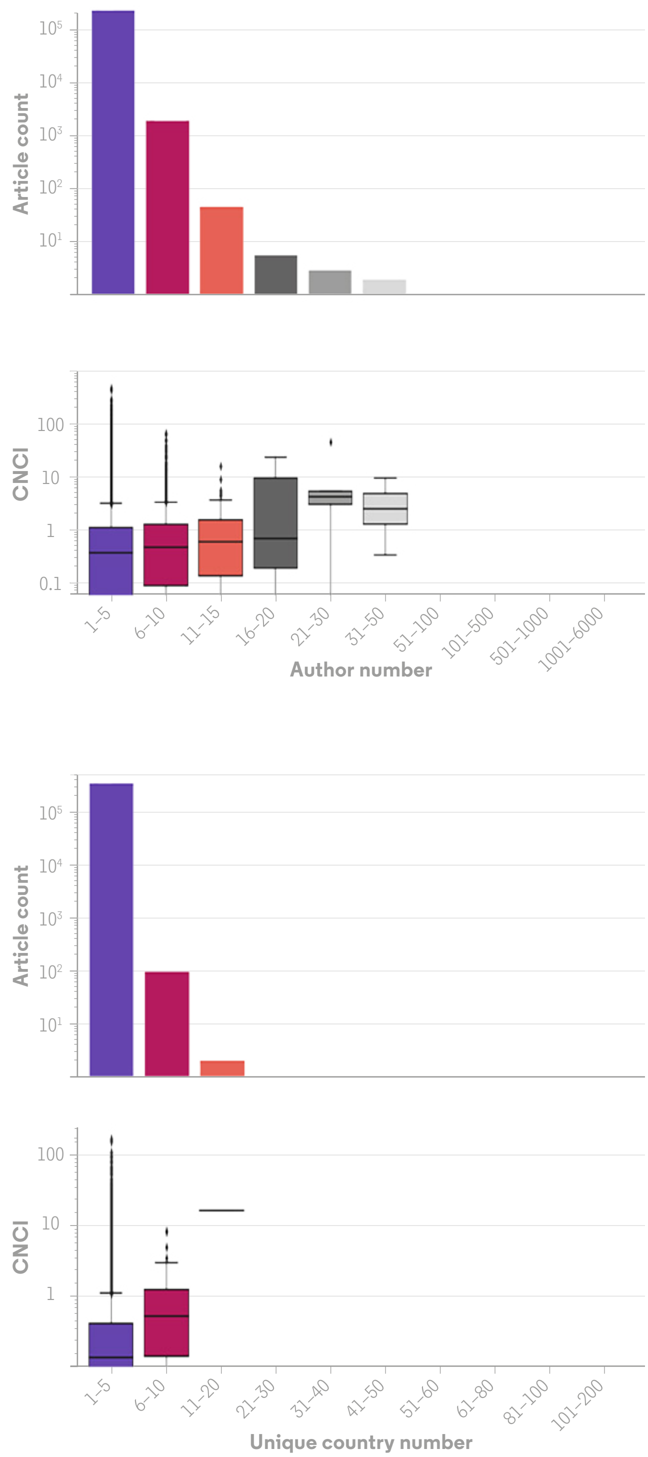
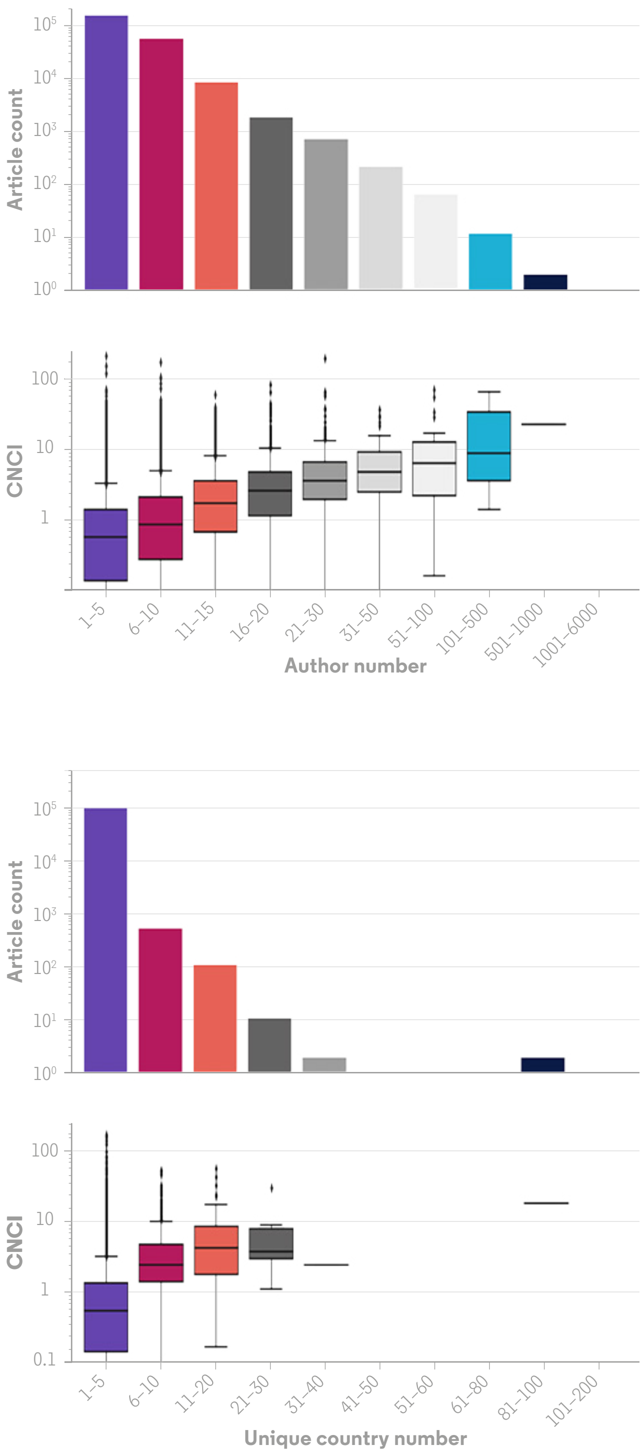


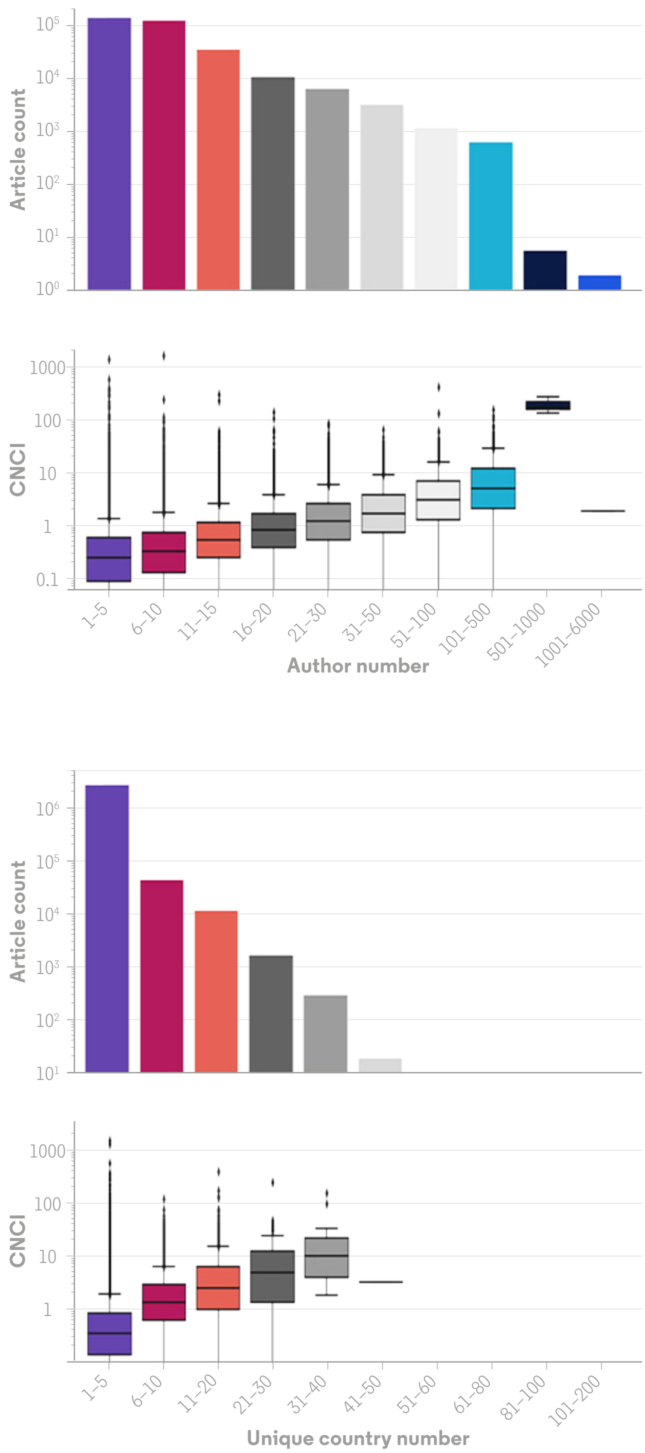
Figure 7.
The data illustrate the altering range of citation impact associated with changes in the frequency of authors and countries for a selection of disciplines at the fine-grained level of the 257 *Web of Science* journal categories.

The CNCI rise for Biology, as in ESI Plant & Animal Science, is coherent and progressive with increasing author and country count. Genetics & Heredity generally looks similar but the spike in authorship over 500 is very clear: the presence of one of these papers in a sample could be significant. Organic Chemistry is a truncated version of its parent ESI Chemistry category. There is no CNCI gain from multi-authorship and international collaboration is rare. By contrast, Physics, Particles & Fields has a spread of authorship and countries. However, whereas authorship count has almost no effect, massively international articles boost CNCI substantially.

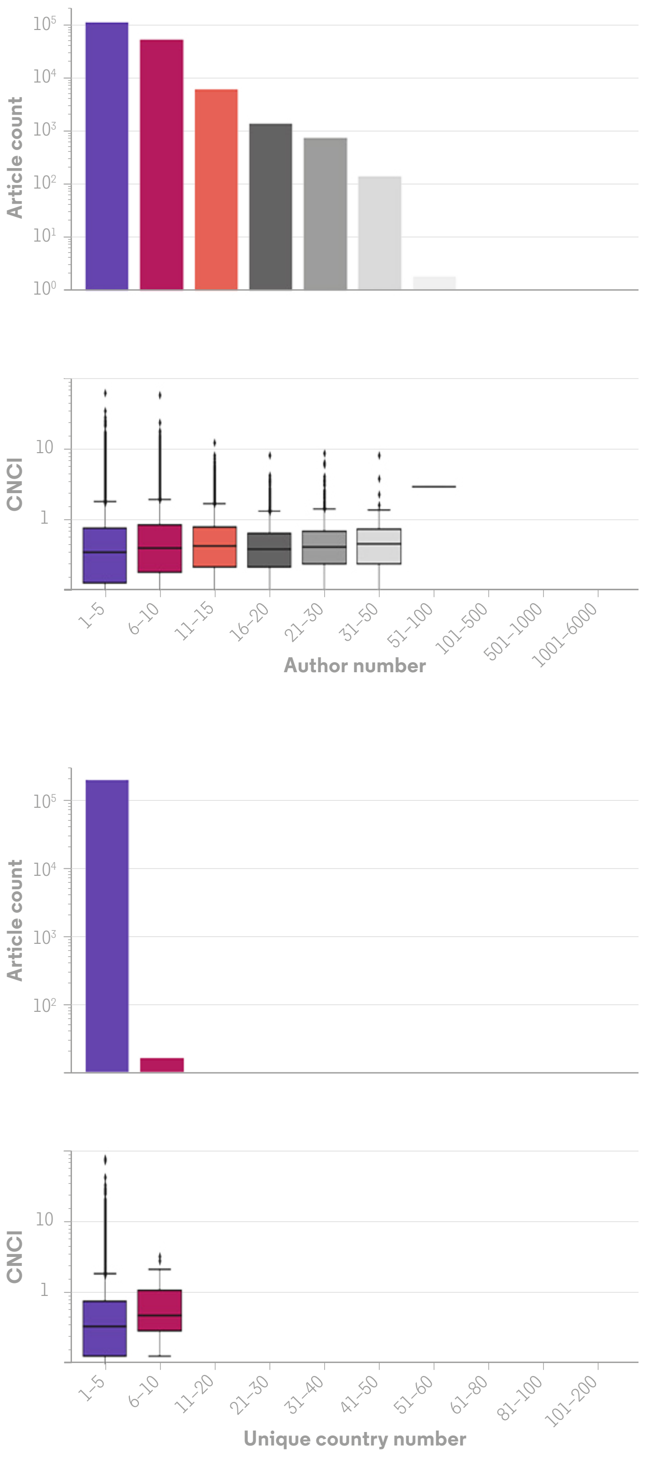
Biology CU-WoS



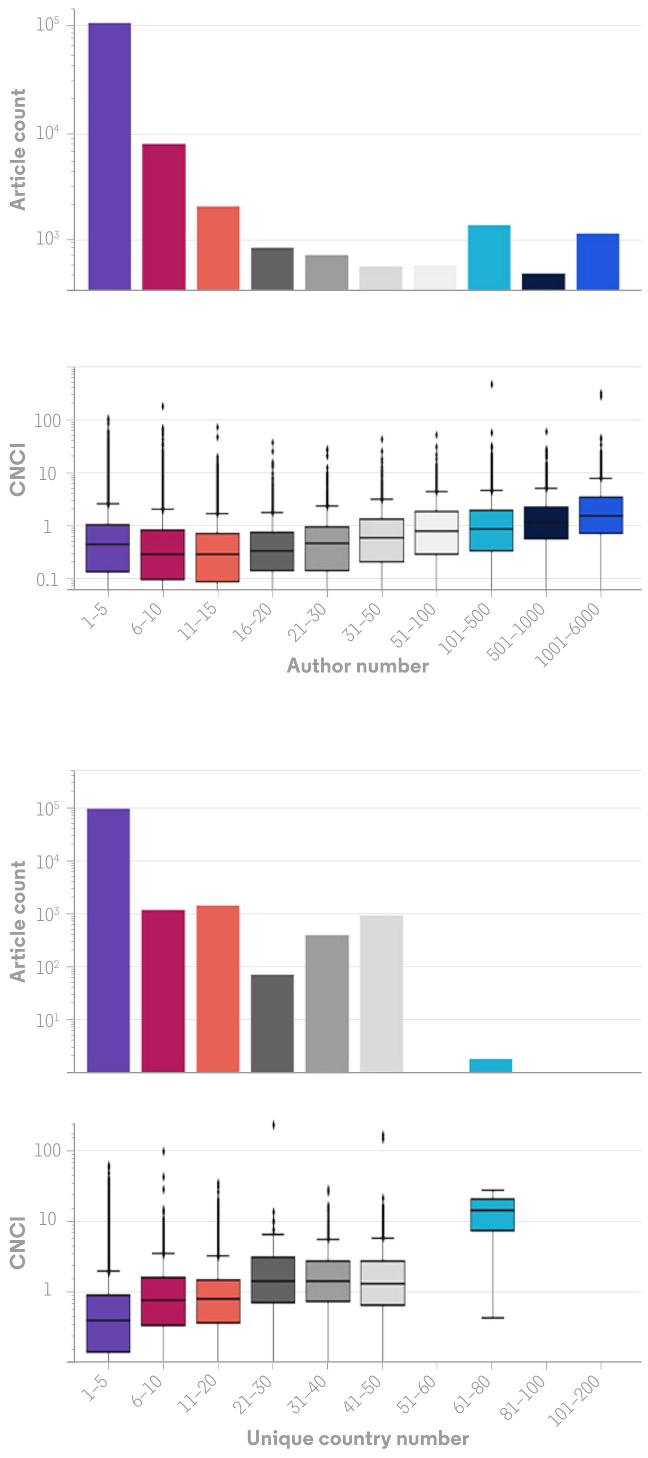
Genetics and Heredity, KM-WoS



Chemistry, Organic EE-WoS



Physics, Particles and Fields, UP-WoS



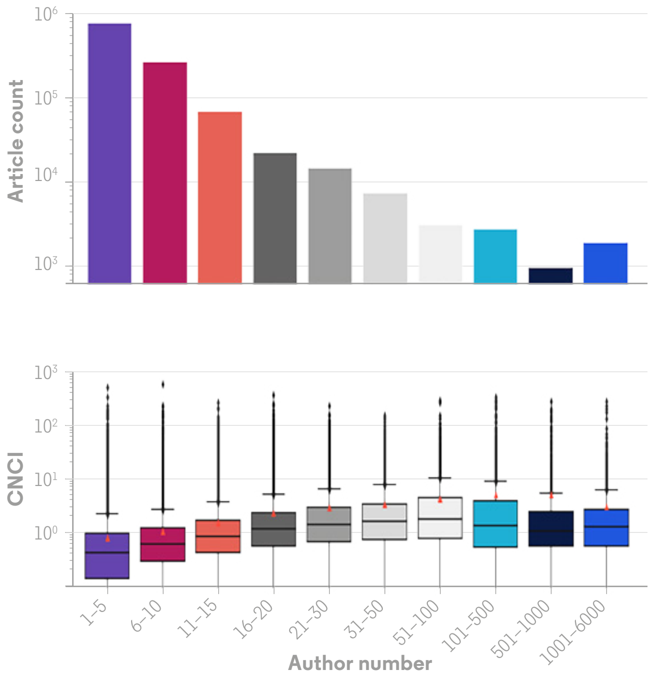
Are some countries affected more than others?

For larger research economies, illustrated by the UK and Germany, sets of articles grouped by the increasing authorship are reasonably large and the pattern is broadly similar (Figure 8). As authorship rises, average CNCI also rises to a maximum around 50 authors, beyond which average CNCI tends to fall slightly.

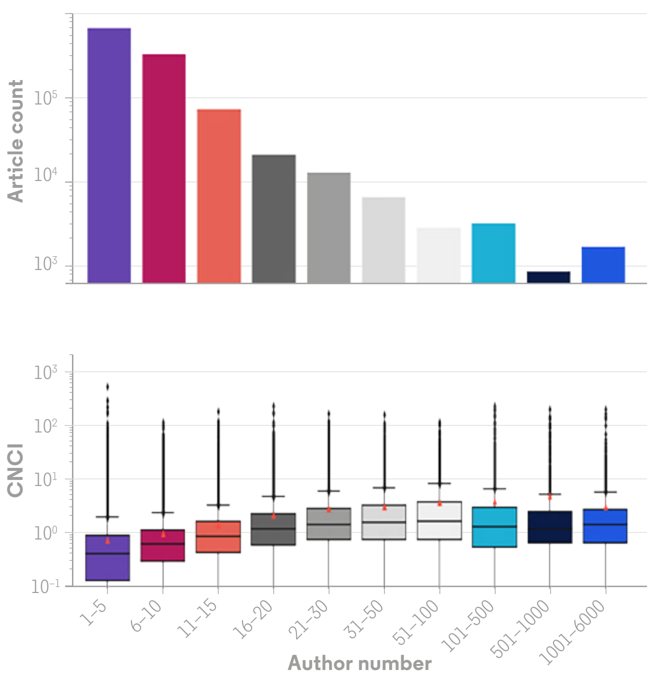
This pattern is not maintained amongst a geographical spread of other research economies. The profile for categories beyond the 50-author point may depend on the degree of involvement in either, or both, of particle physics and medical epidemiology. Bulgaria is linked into the European Research Area and its profile seems similar to Germany, although there are relatively more articles with low CNCI in the categories where author count is ten or fewer, until the 500-1,000 author category when average citation impact spikes upwards. That spike is also seen in the data for the other examples shown.

Figure 8.
The national variation in author count and CNCI (see BOX 1) for a geographical spread of six larger and smaller research economies. The data show the frequency of each grouped author counts across all research areas (upper histogram) and the boxplot (see BOX 2) shows the associated range of citation impact (CNCI) for each group. The triangles on the boxplots represent the mean.

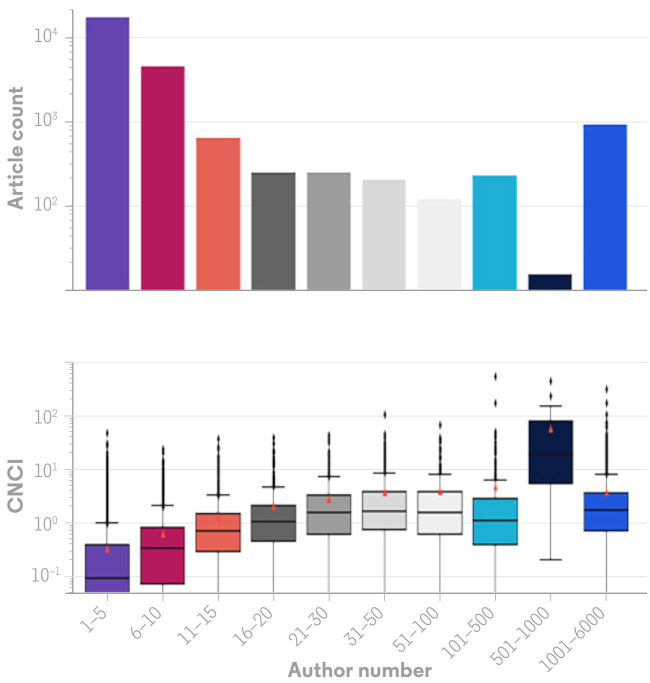
UK



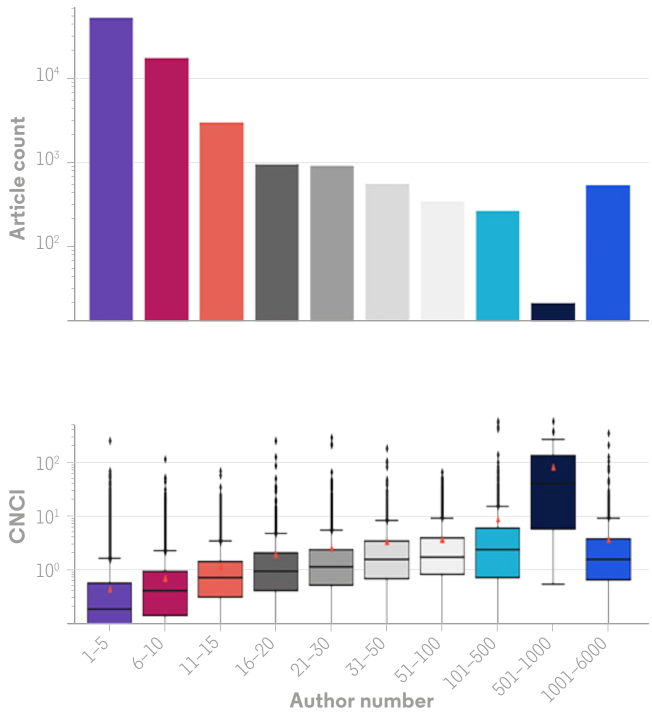
Germany



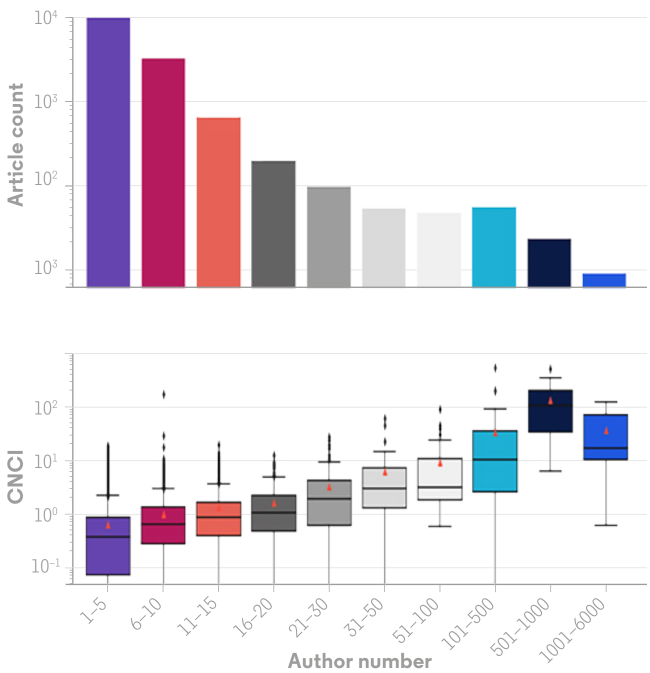
Bulgaria



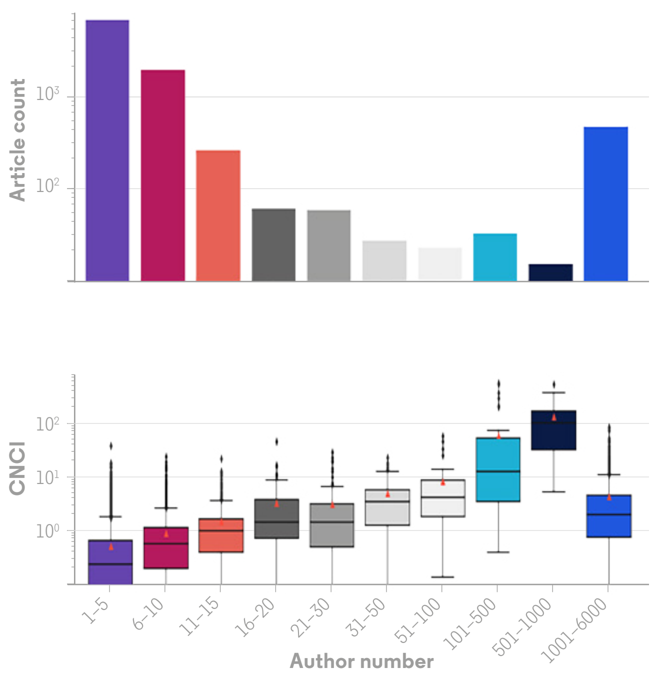
Chile



Ethiopia



Sri Lanka



The small proportion of articles with high author counts may attract relatively high citation counts in many categories, although the effect of this at national level is variable since the average CNCI for such articles is not always greater than that for articles with lower author counts. However, it is also evident that articles with high counts of countries in author address affiliations not only have high CNCI but that this may be erratically different from more conventional articles.

The effect of 'outlier' data points inevitably becomes greater when samples are smaller. Thus, for bibliometric analyses, this would become most apparent for smaller countries (as seen in Figure 8) and at institutional level. Such a possibility was identified by Must (2014). To test the effect of these high author/country count articles, we analyzed the likely citation impact of countries if their portfolio focused on the more frequent articles with lower authorship counts.

In fact, 94.98% of global articles published between 2009-18 had 10 authors or fewer, so we can look at this group as the 'typical' 95% of global low - perhaps better described as normal - authorship. The residual 5% of global articles with higher multi-authorship then form a second, comparator group.

For the large trans-Atlantic research economies, the 5% high-author count articles have around 2.5 times the CNCI of the 95% of more 'typical' articles with ten or fewer authors. These articles are about 10% of their output, because they have the capacity to contribute to many highly collaborative projects, and substantively raise their average national CNCI.

China is different. Its multi-author articles have twice the average CNCI of typical articles but they make up only 5% of national output (about 75% of China's publication output is purely domestic) and so raise the overall CNCI only marginally.

The smaller research economies are affected much more than the larger. The average CNCI of their 'typical' output is usually below world average. The average CNCI of the multi-author articles on which their researchers appear is higher than for the larger economies because they have few domestic multi-author articles and co-author relatively more often on multi-country publications. The effect of these articles is to raise their overall CNCI significantly, usually putting it above world average and in some cases doubling the headline figure.

The average CNCI for Sri Lanka is 0.65 on the 7,436 articles with normal authorship but almost 7.0 on the ~1,000 multi-authored articles raising the country's overall average CNCI above the UK and USA; it is the highest of any country in Table 3. It is essential to recognize, understand and describe this collaborative multi-author effect in order to interpret that headline average.

Table 3.
Count of journal articles (2009-2018) and average CNCI (see BOX 1) for four large and four smaller research economies. Counts are shown for each country's total articles, for the article count on which it has at least one author among the 95% of all global articles with authorship of ten or fewer, and for its count among the 5% of articles with greater authorship. CNCI values are then shown for each sub-set of articles plus the CNCI ratio between the most authored 5% and least authored 95% for that country. Green-red shows low-high values in each column.

	Articles (2009-18)			Citation impact (CNCI)			
	Total	Low 95%	High 5%	Overall	For low 95%	For high 5%	Ratio high/low
USA	3,964,964	3,644,184	320,780	1.32	1.20	2.69	2.24
China	2,469,444	2,334,272	135,172	1.02	0.97	1.89	1.94
UK	1,148,033	1,028,160	119,873	1.42	1.23	3.08	2.51
Germany	1,044,111	920,866	123,245	1.25	1.05	2.73	2.59
Chile	79,253	70,230	9,023	1.03	0.75	3.15	4.19
Bulgaria	29,119	25,523	3,596	0.80	0.45	3.34	7.49
Ethiopia	13,287	12,185	1,102	1.18	0.77	5.66	7.31
Sri Lanka	8,519	7,436	1,083	1.46	0.65	6.96	10.63

References

Aad, G. et al. (ATLAS Collaboration, CMS Collaboration) (2015). Combined measurement of the Higgs boson mass in pp collisions at root s=7 and 8 TeV with the ATLAS and CMS experiments. *Physical Review Letters*, 114, 191803, doi:10.1103/PhysRevLett.114.191803.

Adams J. (2012). Collaborations: the rise of research networks. *Nature*, 490, 335-336, doi:10.1038/490335a.

Adams, J. (2013). The Fourth Age of research. *Nature*, 497, 557-560, doi:10.1038/497557a.

Adams, J. and Gurney, K. A. (2018). Bilateral and multilateral coauthorship and citation impact: patterns in UK and US international collaboration. *Frontiers in Research Metrics and Analytics*, 3, 12, doi:10.3389/frma.2018.00012.

Adams, J., Rogers, G. and Szomszor, M. (2019). The Annual G20 Scorecard – Research Performance 2019. Clarivate Analytics, London UK. ISBN 978 1 9160868 3 8 https://clarivate.com/news/the-first-annual-g20-data-scorecard-report-highlights-the-research-performance-of-the-worlds-leading-economies/.

Bozeman, B., Fay, D. and Slade C. P. (2013). Research collaboration in universities and academic entrepreneurship: the-state-of-the-art. *Journal of Technology Transfer*, 38, 1-67, doi:10.1007/s10961-012-9281-8.

Constantian, M. B. (1999). The Gordian knot of multiple authorship. *Plastic and Reconstructive Surgery*, 103, 2064-2066.

Croll, R. P. (1984). The noncontributing author: an issue of credit and responsibility. *Perspectives in Biology and Medicine*, 27, 401-407, doi:10.1353/pbm.1984.0053.

Cronin, B. (2001). Hyperauthorship: A Postmodern Perversion or Evidence of a Structural Shift in Scholarly Communication Practices? *Journal of the American Society for Information Science and Technology*, 52, 558-569, doi:/10.1002/asi.1097.

Drenth, J. P. H. (1998). Multiple authorship – the contribution of senior authors. *Journal of the American Medical Association*, 280, 219-221, doi:10.1001/jama.280.3.219.

Economist (2016). Why research papers have so many authors. 24 November. https://www.economist.com/science-and-technology/2016/11/24/why-research-papers-have-so-many-authors.

Endersby, J. W. (1996). Collaborative research in the social sciences: Multiple authorship and publication credit. *Social Science Quarterly*, 77, 375-392.

Glänzel, W. and Schubert, A. (2004). Analyzing scientific networks through co-authorship. In Moed, H. F., Glänzel, W. and Schmoch, U. (eds.), *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems*, Dordrecht: Kluwer Academic Publishers, pp 257-276.

Hudson, J. (1996). Trends in multi-authored papers in economics. *Journal of Economic Perspectives*, 10, 153-158, doi:10.1257/jep.10.3.153.

Katz, J. S. and Martin, B. (1997). What is research collaboration? *Research Policy*, 26,1-18, doi:10.1016/S0048-7333(96)00917-1.

King, C. M. (2012). Multi-author papers: onward and upward. Science Watch. http://archive.sciencewatch.com/newsletter/2012/201207/.

Mallapaty, S. (2018). Paper authorship goes hyper. Nature Index. January 30. https://www.natureindex.com/news-blog/paper-authorship-goes-hyper.

McGill, R., Tukey, J. W. and Larsen, W. A. (1978). Variations of Box Plots. *The American Statistician*, 32, 12-16, doi:10.2307/2683468.

Must, U. (2014). The impact of multi-authored papers: the case of a small country. *Collnet Journal of Scientometrics and Information Management*, 8, 41-47, doi:10.1080/09737766.2014.916874.

NCD Risk Factor Collaboration (2017). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. *Lancet*, 390, 2627-2642, doi:10.1016/S0140-6736(17)32129-3.

Price, D. J. de S. (1963). *Little Science, Big Science*, Columbia University Press.

Waltman, L. and van Eck, N. J. (2015). Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics*, 9, 872-894, doi:10.1016/j.joi.2015.08.001.

About the Global Research Report series from the Institute for Scientific Information (ISI)

Our Global Research Reports draw on our unique industry insights to offer analysis, ideas and commentary to enlighten and stimulate debate.

Each one demonstrates the huge potential of research data to inform management issues in research assessment and research policy and to accelerate development of the global research base.

If you would like to receive news and reports from the Institute for Scientific Information, or to find out more about our work of ISI, please do get in touch.

e: ISI@clarivate.com

Download here:
www.webofsciencegroup.com/isi

Previous reports include:

GRR1: Profiles, not metrics

GRR2: The Plan S Footprint: Implications for the Scholarly Publishing Landscape

GRR3: Navigating the Structure of Research on Sustainable Development Goals

GRR4: The Annual G20 Scorecard Research Performance 2019

GRR5: Global Research Report South and Southeast Asia

About the Web of Science Group

The Web of Science Group, a Clarivate Analytics company, organizes the world's research information to enable academia, corporations, publishers and governments to accelerate the pace of research. It is powered by the *Web of Science* – the world's largest publisher-neutral citation index and research intelligence platform. Its many well-known brands also include *Converis*, *EndNote*, *Kopernio*, *Publons*, *ScholarOne* and the Institute for Scientific Information (ISI).

The 'university' of the Web of Science Group, ISI maintains the knowledge corpus upon which the index and related information and analytical content and services are built; it disseminates that knowledge externally through events, conferences and publications and it carries out research to sustain, extend and improve the knowledge base.

For more information, please visit webofsciencegroup.com.

@webofscience

webofsciencegroup.com/isi

WS419558643 / 01