R&D

Publications Output: U.S. Trends and International Comparisons

NSB-2021-4

October 28, 2021

This publication is part of the *Science and Engineering Indicators* suite of reports. *Indicators* is a congressionally mandated report on the state of the U.S. science and engineering enterprise. It is policy relevant and policy neutral. *Indicators* is prepared under the guidance of the National Science Board by the National Center for Science and Engineering Statistics, a federal statistical agency within the National Science Foundation. With the 2020 edition, *Indicators* is changing from a single report to a set of disaggregated and streamlined reports published on a rolling basis. Detailed data tables will continue to be available online.

Table of Contents

List of Tables

List of Figures

Executive Summary

Key takeaways:

- The United States remains a highly influential nation in science and engineering (S&E) research as measured by the volume of conference papers and peer-reviewed journal articles and the citations to those publications.
- The high-income economies (including the United States, the United Kingdom, Germany, France, and Japan) have steadily increased their large base of S&E publications, while lower-middle-income and upper-middle-income economies (including India and China) have rapidly increased production starting from a lower base of S&E publications in 1996.
- In 2016, China exceeded the publication output of the United States; in 2017, India became the third-largest producer.
- Beginning in 2017, articles from China received citations above the number expected, given the number of articles they produced.
- U.S. researchers have increased international collaboration over the last 15 years. Authors with an institutional address in China are the most frequent coauthors with authors whose institutional address is in the United States. In addition, the international citations by authors, whose institutional address is in the United States, most often cite authors with institutional addresses in the United Kingdom, Canada, and Australia.
- In 2020, major research countries—the United States, China, the United Kingdom, and many European Union countries—were hubs of global coronavirus research collaboration.

The primary method of disseminating research findings is through publication of conference proceedings and peer-reviewed journal articles (i.e., publication output). Data on publication output indicate an increase in global research activity, a growth in the involvement and scientific capabilities of middle-income countries, and an internationally connected research ecosystem.

Publication output reached 2.9 million articles in 2020, based on data from the Scopus database of S&E publications. The countries with the largest volume of S&E publications in 2020 were China, with 23% of global output, and the United States, with 16%. The compound annual growth rate of publication output has increased in recent years. The rate was 5% over the last 4 years (2017 to 2020) but was 4% over the longer 11-year period (2010 to 2020).

Internationally collaborative research continues to grow. International collaborations increased in 2020, with nearly one in four articles having coauthors from multiple countries as compared to one in seven in 2000. Collaboration on articles has also grown among relatively small producers, such as Saudi Arabia, Pakistan, and Argentina.

Data on article citations show that the United States and other nations, such as the United Kingdom, the Netherlands, and Switzerland, are among the countries producing highly impactful research papers. The impact of U.S S&E articles has remained steady over the last 24 years, while highly cited articles from China grew dramatically.

Introduction

Research publications and presentations at conferences represent the main mechanisms for disseminating research findings. Presentations are represented in the published research literature as conference proceedings. Published literature is an indicator of scientific activity and global research partnerships. Additionally, analysis of how the published literature is cited provides insight into the impact of research output. Scientific publications are not merely an exercise of ivory tower academics but serve as a key linkage enabling public uses of scientific output (Yin et al. 2021).

This report presents data on research publication output by country, scientific field, international collaboration, and impact measures. The first section examines comparative country data on publication output across science and engineering (S&E) fields. It also includes two sidebars: (1) publications by members of underrepresented groups and the impact on the research and development (R&D) workforce, and (2) measuring cross-disciplinary publication output. The second section focuses on collaboration between researchers in the United States and other regions, countries, and economies through examining coauthoring and citation patterns. This section includes a sidebar on the 2020 coronavirus publication output and collaboration network. The third section provides an analysis of scientific impact as measured by citations in research publications.

The analysis reported here is based on counting publications and citations using bibliometric data in Scopus, a database of scientific literature with English-language titles and abstracts (Science-Metrix 2021a). There are benefits and limitations to counting publications and citations using bibliometric data as an indicator of research output. A benefit is that this approach provides comparable information for analyzing research output across countries. A potential limitation is that country-specific incentive payments for academic publications are not considered (Franzoni, Scellato, and Stephan 2011). Additional limitations are the lack of measurement for the amount of research contained in each article and the contributions of associated data sets (Sugimoto and Larivière 2018).

There are two potential sources of bias in the data: inclusion of non-peer-reviewed articles, and a bias toward English-speaking countries because Scopus has a requirement for articles to contain an English-language title and abstract. The first bias is mitigated by removing articles published in journals that lack substantive peer review, sometimes referred to as *predatory journals* (Grudniewicz et al. 2019). This filtering removed 1 million research articles and conference papers from 2008 to 2020 (see **Technical Appendix**). The potential for a bias toward English-speaking countries is more difficult to solve. One solution undertaken by Elsevier is to increase publications from non-English-speaking countries. Specifically, Scopus has increased Chinese publications 624% from 1996 to 2020 (see **Technical Appendix**).[1]

Over 44 million articles published from 1996 to 2020 are analyzed. The articles include conference papers and research articles (collectively referred to as *articles*) published in conference proceedings and peer-reviewed scientific and technical journals. The articles exclude editorials, errata, letters, and other material that do not typically present new scientific data, theories, methods, apparatuses, or experiments. The articles also exclude working papers and preprints, which are not generally peer reviewed yet.

Articles with authors working in multiple countries are used for both counting publication output by country and for determining international collaborations. The country is determined by the institutional address of each author as listed in the article. For counting country output, each country receives a fractional contribution based on the number of authors. For determining international collaboration, each country or region represented by one or more authors is counted once. Because whole counting is used for international collaboration and fractional counting is used for publication output, those values are not directly comparable.

Assignment of articles to S&E fields uses the 14 fields of science in the National Center for Science and Engineering Statistics (NCSES) Taxonomy of Disciplines (TOD) (Science-Metrix 2019). The categorization is done by first assigning the journal to one of the 176 subfields in the Science-Metrix classification and then to the TOD. This approach works well for most journals and fields; for example, all of *dentistry* gets assigned to *health sciences*. Challenges arise for subfields that are more general, such as *energy*, and multidisciplinary journals, such as *Science* or *Nature*. For these fields and journals, classification occurs at the article level based on an algorithm using author affiliations, the names of journals referenced in the bibliography, the titles of the references, the publication's abstract, the publication's author-defined keywords, the publication's title, and the scientific field of references.

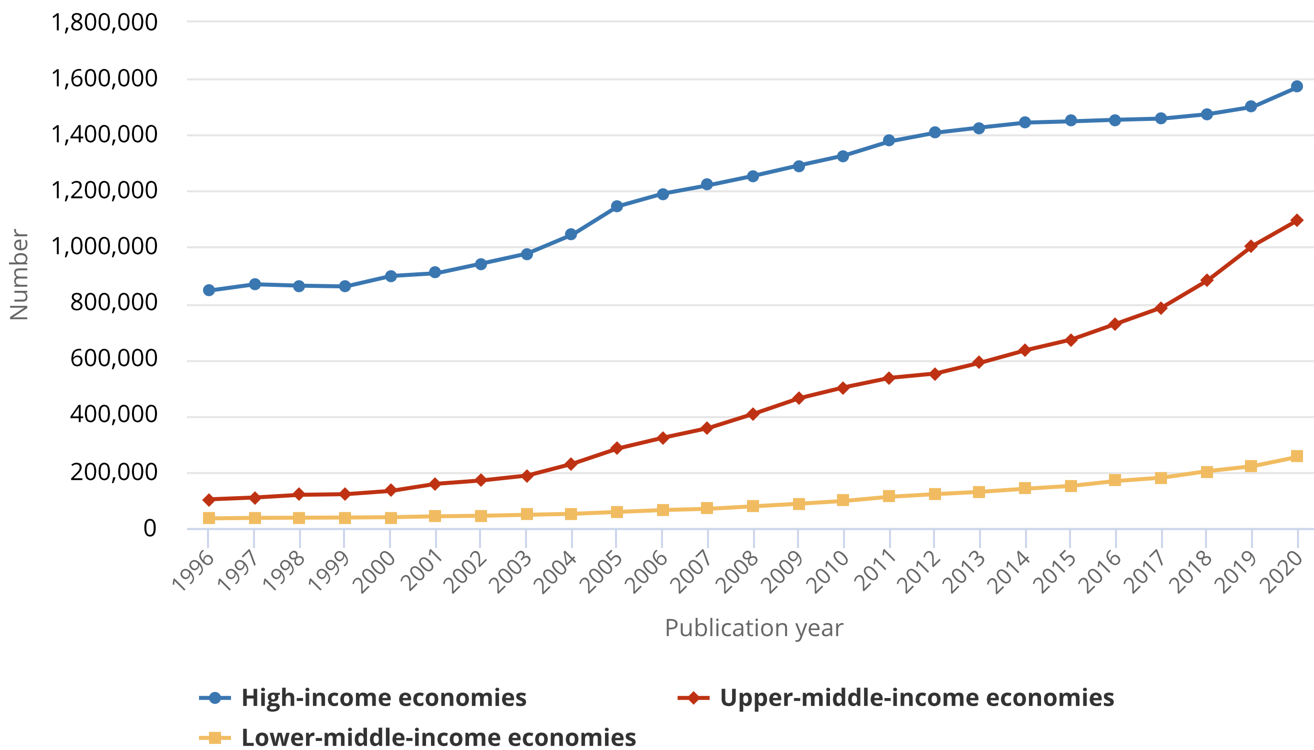
Publications-related data are best viewed as trends. Year-to-year differences are often not indicative of a pattern due to the process by which the information is indexed in Scopus. Additional details regarding document selection, limitations, and sources of bias are available in the **Technical Appendix**.

Publication Output by Country, Region, or Economy and Scientific Field

Publication output reached 2.9 million articles in 2020 with over 90% of the total from countries with high-income and upper middle-income economies (Figure PBS-1).[2] Since 1996, output has consistently grown for countries with high-income economies, such as the United States, Germany, and the United Kingdom (UK), expanding from a large base number of publications (Table SPBS-2).[3] Countries with upper-middle-income economies, such as China, Iran, Russia, and Brazil, have had a more rapid pace of growth since 1996, expanding from a relatively smaller base number of publications. Overall, the publication compound annual growth rates of countries with upper middle-income and high-income economies have been 10% and 3%, respectively, for the 25-year period covering 1996–2020 (Figure PBS-1).

Figure PBS-1

S&E articles, by income group: 1996–2020



Note(s):
Article counts refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a fractional count basis (i.e., for articles produced by authors from different countries, each country receives fractional credit on the basis of the proportion of its participating authors). Data are not directly comparable to *Science and Engineering Indicators 2020*; see the Technical Appendix for information on data filters. Low-income economies are not included in this figure because of their low publication output. Data by country and income groups are available in Table SPBS-2.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed May 2021; World Bank Country and Lending Groups, accessed March 2021.

Science and Engineering Indicators

More recently, the compound annual growth in publication output for the world was 4% from 2010 to 2020 (Table PBS-1). Country-specific growth rates vary widely by country. Among the 15 largest publication producers, countries with compound annual growth rates above the world average were Russia (10%), Iran (9%), India (9%), China (8%), and Brazil (5%); those with the lower growth rates were Japan (-1%), France (-0.3%), the United States (1%), the UK (1%), and Germany (1%).[4] The countries with low growth rates are those that built their scientific capacity decades ago and continue to maintain their scientific research. The worldwide growth of publication output, from 1.9 million in 2010 to 2.9 million in 2020, was led by four geographically large countries. China (36%), India (9%), Russia (6%), and the United States (5%) together accounted for about half the increase in publications over this time period.

Table PBS-1

S&E articles in all fields for 15 largest producing regions, countries, or economies: 2010 and 2020

(Number and percent)

Rank	Region, country, or economy	2010	2020	2020 world total (%)	2020 cumulative total (%)
na	World	1,938,121	2,940,807	na	na
1	China	308,769	669,744	22.77	22.77
2	United States	409,512	455,856	15.50	38.28
3	India	60,555	149,213	5.07	43.35
4	Germany	97,255	109,379	3.72	47.07
5	United Kingdom	94,081	105,564	3.59	50.66
6	Japan	108,534	101,014	3.43	54.09
7	Russia	33,855	89,967	3.06	57.15
8	Italy	58,252	85,419	2.90	60.06
9	South Korea	50,224	72,490	2.46	62.52
10	Brazil	41,501	70,292	2.39	64.91
11	France	68,300	66,479	2.26	67.17
12	Canada	56,445	65,822	2.24	69.41
13	Spain	49,031	65,638	2.23	71.64
14	Australia	41,661	60,891	2.07	73.71
15	Iran	24,694	57,755	1.96	75.68

na = not applicable.

Note(s):
The countries or economies are ranked based on the 2020 total. Article counts refer to publications from conference proceedings and peer-reviewed journal articles in S&E and indexed in Scopus (see Technical Appendix for more details). Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a fractional count basis (i.e., for articles from multiple countries or economies, each country or economy receives fractional credit on the basis of the proportion of its participating authors). Detail may not add to total because of countries or economies that are not shown. Proportions are based on the world total excluding unclassified addresses (data not presented). Details and other countries are available in Table SPBS-2.

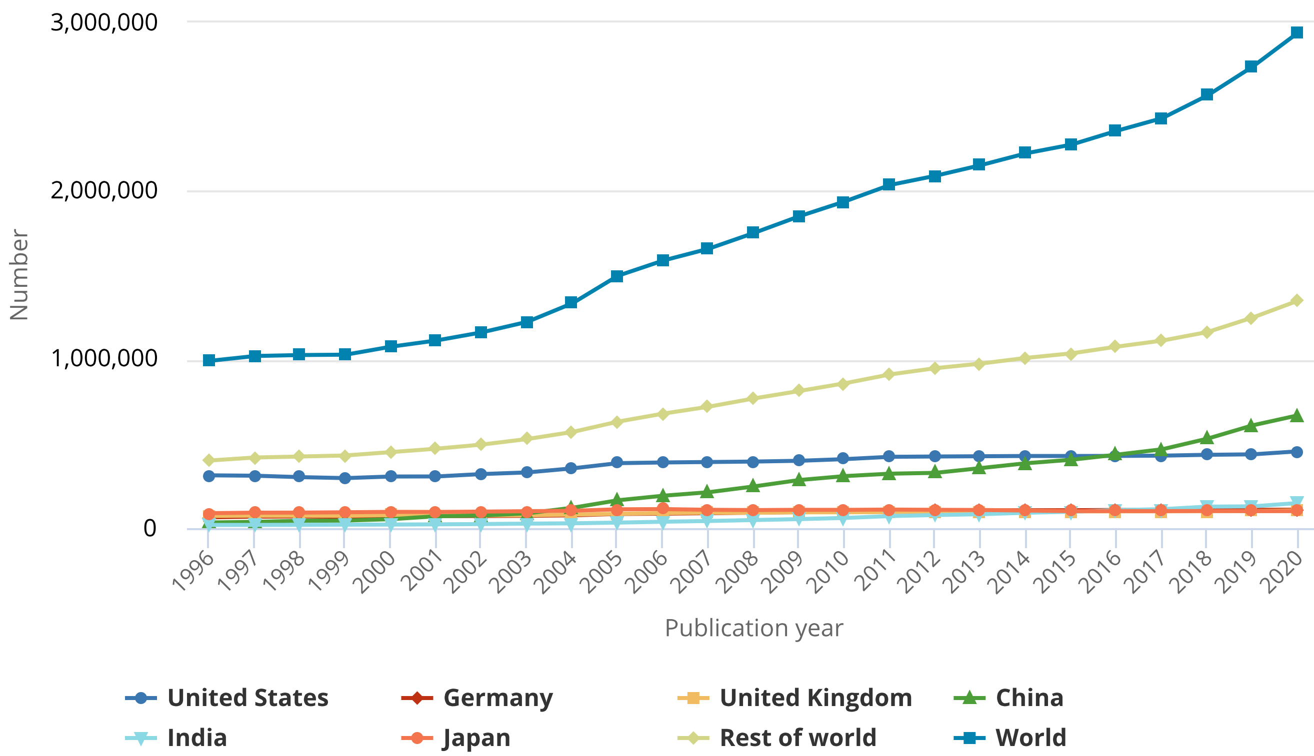
Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed May 2021.

Science and Engineering Indicators

Collectively, the top 15 countries produced 76% of the world's publication output of 2.9 million articles in 2020 (Table PBS-1).[5] The two countries producing the most S&E publications in 2020 were China (669,744, or 23%) and the United States (455,856, or 16%) (Figure PBS-2). With the exception of Iran replacing Taiwan beginning in 2014, the top 15 producers of S&E articles have been the same over the last 10 years (NSB 2016).

Figure PBS-2

S&E articles, by selected region, country, or economy and rest of world: 1996–2020



Note(s):
Article counts refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a fractional count basis (i.e., for articles produced by authors from different countries, each country receives fractional credit on the basis of the proportion of its participating authors). Data for all regions, countries, and economies are available in Table SPBS-2.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed May 2021.

Science and Engineering Indicators

The U.S. trend of moderate but increasing publication output varies by state. The National Science Board's (NSB's) **State Indicators data tool** provides state-level data based on each state's doctorate population and R&D funding, including academic S&E article output per 1,000 science, engineering, and health doctorate holders in academia (NSB 2021a) and academic S&E article output per $1 million of academic S&E R&D (NSB 2021b).

The U.S. trend of publication output varies across race or ethnicity and sex, which impacts R&D careers (see sidebar **Publication Output by Underrepresented Groups and Impact on R&D Careers** and *Indicators 2022* report "[2022] The STEM Labor Force of Today: Scientists, Engineers, and Skilled Technical Workers").

SIDEBAR

Publication Output by Underrepresented Groups and Impact on R&D Careers

The National Science Board stated in its Vision 2030 report that "women and underrepresented minorities remain inadequately represented in S&E relative to their proportions in the U.S. population" (NSB 2020). These disparities have also been found in the publication of peer-reviewed articles (Hopkins et al. 2013). The National Center for Science and Engineering Statistics (NCSES) has undertaken research to examine linkages between publication output and careers in research (Chang, White, and Sugimoto forthcoming).

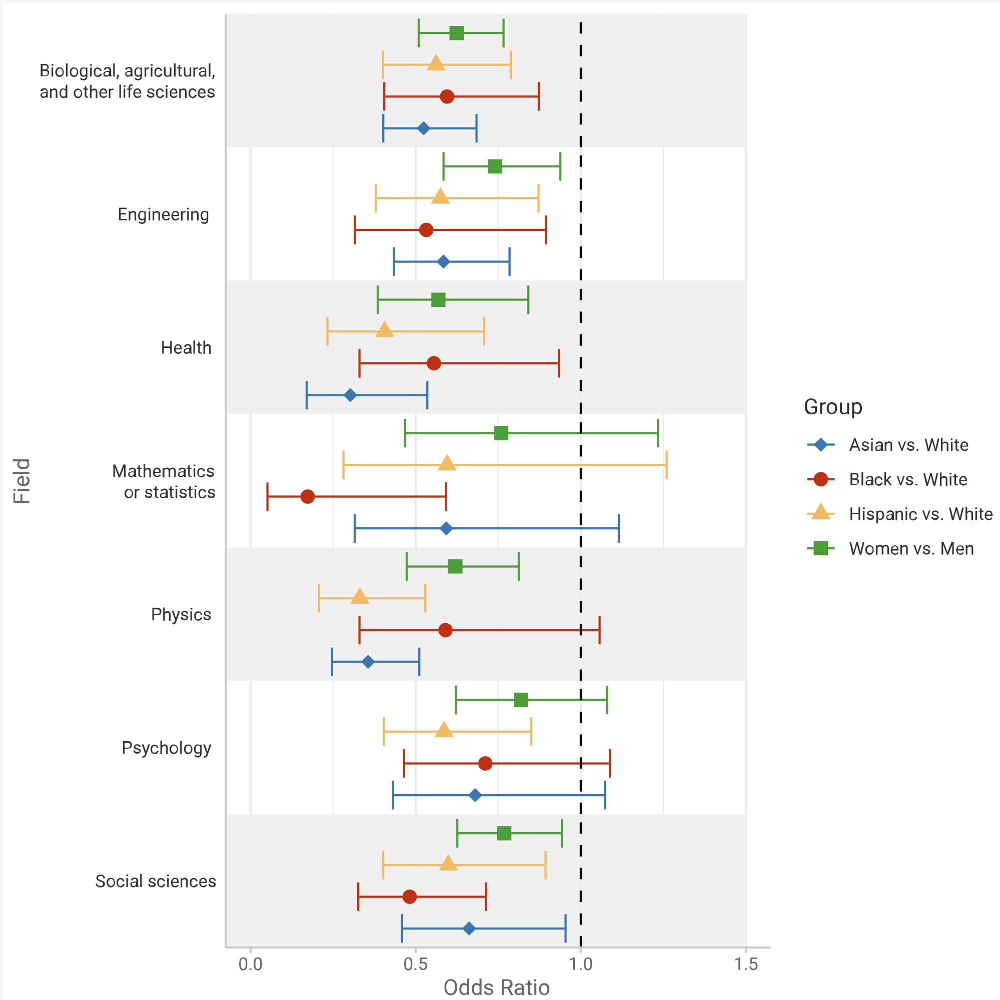
Matching publication output data to demographic survey data provides a key to understanding publication output in conjunction with authors' demographic, training, and career information. Prior researchers have attempted to add author demographics using various methods, such as sex and race disambiguation algorithms (e.g., NamSor, Ginni, Ethnicolr, OriginsInfo), that estimate the probability of race or sex from given names (or, in the case of Face++, from images). The accuracy of these matches varies dramatically by country and field; sex disambiguation algorithms perform better for western countries and poorly for other countries, specifically in Asia and South America (Karimi et al. 2016). In addition, some scientific fields, such as astronomy and astrophysics, generally use initials rather than given names. Despite these limitations, researchers have observed sex and race disparities in publication output (Hopkins et al. 2013; Larivière et al. 2013; Marschke et al. 2018; and **NSB Indicators 2018**: **S&E Publication Patterns, by Gender**).

The limitations associated with the earlier approaches can be overcome using data directly collected from the authors. One such source is the NCSES Survey of Doctorate Recipients (SDR),[*] which provides demographic, education, and career history information from a sample of individuals with a U.S. research doctoral degree in a science, engineering, or health field (NCSES 2021). Clarivate, the architect of Web of Science (WoS),[†] matched SDR respondents to publication records in the WoS publication output database. The results provide demographic information, such as sex and race or ethnicity of publication authors.

These data shed light on publication output differences between groups defined by race or ethnicity and sex, by discipline, and by impacts to R&D career paths (Chang, White, and Sugimoto forthcoming).[‡] The point estimates in Figure PBS-A show the odds of pre-doctorate student publishing by ethnic group or sex relative to White students (or men, for the sex comparison) while the error bars show the confidence around that point estimate (95% confidence interval). The confidence interval is closely linked to the size of the sample. In the SDR-WoS data, the number of minorities and women receiving degrees in the population influences the sample size—and, consequently, the ability to measure odds ratios. For example, there are 3,750 women who received mathematics or statistics PhDs compared to 10,450 men (Table SPBS-32). A similar issue arises for mathematics or statistics PhDs by race or ethnicity (Table SPBS-33). Overall, compared to White graduates, Asian, Black, or Hispanic graduates are less likely to publish before their doctorate in biological, agricultural, and other life sciences; engineering; health sciences; and social sciences.

Figure PBS-A

S&E pre-doctorate publishing odds ratio, by sex and selected race or ethnicity: 1995–2006



Note(s):
S&E doctorates include science, engineering, and health PhD candidates at U.S. research doctorate institutions. Computer sciences is not included in the figure because the odds ratio and confidence interval show no conclusive results for any demographic group or sex. Table shows the estimated odds ratios of publishing at least one article or conference proceeding during the five years before receiving a doctorate in the combined Web of Science and Survey of Doctorate Recipients database. For more detail, see Table SPBS-32 and Table SPBS-33.

Source(s):
National Center for Science and Engineering Statistics, Survey of Doctorate Recipients; Clarivate, Web of Science.

Science and Engineering Indicators

Compared to men, women are less likely to publish before graduation in the biological sciences, agriculture, engineering, health sciences, physics, and social sciences. Pre-doctorate publications appear to factor into obtaining a job in which research is the primary activity.[§] For those with at least one pre-PhD publication, 56% reported that their first job has research as its primary activity compared to 37% of those without a publication (Chang, White, and Sugimoto forthcoming).

* A machine learning approach matches the SDR respondents to the authors of publications indexed by the Web of Science (WoS). The matching algorithm incorporates name commonality, research field, education, employment affiliations, coauthorship network, and self-citations to predict matches from the SDR respondents to the WoS.

† WoS is a bibliometric database of conference proceedings and peer-reviewed literature with English-language titles and abstracts.

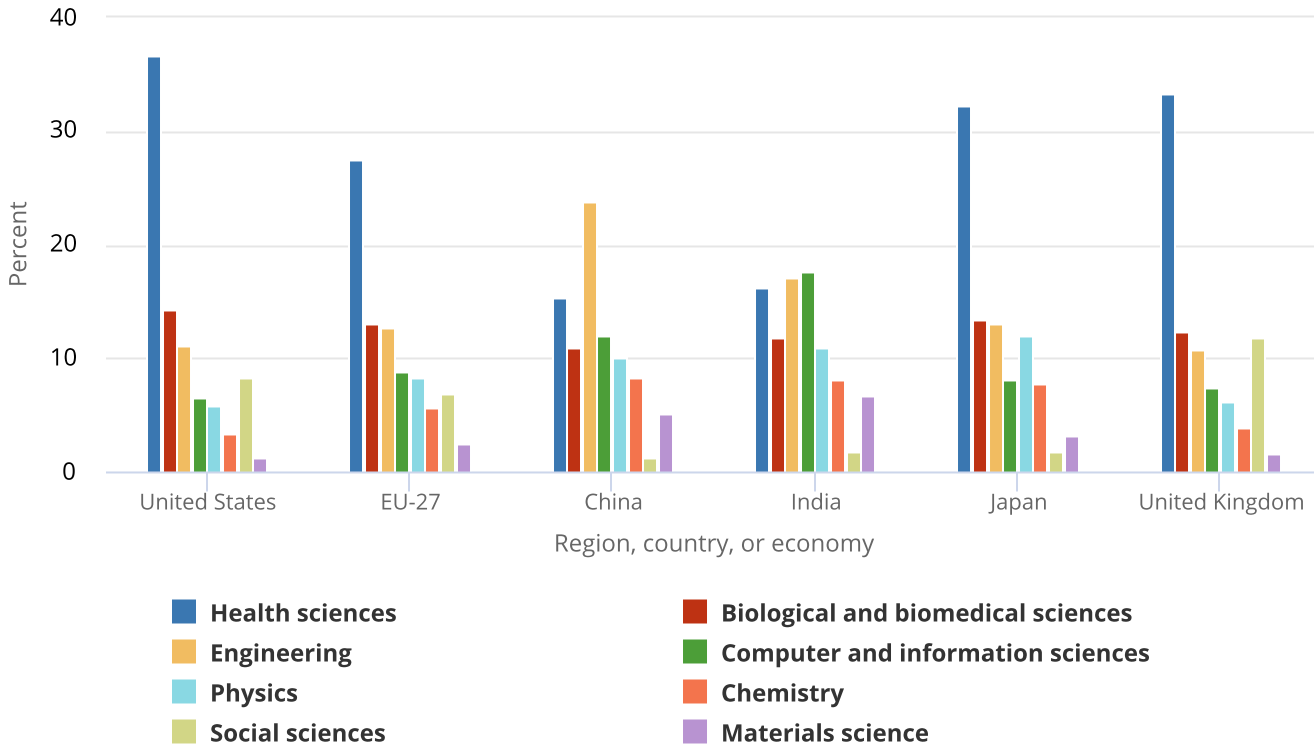
‡ To predict pre-doctorate publishing propensity, separate models were fitted for each doctoral field, and the following factors from the NCSES's Survey of Earned Doctorates were controlled: doctorate award year, type of PhD-awarding institution, source of primary support, community college experience, U.S. citizenship status at the time of degree award, level of parental education, marital status, dependents under 18 years old, disability, graduate debt, and name commonality.

§ The model controls for critical factors, such as the PhD institutions ranking as a high research institution, year of graduation, citizenship, parental degree, and student debt. The model does not measure article submissions or rejections.

Distribution of publications by field of science and region, country, or economy can indicate research priorities and capabilities. Health sciences is the largest field of science globally (25% of publications in 2020) (Table SPBS-2 and Table SPBS-10). Likely due to COVID-19, health sciences publications grew 16%, and biological and biomedical sciences publications grew 15% from 2019 to 2020, far surpassing their previous 2009–19 compound annual growth rates of 3% for each (Table SPBS-5 and Table SPBS-10). In the United States, the European Union (EU-27), the UK, and Japan, health sciences publication output far exceeds that of any other field (Figure PBS-3).[6] The United States, the UK, and the EU-27 have the highest proportions of articles in the social sciences of the six countries and regions shown. In China, the largest research area is engineering (24%), followed by health sciences (15%) and computer and information sciences (12%). The largest scientific field for publication output in India is computer sciences (18%). Japan has a portfolio with health sciences (32%) at the top, followed by biological and biomedical sciences (13%) and engineering (13%).

Figure PBS-3

S&E research portfolios, by eight largest fields of science and by selected region, country, or economy: 2020



EU = European Union.

Note(s):
Articles refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a fractional count basis (i.e., for articles from multiple countries, each country receives fractional credit on the basis of the proportion of its participating authors). See Table SPBS-1 for countries included in the EU; beginning in 2020, the United Kingdom was no longer a member of the EU. See Table SPBS-2 for all fields of science. See Table SPBS-2 through Table SPBS-16 for data on all regions, countries, and economies and all fields of science.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed May 2021.

Science and Engineering Indicators

There is increasing interest in measuring publication output that crosses or combines the standard scientific fields for solving boundary-defying issues, such as climate change or poverty reduction (NRC 2014, NASEM 2021). While publication output provides a potential avenue for measuring cross-disciplinary research output, there are challenges for national-level measures. (See sidebar **Measuring Cross-Disciplinarity Using Publication Output**.)

SIDEBAR

Measuring Cross-Disciplinarity Using Publication Output

This sidebar uses cross-disciplinarity as an envelope term that includes convergent, multidisciplinary, and interdisciplinary research because the measurement techniques for examining publication output are similar. Cross-disciplinary research includes the following:

- Convergent research that is driven by a specific and compelling problem requiring deep integration across disciplines (NSF 2019). Convergent science is a team-based approach to problem solving cutting across fields of inquiry and institutional frontiers to integrate areas of knowledge from multiple fields to address specific scientific and societal challenges.
- Multidisciplinary research (MDR) that "juxtaposes two or more disciplines focused on a question … [where] the existing structure of knowledge is not questioned" (NRC 2014:44).
- Interdisciplinary research (IDR) that "integrates information, data, methods, tools, concepts, and/or theories from two or more disciplines focused on a complex question, problem, topic, or theme" (NRC 2014:44).

Efforts using publication output to measure cross-disciplinary research yields results that are not suitable for comparing at the country level (Wagner et al. 2011; Wang and Schneider 2020). This finding is similar to sidebars in previous *Indicators* reports (NSB 2010; NSB 2016; Wagner et al. 2009). This sidebar explains the ongoing methodological issues with measuring convergence, MDR, and IDR at the country level and provides potential directions for future research.

For measurement at the country level, researchers have analyzed cross-disciplinary research using various bibliometric measures. Some have used article citations (Campbell et al. 2015; Porter and Chubin 1985), coauthor fields of specialization (Porter et al. 2007), text mining of abstracts or keywords listed on each article (Del Rio et al. 2001), or network analysis (Leydesdorff and Rafols 2011). An analysis of various approaches for measuring interdisciplinarity revealed a lack of consistent measurement outcomes across scientific fields, over time, and for countries or economies (Digital Science 2016).

Measuring cross-disciplinarity is challenging because indicators that are valid by one measure (e.g., citation counts), are not stable in another scientific area. For example, looking within the broad field of health sciences, health economics uses fewer citations, while biomedicine uses many more. When attempting to measure cross-disciplinarity for health sciences, the differences between health economics and biomedicine are, at least in part, related to different citation habits and not necessarily to differences in the cross-disciplinarity of the research.

Although research has not uncovered robust cross-disciplinary measures for countries, there are insights into the growth and influence of convergence, IDR, and MDR. Measured broadly, researchers find growth in cross-disciplinarity: "from about the mid-1980s, both natural sciences and engineering (NSE) and medical fields (MED) raised their level of interdisciplinarity at the expense of a focus on specialties" (Larivière and Gingras 2014:197). The team also found that the social sciences, as well as the arts and humanities, were the most open to collaborating with other disciplines. While cross-disciplinarity has grown, citation lags are associated with cross-disciplinary research papers. Specifically, they garner fewer than the normal number of citations for the first 3 years but pick up more citations than normal over 13 years (Wang, Thijs, and Glänzel 2015).

Recently, Digital Science prepared a report for the Research Councils of the United Kingdom (RCUK) that scanned the current literature and measurement approaches (Digital Science 2016). RCUK concluded that "no single indicator of interdisciplinarity (either MDR or IDR) analysed here should, used alone, satisfy any stakeholder. They show diverse inconsistency—in terms of change over time, difference between disciplines and trajectory for countries—that raises doubts as to their specific relevance" (Digital Science 2016:8). The RCUK report suggested that combining bibliometric IDR measures with other data, such as award information, could create a framework for expert analysis of IDR. Among the recommendations were continued exploration of text analysis and the inclusion of departmental affiliations in award information.

Similarly, the 2021 National Academies of Sciences, Engineering, and Medicine workshop on Measuring Convergence in Science and Engineering found that "using a single or even a few atomistic indicators to measure complex research activities capable of addressing societal problems is misguided" (NASEM 2021:49). Workshop participant Ismael Rafols suggested shifting from an atomistic to a portfolio approach, investigating the entire landscape that makes convergence possible.

International Collaboration and Citations

S&E research has steadily become more global over the past decade; this trend can be measured by examining cross-national coauthorships and citations (Luukkonen et al. 1993; Glänzel and Schubert 2005; Royal Society 2011). Researchers gather scientific expertise beyond their country's borders through collaboration, both direct (working to coauthor articles) and indirect (through referencing articles from other countries). Measured at the country level, international collaboration and citation are strongly influenced by the size of the country as well as the policies of the country. For example, some countries provide preferential funding for research within a region through programs such as the European Commission's Horizon 2020.[7]

Researchers collaborate for a number of reasons, including to develop a scientific relationship with another researcher, to gain access to costly or shared equipment, or to meet conditions attached to research funding that require international collaboration (Wagner 2018). The benefits of international collaboration include achieving outcomes that no one nation could achieve; training a robust S&E workforce, including researchers from developing countries; advancing domestic science excellence; increasing the impact of discoveries through better distribution of knowledge; strengthening scientific and diplomatic relations; and enhancing a sense of shared responsibility for future action (Lyons et al. 2016). Also, international collaborations increase the impact of research, as measured by citations (Narin, Stevens, and Whitlow 1991; Sugimoto et al. 2017; Chinchilla-Rodríguez, Sugimoto, and Larivière 2019). Domestic collaboration among researchers in the academic, government, and industry sectors also receives higher citation rates when compared to articles from a single author or authors from only one sector (see section Business Collaboration in Peer-Reviewed Publications in forthcoming *Indicators 2022* report "[2022] Invention, Knowledge Transfer, and Innovation").

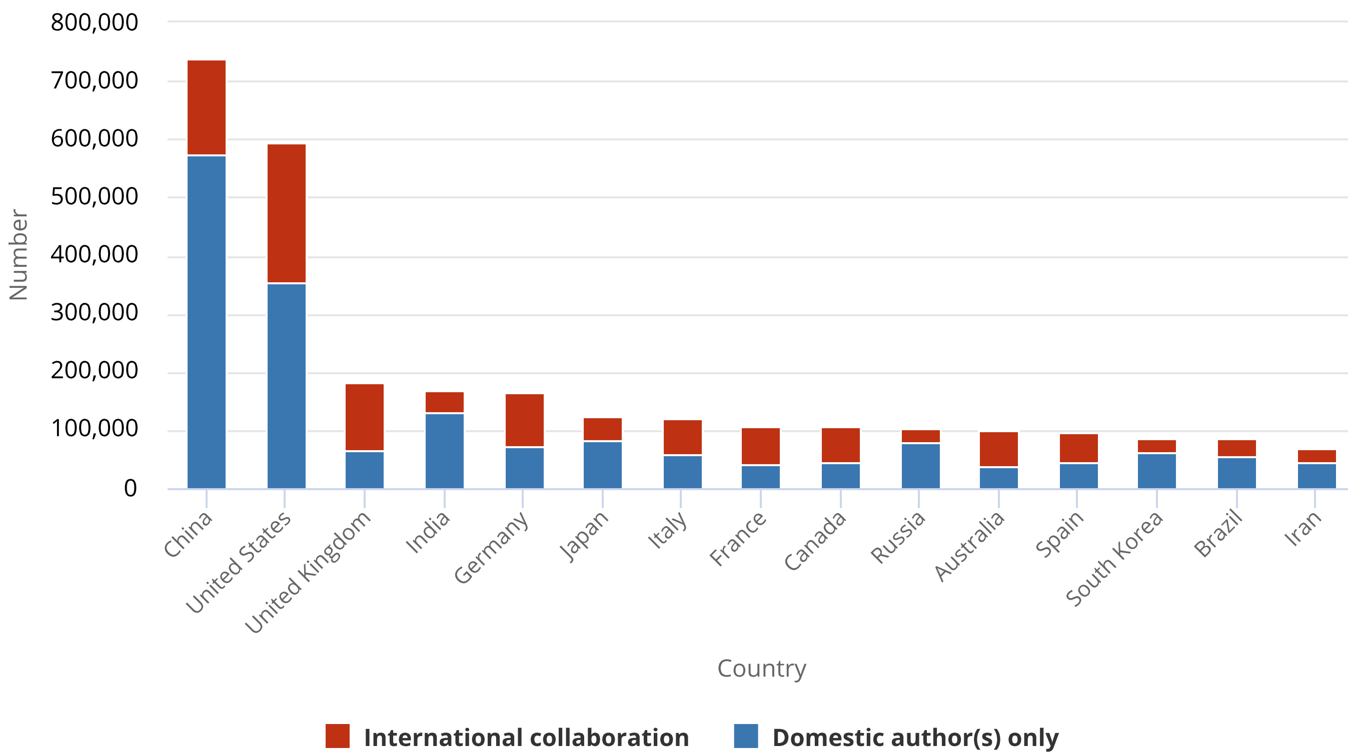
International Collaboration Patterns

Articles that have authors from universities or research institutions in at least two countries indicate international collaboration. The percentage of worldwide S&E articles produced with international collaboration increased from 18% to 23% from 2010 to 2020 (Table SPBS-35).[8] International connections among researchers indicate the expanding research capabilities around the world. Authors with their institutional address in the United States were the largest group participating in international collaboration, cowriting 239,344 articles, or roughly 35% of the world's internationally coauthored articles in 2020 (Table SPBS-35). The United States coauthors the most articles in collaboration with China, the UK, Germany, and Canada (Table SPBS-37).

The United States collaborated with international partners on 40% of the U.S. articles in 2020 (Table SPBS-35). International collaboration rates vary by country. There is a higher percentage of international collaboration for the UK (65%), Australia (62%), France (61%), and Canada (58%), with lower rates for India (22%), China (22%), and Russia (25%) (Figure PBS-4). Beyond the 15 largest producers, the 2020 international collaboration rates vary: Saudi Arabia (76%), Switzerland (73%), and Belgium (72%) have relatively high collaboration rates, while Brazil (35%) and Poland (38%) are lower (Table SPBS-35).[9]

Figure PBS-4

International coauthorship of S&E articles for the 15 largest producing countries of S&E articles, by country: 2020



Note(s):
Articles refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a whole count basis (i.e., each collaborating country or economy is credited with one count). Articles without international coauthorship are counts of articles with one or more institutional addresses all within a single region, country, or economy, which include single-author articles and articles coauthored under the same institutional address. International articles are articles with institutional addresses from more than one country or economy. The numbers of articles from the International collaboration and Domestic author(s) only categories may not sum to the total article number because some coauthored publications have incomplete address information in the Scopus database. These publications often cannot be reliably identified as international or domestic collaborations. For this reason, they are not included in either subcategory but are still counted toward the total number of articles. For more detail, see Table SPBS-35.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed May 2021.

Science and Engineering Indicators

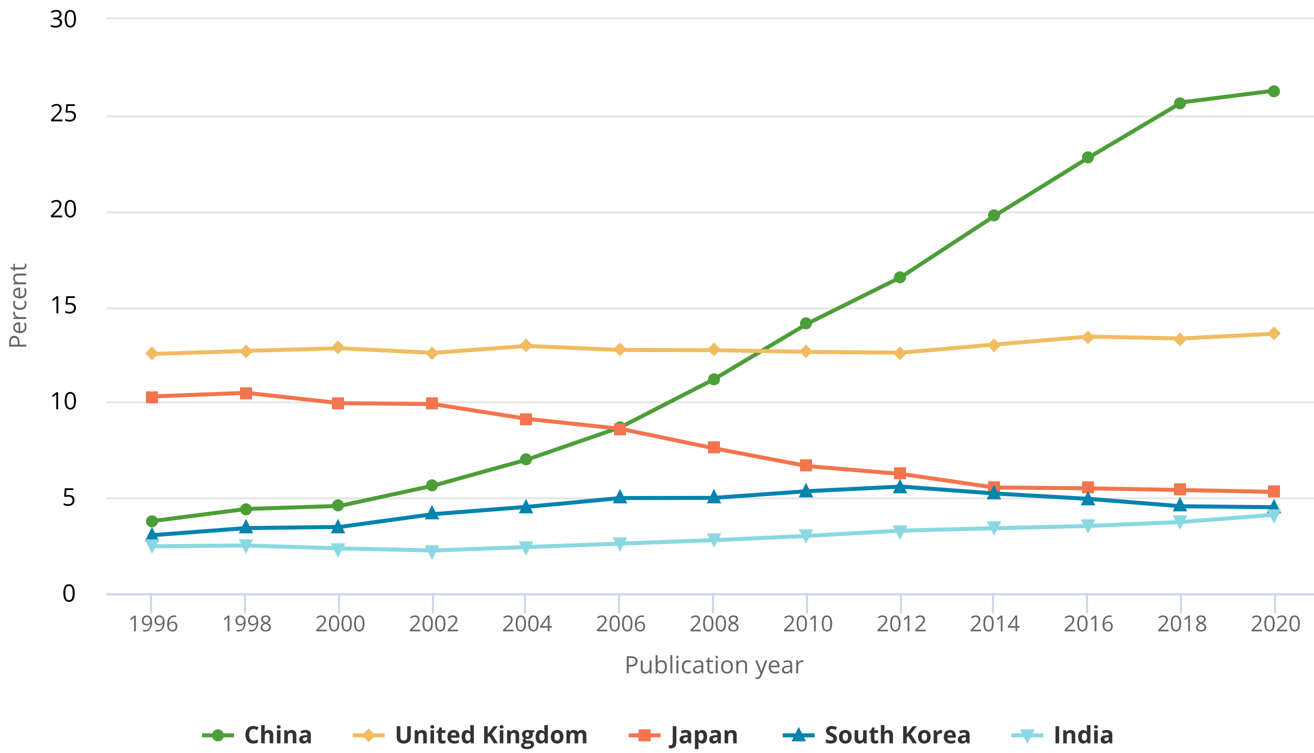
The international collaboration rate was stable for 2018, 2019, and 2020, possibly indicating that international collaborations were not impacted by COVID-19 (Table SPBS-35). However, there is a potential for a lagged effect on published articles because the appearance of research in peer-reviewed journals typically evolves over multiple years (Fry et al. 2020).

Globally, publication output shares have changed (Figure PBS-2), and so have the U.S. international collaboration partnerships, especially with Asian countries. In 1996, the largest collaboration country for the United States was the UK, which accounted for 13% of U.S. internationally collaborated articles; by comparison, China (26%) was the largest collaborator with the United States in 2020 (Figure PBS-5). In 2020, roughly one of every four internationally coauthored U.S. articles was with an author with an institutional address in China. The U.S.-China collaboration on publications has benefited both countries as well as global science through the amount of published research collaborations, funding agency support from U.S. and China, and the role of collaborators on papers (Lee and Haupt 2020). The rapid growth in

U.S.-China collaborations coincided with the growing scientific and technological capabilities of China, such as rising R&D spending (see forthcoming *Indicators 2022* report "[2022] Research and Development: U.S. and International Trends") and university degree awards (see forthcoming *Indicators 2022* report "[2022] Higher Education in Science and Engineering"), which may be contributing factors to the U.S.-China collaboration pattern.

Figure PBS-5

U.S. international articles with coauthor(s) from the United Kingdom and Asian countries: Selected years, 1996–2020



Note(s):
Articles refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a whole count basis (i.e., each collaborating country or economy is credited with one count). Articles with international institutions are counts of articles with institutional addresses from more than one country or economy. For more detail, see Table SPBS-37.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed May 2021.

Science and Engineering Indicators

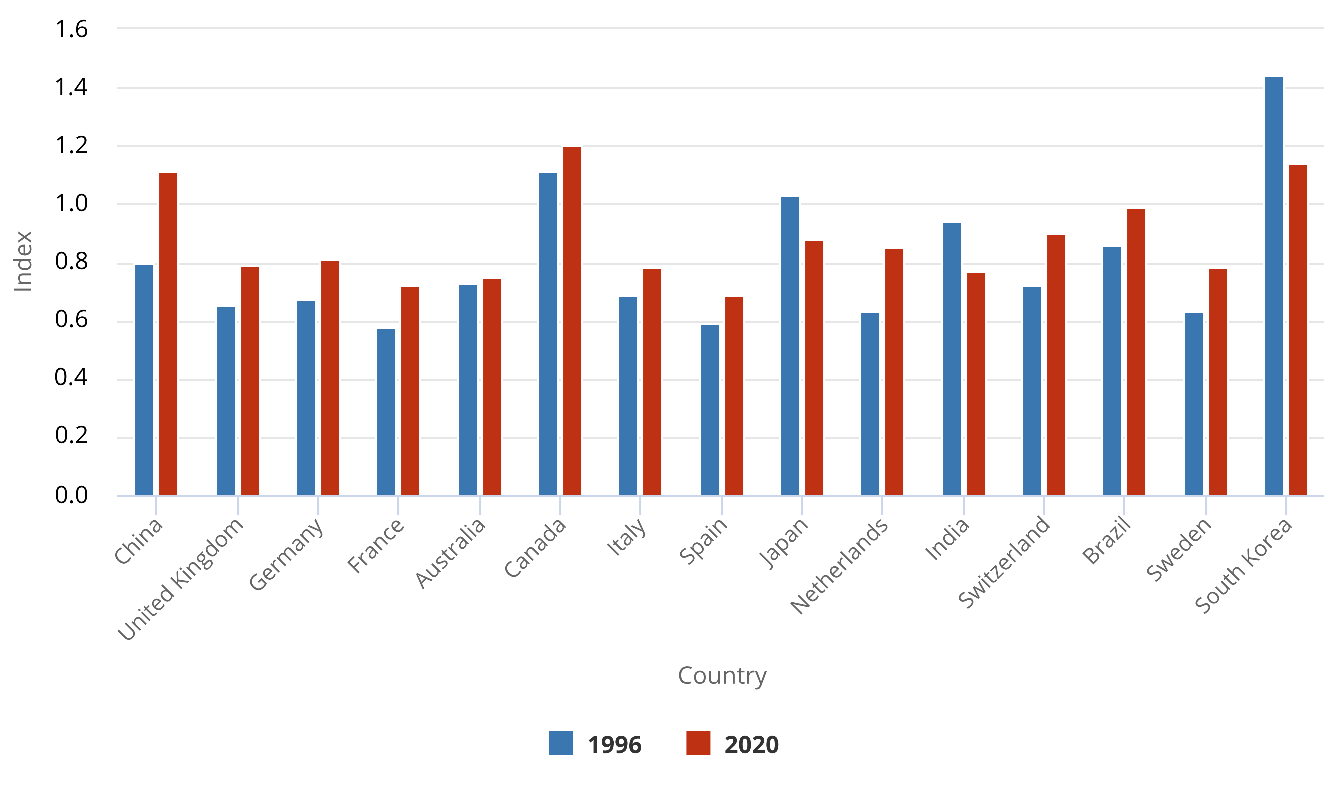
U.S. international articles with other Asian countries have also changed over time. From 1996 to 2020, U.S. articles with Japan generally declined but gradually increased with India. South Korea presents a case in which U.S. partnerships grew from 1996 to 2012 and then declined from 2014 to 2020 as South Korea increased its partnerships with other Asian countries. Meanwhile, there has been little change to the percentage of U.S. coauthorships with Canada and European countries, such as the UK, Germany, France, Switzerland, Italy, the Netherlands, and Sweden (Table SPBS-37).

Since 1996, the countries that produce over 75% of the world's publication output partner with at least one U.S. author for a quarter or more of their international articles, with the exceptions of Russia and Spain from 2018 to 2020, France in 2008, and Iran for most years (Table SPBS-38). From 1996 to 2020, Canada, China, and South Korea have partnered with the United States the most, but coauthor percentages have declined in recent years as those countries have increased their coauthorship with other countries, primarily in the Pacific region.

Normalizing the international collaboration by a country's publication output enables comparisons independent of country size (He 2009). The international collaboration index is obtained by dividing a country's share of collaboration with another country by its overall share of international collaborations with all countries. An index of 1.0 occurs when coauthorship between two countries is exactly proportional to their overall shares of international collaborations. Index values above 1.0 indicate stronger ties, and scores below 1.0 indicate weaker collaborative ties. The index of internationally coauthored S&E articles increased between the United States and most other major research countries from 1996 to 2020, with the exceptions of South Korea (1.4 to 1.1), Japan (1.0 to 0.9), and India (0.9 to 0.8) (Figure PBS-6).

Figure PBS-6

Relative international collaboration specialization of select large producing countries with the United States: 1996 and 2020



Note(s):
Article counts for computing the index refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a whole count basis (i.e., each collaborating region, country, or economy is credited with one count). Regions, countries, or economies that have contributed to less than 1% of all internationally coauthored articles in 2018 are omitted. The index of collaboration is calculated as follows: $ICxy = (Cxy / Cx) / (Cy / Cw)$, where *ICxy* is the index of collaboration between country *x* and country *y*, *Cxy* is the number of papers coauthored between country *x* and country *y*, *Cx* is the total number of international coauthorships by country *x*, *Cy* is the total number of international coauthorships by country *y*, and *Cw* is the total number of international coauthorships in the database. For additional countries, see Table SPBS-39.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed May 2021.

International collaboration on COVID-19 research shows the role of the United States as a leading research producer and collaborator. (See sidebar **Coronavirus Publication Output and International Collaboration**.)

SIDEBAR

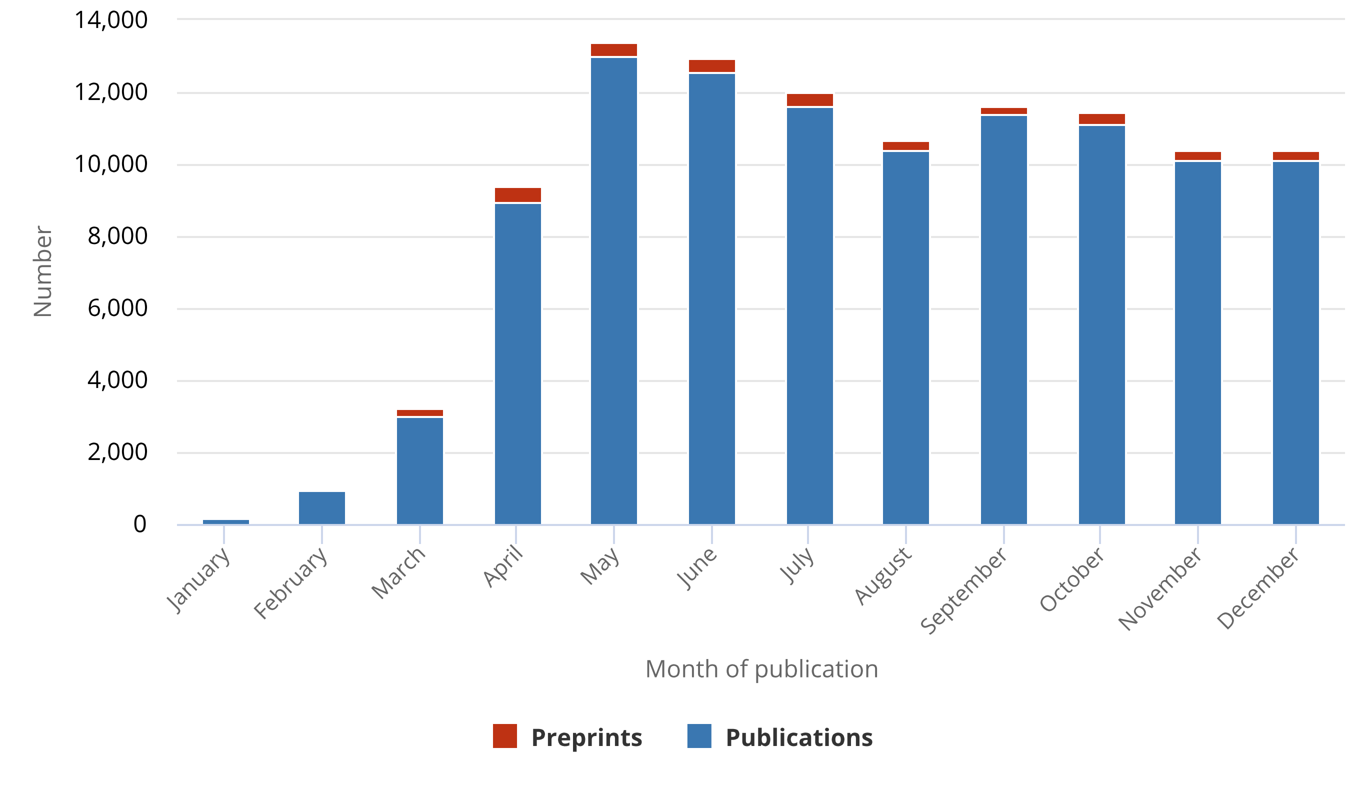
Coronavirus Publication Output and International Collaboration

The onset of the COVID-19 pandemic caused by the novel coronavirus, SARS-CoV-2, marked a dramatic growth in research articles as well as actions to expand sharing of that research. This sidebar reviews publication output data from 1 January to 31 December 2020, finding dramatic month-to-month increases in coronavirus-related publication output. The sidebar also highlights international research collaboration.

Early in the pandemic, the White House Office of Science and Technology Policy (OSTP) issued "a call to action to the Nation's artificial intelligence experts to develop new text and data mining techniques that can help the science community answer high-priority scientific questions related to COVID-19" (Karami et al. 2021:1).* OSTP partnered with the Allen Institute for Artificial Intelligence (AI), the Chan Zuckerberg Initiative, Georgetown University's Center for Security and Emerging Technology, Microsoft, and the National Library of Medicine at the National Institutes of Health. This consortium launched the COVID-19 Open Research Dataset (CORD-19) with articles and preprints covering the novel coronavirus as well as research on influenzas and coronaviruses, such as SARS and MERS (Wang et al. 2020).[†] Adoption of CORD-19 was swift; in its first month of release, CORD-19 received over 1.5 million web views and over 75,000 downloads (Wang et al. 2020). The number of peer-reviewed articles and preprints entering the data set peaked in May, totaling more than 107,000 articles for 2020 (Figure PBS-B). The rapid growth and sustained publication output indicate a previously unseen international research focus (Aviv-Reuven and Rosenfeld 2020).[‡] Besides removing duplicates, the CORD-19 database was not filtered to remove letters, notes, editorials, and so forth to show activity in the field.

Figure PBS-B

S&E articles on coronavirus in CORD-19: 2020



Note(s):
Records are as contained in the COVID-19 Open Research Dataset (CORD-19). Counts are filtered to those that contain URL addresses and are deduplicated based on the database-provided unique ID. Figure is currently being updated to use the keywords to build the data set: "covid"; "coronavirus*"; "corona virus*"; "severe acute respiratory syndrome"; "middle east respiratory syndrome"; "SARS-CoV-2"; "spike protein*"; "mers-cov"; "nl63"; "229e"; "oc43"; "hcov"; "alphacoronavirus*"; "covid19"; "covid-19"; "SARSCoV2." Note that the * character is used to specify that the words should serve as prefixes, such that "coronavirus*" also captures articles that used "coronaviruses," for example. Preprints come from preprint servers (e.g., arXiv, bioRxiv, and medRxiv).

Source(s):
COVID-19 Open Research Dataset, accessed 12 May 2021.

Science and Engineering Indicators

The United States and China produced the largest portion of the coronavirus papers in 2020, as shown in the Scopus database (Table PBS-A).[§] The count of coronavirus articles from Scopus, 59,000, was filtered to remove preprints, letters, notes, editorials, etc.[‖] This filtering matches that on the Scopus database used in other parts of this report. Therefore, the 59,000 articles represent peer-reviewed coronavirus publications in 2020. Italy and Spain, sites of high early outbreaks of coronavirus, ranked higher, at 3rd and 6th by volume of coronavirus articles, compared to their rankings of 8th for Italy and 13th for Spain in the overall publication count (Table PBS-1). The publication output of Italy and Spain matches a broader trend of more coronavirus articles from countries with higher rates of COVID-19 infection in 2020 (Cai, Fry, and Wagner 2021).

Table PBS-A

Coronavirus articles for 15 largest coronavirus article producing regions, countries, or economies: 2020

(Number of articles)

Region, country, or economy	Coronavirus article count
World	58,627.0
United States	12,199.6
China	7,529.1
Italy	4,070.7
India	3,528.0
United Kingdom	3,518.6
Spain	1,930.9
Germany	1,673.4
Brazil	1,634.6
France	1,542.6
Canada	1,426.4
Iran	1,227.9
Australia	1,139.8
Russia	975.2
Turkey	968.4
South Korea	708.4

Note(s):
Number of publications uses fractional counting. The list is sorted based on decreasing fractional counting numbers. The data were filtered in a similar manner as the Scopus data used throughout this report. The filters removed preprints, letters, notes, editorials, errata, surveys, low-quality journals, or lack of institutional addresses. Any article that had any of the following keywords in its title, abstract, or author-defined keywords were used to build the data set: "covid"; "coronavirus*"; "corona virus*"; "severe acute respiratory syndrome"; "middle east respiratory syndrome"; "SARS-CoV-2"; "spike protein*"; "mers-cov"; "nl63"; "229e"; "oc43"; "hcov"; "alphacoronavirus*"; "covid19"; "covid-19"; "SARSCoV2." The * character is used to specify that the words should serve as prefixes, such that "coronavirus*" also captures articles that used "coronaviruses," for example. Articles, conference proceedings, and reviews indexed in the Scopus database were searched. For additional detail, see Table SPBS-55.

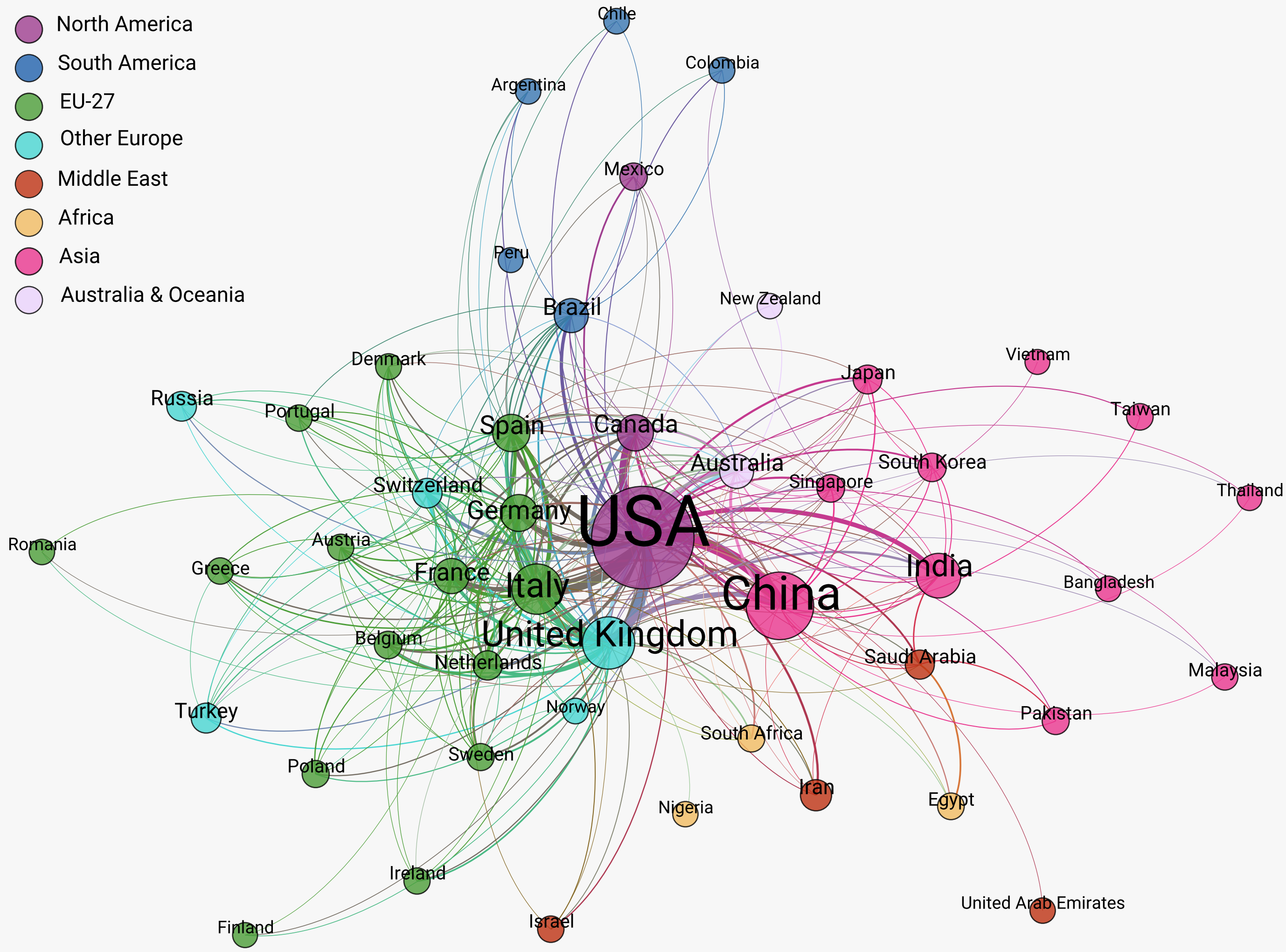
Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed May 2021.

Science and Engineering Indicators

International collaboration on coronavirus-related research is 22% overall,[#] which is virtually the same as the 23% calculated for all S&E articles (Table SPBS-35 and Table SPBS-55). United States–China and United States–United Kingdom (UK) were strong international coauthorship partners on COVID-19 publications (Figure PBS-C). In the network figure, the color and symbol indicate the region or the country; the size is proportional to the total number of coronavirus-related articles written by each country; the thickness of the links between nodes is proportional to the quantity of cowritten papers; and the distance between nodes indicates the relatedness (similarity in terms of network properties) of the countries. The network analysis shows the centrality of the major research countries—United States, China, the UK, European Union (EU-27) countries, Japan—while more evolving R&D countries, such as Iran and Russia, are less integrated into the network. The diagram also shows strong collaboration between the United States and authors in China, the UK, and Canada. More information on the network measures and definition are available in this report's **Technical Appendix** and Table SPBS-57.

Figure PBS-C

Coronavirus collaboration network, by country: 2020



Note(s):
Network diagram shows the number of cowritten articles by all pairs of countries within the top 50 producers of coronavirus-related research based on whole counting for those pairs that cowrote 50 articles or more. Coronavirus article counts refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a country on the basis of the institutional address(es) of the author(s) listed in the article. Links are only shown in a single direction, dictated by alphabetical order. The size of the nodes is proportional to the total number of coronavirus-related articles written by each country. The width of the links between nodes is proportional to the quantity of articles both countries have cowritten. Positioning of nodes is defined using the ForceAtlas2 algorithm.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed May 2021.

Science and Engineering Indicators

The research pattern seen in publication output during the COVID-19 pandemic is different than the pre-pandemic pattern. The “proportion of COVID-19 papers with a woman first author was 19% lower than that for papers published in the same journals in 2019” (Andersen et al. 2020). This decline in female research participation is also shown in an April 2020 survey of 4,535 U.S.- and European-based scientists. They found a decline of women’s work hours and a disproportionately larger share of home and family care relative to male counterparts during the pandemic (Myers et al. 2020). In addition, junior faculty and PhD students who responded to the American Finance Association Survey in October 2020 reported negative effects on time spent on research, self-reported research productivity, health, isolation, and other similar measures when compared to senior faculty (Barber et al. 2021; Woolston 2020).

[*] COVID-19 Open Research Dataset Challenge (CORD-19), **https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge**.

[†] OSTP worked with publishers to provide full-text coverage of relevant papers available in its catalog. Papers were made available under special COVID-19 open-access licenses.

[‡] CORD-19 includes COVID-19 research from the bioRxiv and medRxiv servers. Preprints have not received peer review but can contain the earliest sources of novel findings (Wang et al. 2020). Preprints are included to show activity, not peer-reviewed science.

[§] Country-level COVID-19 analysis uses the Scopus database because CORD-19 does not contain uniform data on authors' institutions. Retrieval of articles from both Scopus and CORD-19 used the keywords "covid"; "coronavirus*"; "corona virus*"; "severe acute respiratory syndrome"; "middle east respiratory syndrome"; "SARS-CoV-2"; "spike protein*"; "mers-cov"; "nl63"; "229e"; "oc43"; "hcov"; "alphacoronavirus*"; "covid19"; "covid-19"; "SARSCoV2." The * character is used to specify that the words should serve as prefixes, such that "coronavirus*" also captures articles that used "coronaviruses" (Science-Metrix 2021b; Wang et al. 2020).

[‖] The initial query of the Scopus database found 108,000 publications, close to the 107,000 found in CORD-19. The Scopus database applied filters to remove preprints (15,500), letters (12,800), notes (6,300), editorials (5,200), errata or surveys (2,000), low-quality publications (2,300), and other publications, such as those not being attributable to a country or of a nonstandard publication type (5,000).

[#] This is computed as the [papers involving international institutions] / [total number of papers] using whole counting. Some coauthored publications have incomplete address information in the Scopus database, so some papers cannot be reliably categorized as international or domestic collaborations. They are not included in either subcategory but are still counted toward the total number of articles.

International Citation Patterns

Like the indicators of international coauthorship, cross-national citations provide evidence that S&E research is increasingly international in scope. Citations to a country's publications that come from publications authored outside that country are referred to as international citations. This section provides a relative measure normalized by the publication output by country. If the international citations were simply summed by country, the United States, the UK, Germany, and China would account for the largest shares of the international citations simply because they produce so many articles.

The relative citation index (RCI) provides a method to group the citations between countries as above (greater than 1), near (about 1), or below (less than 1) the expected rate of citation relative to each country's output level.[10] Using the RCI, U.S. authors display a proclivity for citing English-speaking countries with a citation rate above the expected rate with those countries: for example, Canada (1.22), the UK (1.25), and Australia (1.06) (Table PBS-2). In the next grouping of those near the expected rate of citation are countries with historically large research communities: for example, Germany (0.91), Italy (0.71), and France (0.85). Finally, U.S. authors cite countries with emerging research communities less than expected given the number of articles produced by those countries: for example, China (0.39), India (0.21), and Iran (0.26).

Table PBS-2

Relative citation index for 15 largest producing regions, countries, or economies: 2018

(Index)

Citing region, country, or economy	Cited country or economy														
	China	United States	United Kingdom	India	Germany	Japan	Italy	France	Canada	Russia	Australia	Spain	South Korea	Brazil	Iran
China	2.60	0.74	0.64	0.50	0.55	0.52	0.54	0.54	0.67	0.13	0.89	0.57	0.99	0.30	0.77
United States	0.39	3.16	1.25	0.21	0.91	0.52	0.71	0.85	1.22	0.11	1.06	0.62	0.54	0.33	0.26
United Kingdom	0.40	1.29	6.68	0.27	1.18	0.50	1.01	1.09	1.21	0.14	1.64	0.89	0.47	0.39	0.33
India	0.87	0.60	0.70	6.42	0.53	0.42	0.65	0.54	0.62	0.18	0.77	0.66	0.95	0.55	1.62
Germany	0.37	1.22	1.52	0.22	6.81	0.62	1.05	1.23	0.99	0.23	1.00	0.91	0.50	0.34	0.25
Japan	0.55	1.09	1.00	0.30	1.00	8.50	0.78	0.95	0.76	0.16	0.80	0.68	0.90	0.30	0.28
Italy	0.39	1.05	1.32	0.32	1.07	0.54	9.85	1.25	0.92	0.16	0.97	1.40	0.60	0.54	0.52
France	0.41	1.23	1.50	0.29	1.34	0.65	1.30	8.38	1.10	0.23	1.09	1.13	0.54	0.49	0.34
Canada	0.50	1.51	1.49	0.28	0.92	0.47	0.78	0.96	9.30	0.11	1.45	0.75	0.57	0.44	0.55
Russia	0.50	0.70	0.82	0.46	0.95	0.54	0.85	0.84	0.65	12.96	0.69	0.72	0.55	0.42	0.60
Australia	0.61	1.17	1.75	0.30	0.84	0.45	0.73	0.80	1.32	0.11	11.35	0.76	0.59	0.42	0.55
Spain	0.47	1.01	1.40	0.38	1.06	0.50	1.60	1.19	0.99	0.16	1.21	9.84	0.63	0.74	0.61
South Korea	1.00	0.99	0.83	0.58	0.71	0.78	0.75	0.65	0.77	0.13	0.88	0.66	9.53	0.36	0.72
Brazil	0.50	0.81	1.00	0.69	0.72	0.40	1.07	0.88	0.89	0.16	1.08	1.30	0.61	11.55	0.99
Iran	0.91	0.52	0.59	1.09	0.43	0.33	0.83	0.49	0.74	0.16	0.90	0.73	0.81	0.66	16.30

Note(s):
Citations refer to publications from a selection of journals, books, and conference proceedings in S&E from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) listed in the article. Articles are credited on a fractional count basis (i.e., for articles with collaborating institutions from multiple countries or economies, each country or economy receives fractional credit on the basis of the proportion of its participating institutions). Citation counts are based on all citations made to articles in their publication year and in the following two years (i.e., 3-year citation window; scores in 2018 are based on citations to articles published in 2018 that were made in articles published in 2018–20). The relative citation index (RCI) normalizes cross-national citation data for variations in relative size of publication output. RCI is computed by dividing the share of the citing country's outgoing citations going to the cited country and then dividing that amount by the share of publications attributed to the cited country. An RCI of 1.00 means that the citing country cites publications from the cited country as much as would be expected to happen randomly, showing no particular affinity between the countries. Scores higher than 1.00 mean that the citing country has a higher-than-expected tendency to cite the cited country's S&E literature. For more detail, see Table SPBS-40. Cells in which the region, country, or economy collaborates at or above the world average for that year are shaded green.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed May 2021.

Science and Engineering Indicators

The RCI also provides a view into the international nature of the scientific research in a country by presenting the number of other countries from which the country draws its research (i.e., citations). For example, authors from China predominantly cite articles from China (2.6), and only South Korea, Saudi Arabia, and Australia are cited by China near the expected rate (Table PBS-2 and Table SPBS-40). Russia (12.9) and Iran (16.3) tend to cite within their own publications (Table PBS-2). On the other hand, France cites many other countries at a greater than expected rate—the United States (1.2), the UK (1.5), Germany (1.3), Italy (1.3), Canada (1.1), Spain (1.1), and Australia (1.1) (Table PBS-2). Many other European countries display a similar pattern. Supplementary tables provide equivalent information for each of the 14 scientific fields (Tables SPBS-41 through Table SPBS-54).

Impact of Published Research

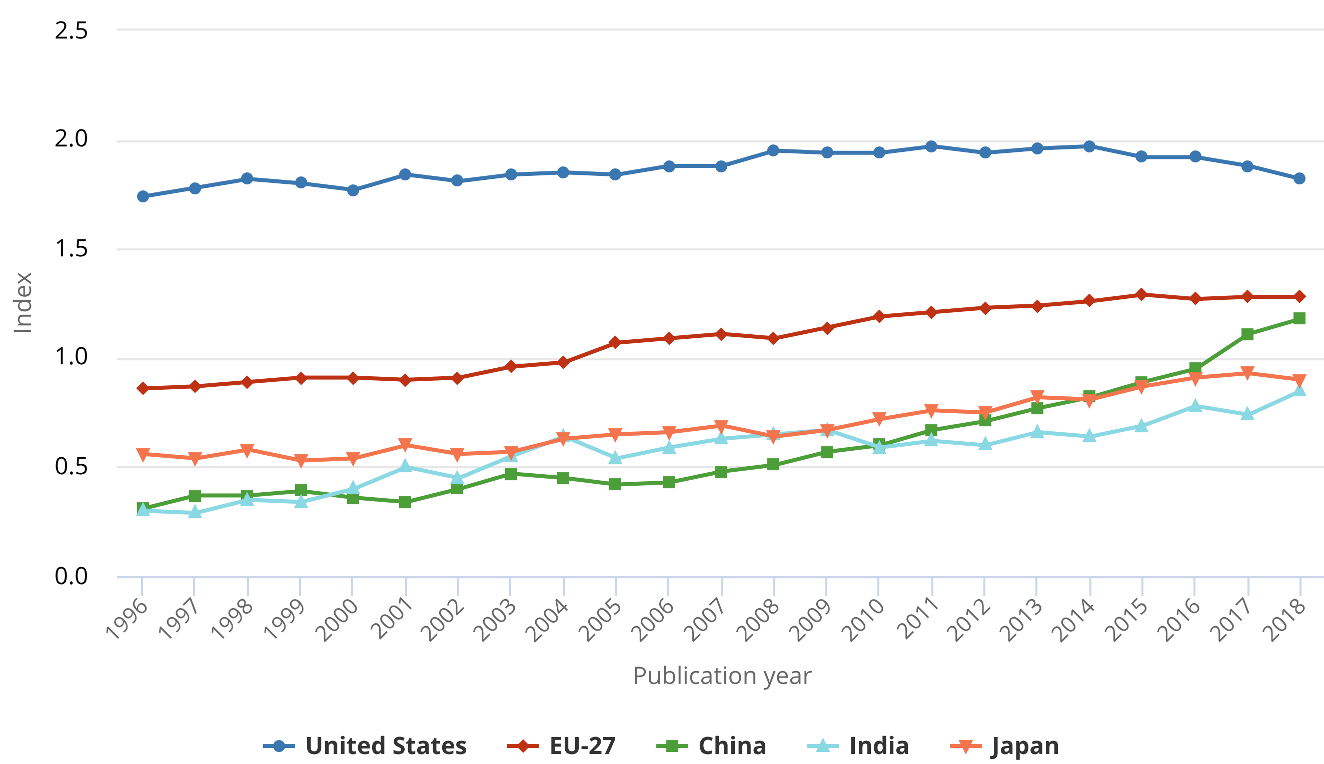
Data from citations provide a window into the impact of specific articles and authors, as well as the return on a country's scientific investment (Narin and Hamilton 1996). Research embodied in publications also impacts invention through patent citations, which will be covered in the forthcoming *Indicators 2022* report "[2022] Invention, Knowledge Transfer, and Innovation."

Publications with more citations are considered to have more impact (Garfield 1955; Waltman, van Eck, and Wouters 2013).[11] A small subset within the count of cited articles is categorized as highly cited articles (HCA), or the publications most frequently cited in other articles. This report presents data on the top 1% of HCA. The index of a country's number of HCA relative to the world's HCA (i.e., the HCA score) creates a comparable cross-country measure.[12] Conference proceedings are excluded from the HCA count because proceedings are not cited as often as journal articles.[13] In addition, conference papers are often later converted into articles in journals, and then it is the journal article that receives the majority of the citations. A minimum of 2 years from the publication date is required for calculation of HCA. This delay provides an opportunity for articles to accumulate citations after the publication (Wang 2012). This report provides data for articles published through 2018 because citations are calculated based on data up to 2020.

The United States contributed more than the expected number of articles to the world's pool of HCA in 2018 (HCA score of 1.8) (Figure PBS-7).[14] China's HCA score is above the expected share of HCA given their level of 2018 articles (HCA score of 1.2). The HCA scores for China track its emergence as a major S&E research country; the HCA for China in 1996 was 0.3. Not all major article-producing countries received above the expected HCA score; Japan and India each had HCA scores less than 1.0 in 2018 (both with a score of 0.9).

Figure PBS-7

Share of S&E articles in the top 1% most-cited journal articles, by selected region, country, or economy: 1996–2018



EU = European Union.

Note(s):
Articles refer to articles from peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Citation data are based on all citations made to articles in their publication year and all following years and are normalized by subfield and publication year to allow for comparisons across subfields and over time, resulting in the world level standing at 1.00 for each subfield and year. A minimum 2-year citation window is needed for a highly cited articles (HCA) score to be computed. This results in scores regarding HCA not being computed after 2018 because the citation window for more recent years is not yet complete. The share of articles in the top 1% is computed as follows: $Sx = HCAx / Ax$, where Sx is the share of output from country x in the top 1% most-cited articles; $HCAx$ is the number of articles from country x that are among the top 1% of most-cited articles (using full counting, with the exception of articles at the limit of the top 1%, which are fractioned so the world average can stand at 1%); and Ax is the total number of articles from country x with a relative citation score, which excludes articles released after 2018 and unclassified publications. The world average is 1.00 for each year. For more details, see Table SPBS-74. See Table SPBS-1 for countries included in the EU; beginning in 2020, the United Kingdom was no longer a member of the EU.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed May 2021.

Science and Engineering Indicators

The trends in HCA scores provide an opportunity to understand the changes in a country's research impact. From 2008 to 2018, the impact of Chinese researchers grew faster than the impact of researchers in the United States and other countries (Conte et al. 2017; Xie and Freeman 2018). The HCA score of the United States was 2.0 for 2008 and 1.8 for 2018. Over the same decade, China's HCA score more than doubled (from 0.5 in 2008 to 1.2 in 2018). The HCA scores for articles with authors in India and Japan also increased from 2008 to 2018; India's HCA score increased from 0.7 to 0.9, and Japan increased from 0.6 to 0.9.

Changes over time in the HCA by field of science provide a more nuanced story of research across countries. The normalization of the HCA by output and scientific field using fractional counting spotlights countries with relatively low publication output but consistently high rankings among the most cited articles in a field:[15] for example, agricultural sciences in Ireland (2.51), engineering in Singapore (2.87), and materials science in the Netherlands (2.43) (Table SPBS-59, Table SPBS-64, Table SPBS-67).

Conclusion

Based on the overall size of the U.S. contribution to S&E research publication output and the relative impact, as measured by citations to its S&E publications, the United States remains a highly influential nation. The publication output of the United States, Japan, the UK, and the countries of the EU-27 is concentrated in health sciences, while China and India focus more on engineering and computer and information sciences, respectively. In terms of S&E publication quantity, China's output has grown rapidly and has now surpassed that of the United States. In terms of impact among S&E publications, China has increased rapidly in the last decade. The high-income economies (including the United States, the EU-27, and Japan) have slowly increased their large base of S&E publications, while middle-income and upper-middle-income economies have increased rapidly across production, collaboration, and impact beginning from a smaller S&E publications base.

International research collaboration is increasing, reflecting traditional cross-country ties and new relationships that stem from growing capabilities in the middle-income economies. Greater publication output—with greater and more diverse collaborations—means more countries are contributing, and many are doing so with U.S. authors.

Glossary

Definitions

Citations: Citations, generally at the end of each article, provide researchers with the list of the prior research relied on for the article. Citations of S&E publications by other S&E publications provide an indication of the impact of publications and of the flow of knowledge or linkage between sectors or geographic locations.

Coauthorship: Coauthorship refers to cases in which there is more than one author listed on a publication. Data on coauthorship can be used to measure collaboration across regions, countries, economies, and institutional sectors. Publication counts of coauthorship use whole counting, resulting in a full count assignment for each country contributing to the article. An article is considered to contain an international coauthorship when there are institutional addresses for authors from two or more different countries. Table SPBS-36 shows international coauthorship from 1996 to 2020.

European Union (EU-27): The European Union (EU) comprises 27 member nations: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, and Sweden. Bulgaria and Romania joined the EU in 2007, and Croatia joined in 2013; they are included in the EU grouping for all years analyzed in this report. In 2020, the United Kingdom (UK) left the EU, and the UK is excluded from the EU in the text of the report but is available in the supplemental tables as EU-28.

Fractional counting: Method of counting S&E publications in which credit for coauthored publications is divided among the collaborating institutions or countries based on the proportion of their participating authors. Fractional counting allocates the publication count by the proportion of each of the countries or institutional coauthors named on the article. Fractional counting enables the counts to sum up to the number of total articles. For example, if a paper were authored by two researchers from the University of Oslo, one from the University College London, and one from the University of Washington, half of the paper would be attributed to Norway, and one quarter each to the UK and the United States when the fractions are calculated at the level of researchers. For this report, fractions were calculated at the level of researchers. If an author provides multiple institutions, and those institutions are in different countries, then each country receives an appropriate fraction of the count.

Highly cited articles (HCA): An HCA score is not a perfect measure of influence, but it provides an indication of scientific impact (Waltman, van Eck, and Wouters 2013). The first step is to create a data set of the top 1% most-cited publications in each field and for each year. The HCA score for a country is the share of authors with institutional addresses within that country that have articles among the top 1% of the world's highly cited articles, relative to all the articles ascribed to that country up to 2018. The HCA score is indexed to 1, so that a country whose authors produce highly cited articles at the expected (i.e., global average) rate has an HCA score of 1.0; that is, 1% of the country's articles are among the top 1% of the world's highly cited articles. Countries with authors producing highly cited articles at greater than the expected rate have HCA scores greater than 1, and countries with authors producing influential articles at lower than the expected rate have HCA scores less than 1. For example, assume a world of two countries produced an output of 10,000 articles, with country *x* authors producing 7,000 articles and country *y* authors producing 3,000 articles. If both countries had the same influence in the citation records, then country *x* would have 70 highly cited articles, and country *y* would have 30 highly cited articles in the top 100 most-cited articles in the world. Each country would have an HCA score of 1. The scores would be different if authors in one of the countries produced a higher proportion of the highly cited articles. For example, if authors in country *y* produced 50 of the most highly cited articles, then their HCA score would be 1.7, indicating that 1.7% of the articles of country *y*'s authors (50 out of 3,000) are among the top 1% of the world's highly cited articles. For country *x*, the HCA score would be 0.7 (50 out of 7,000).

International collaboration index: The international collaboration index is calculated as follows: $ICxy = (Cxy / Cx) / (Cy / Cw)$, where $ICxy$ is the index of collaboration between country x and country y, Cxy is the number of papers coauthored between country x and country y, Cx is the total number of international coauthorships by country x, Cy is the total number of international coauthorships by country y, and Cw is the total number of international coauthorships in the database. An index greater than 1.0 means that a country-country pair has a stronger-than-expected tendency to collaborate; an index less than 1.0 indicates a weaker-than-expected tendency to collaborate (Table SPBS-39).

Relative citation: The relative citation is a normalization of the relative scientific impact of papers produced by a given country that takes into consideration the fact that citation behavior varies between fields and years of publication. For a paper in a given scientific field and publication year, the citation count is then divided by the average count of all papers in the relevant field and publication year.

Relative citation index: The relative citation index normalizes cross-national citation data for variations in relative size of publication output. It is computed by dividing the share of the citing country's outgoing citations going to the cited country and then dividing this by the share of publications attributed to the cited country.

Whole counting: This measure (also called *full* or *integer* counting) assigns one count to each country or institutional sector involved in coauthoring the article, irrespective of its proportionate involvement in authorship. Whereas fractional counting aims to assess the proportionate contributions of countries or sectors, whole counting aims instead to assess the participation of countries or sectors. One result of this difference is that the sum of articles from countries or institutional sectors will exceed the total number of articles when whole counting is used. For the United States in 2018, there were 435,000 publications in the Scopus database as measured on a fractional count basis and 563,000 as measured on a whole count basis (Table SPBS-2 and Table SPBS-17). In the full-counting method, each paper is counted once for each entity listed in the address field. For example, if a paper was authored by two researchers from the University of Oslo, one from the University College London, and one from the University of Washington, the paper would be counted once for Norway, once for the UK, and once for the United States. When it comes to aggregating groups of institutions (e.g., research consortia) or groups of countries (e.g., the EU), double counting is avoided. This means that if authors from Croatia and France co-published a paper, this paper would be credited only once when counting papers for the EU, even though each country had been credited with one publication count.

Key to Acronyms and Abbreviations

CORD-19: COVID-19 Open Research Dataset

EU: European Union

HCA: highly cited articles

IDR: interdisciplinary research

MDR: multidisciplinary research

NASEM: National Academy of Sciences, Engineering, and Medicine

NRC: National Research Council

NSF: National Science Foundation

OSTP: Office of Science and Technology Policy

R&D: research and development

RCI: relative citation index

RCUK: Research Councils of the United Kingdom

S&E: science and engineering

SDR: NCSES Survey of Doctorate Recipients

TOD: Taxonomy of Disciplines

UK: United Kingdom

WoS: Web of Science

References

Andersen JP, Nielsen MW, Simone NL, Lewiss RE, Jagsi R. 2020. *COVID-19 Medical Papers Have Fewer Women First Authors Than Expected.* arXiv. Available at **https://arxiv.org/abs/2005.06303v3**. Accessed 24 May 2021.

Aviv-Reuven S, Rosenfeld A. 2020. *Publication Patterns' Changes Due to the COVID-19 Pandemic: A Longitudinal and Short-Term Scientometric Analysis.* Available at **https://arxiv.org/abs/2010.02594**. Accessed 24 May 2021.

Barber BM, Jiang W, Morse A, Puri M, Tookes H, Werner IM. 2021. *What Explains Differences in Finance Research Productivity during the Pandemic?* Working Paper 28493. Cambridge, MA: National Bureau of Economic Research. Available at **https://www.nber.org/papers/w28493**. Accessed 12 May 2021.

Cai X, Fry CV, Wagner CS. 2021. International Collaboration during the COVID-19 Crisis: Autumn 2020 Developments. *Scientometrics* 126:3683–92. Available at **https://doi.org/10.1007/s11192-021-03873-7**. Accessed 24 May 2021.

Campbell D, Deschamps P, Côté G, Roberge G, Lefebvre C, Archambault É. 2015. *Application of an "Interdisciplinarity" Metric at the Paper Level and Its Use in a Comparative Analysis of the Most Publishing ERA and Non-ERA Universities.* Proceedings of the 20th International Conference on Science and Technology Indicators. Available at **https://www.science-metrix.com/?q=en/publications/conference-presentations/application-of-an-interdisciplinarity-metric-at-the-paper**. Accessed 24 May 2021.

Chang WY, White K, Sugimoto CR. The Critical Role of Doctoral Research Assistantships for Students of Color. *eLife*. Forthcoming.

Chinchilla-Rodríguez Z, Sugimoto CR, Larivière V. 2019. Follow the Leader: On the Relationship between Leadership and Scholarly Impact in International Collaborations. *PLOS ONE* 14(6):e0218309. Available at **https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0218309**. Accessed 24 May 2021.

Conte ML, Liu J, Schnell S, Omary MB. 2017. Globalization and Changing Trends of Biomedical Research Output. *JCI Insight* 2(12):e95206. Available at **https://doi.org/10.1172/jci.insight.95206**. Accessed 24 May 2021.

Del Rio JA, Kostoff RN, Garcia EO, Ramirez AM, Humenik JA. 2001. Citation Mining: Integrating Text Mining and Bibliometrics for Research User Profiling. *Journal of the American Society for Information Science and Technology* 52(13): 1148–56.

Digital Science. 2016. *Interdisciplinary Research: Methodologies for Identification and Assessment.* Available at **https://www.mrc.ac.uk/documents/pdf/assessment-of-interdisciplinary-research/**. Accessed 24 May 2021.

Franzoni C, Scellato G, Stephan P. 2011. Changing Incentives to Publish. *Science* 6043:702–3. Available at **https://science.sciencemag.org/content/333/6043/702.full**. Accessed 30 June 2021.

Fry CV, Cai X, Zhang Y, Wagner CS. 2020. Consolidation in a Crisis: Patterns of International Collaboration in Early COVID-19 Research. *PLOS ONE* 15(7):e0236307. Available at **https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0236307**. Accessed 24 May 2021.

Garfield E. 1955. Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science* 122(3159):108–11.

Glänzel W, Schubert A. 2005. Domesticity and Internationality in Co-Authorship, References and Citations. *Scientometrics* 65(3):323–42.

Grudniewicz A, Moher D, Cobey KD, Bryson GL, Cukier S, Allen K, Ardern C, Balcom L, Barros T, Berger M, Ciro JB, Cugusi L, Donaldson MR, Egger M, Graham ID, Hodgkinson M, Khan KM, Mabizela M, Manca A, Milzow K, Mouton J, Muchenje M, Olijhoek T, Ommaya A, Patwardhan B, Poff D, Proulx L, Rodger M, Severin A, Strinzel M, Sylos-Labini M, Tamblyn R, van Niekerk M, Wicherts JM, Lalu MM. 2019. Predatory Journals: No Definition, No Defence. *Nature* 576:210–12 Available at **https://www.nature.com/articles/d41586-019-03759-y?fbclid=IwAR30GTZSaMn_D7bPyeQtk5hHdDXlcD4-hq8McI37TXX0PpjB1Tw4sYpSxjA**. Accessed 30 June 2021.

He T. 2009. International Scientific Collaboration of China with the G7 Countries. *Scientometrics* 80(3):571–82.

Hopkins AL, Jawitz JW, McCarty C, Goldman A, Basu NB. 2013. Disparities in Publication Patterns by Gender, Race and Ethnicity Based on a Survey of a Random Sample of Authors. *Scientometrics* 96:515–34.

Karami A, Bookstaver B, Nolan M, Bozorgi P. 2021. Investigating Diseases and Chemicals in COVID-19 Literature with Text Mining. *International Journal of Information Management Data Insights* 1(2). Available at **https://doi.org/10.1016/j.jjimei.2021.100016**. Accessed 1 July 2021.

Karimi F, Wagner C, Lemmerich F, Jadidi M, Strohmaier M. 2016. *Inferring Gender from Names on the Web: A Comparative Evaluation of Gender Detection Methods.* Available at **https://arxiv.org/pdf/1603.04322.pdf**. Accessed 24 May 2021.

Larivière V, Gingras, Y. 2014. Measuring Interdisciplinarity. *In* Cronin B, Sugimoto C, editors, *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact,* pp. 187–200. Cambridge, MA: MIT Press.

Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto CR. 2013. Bibliometrics: Global Gender Disparities in Science. *Nature* 504:211–13. Available at **https://www.nature.com/news/bibliometrics-global-gender-disparities-in-science-1.14321**. Accessed 24 May 2021.

Lee JJ, Haupt JP. 2020. Winners and Losers in U.S.-China Scientific Research Collaborations. *Higher Education* 80:57–74. Available at **https://doi.org/10.1007/s10734-019-00464-7**. Accessed 12 May 2021.

Leydesdorff L, Rafols I. 2011. Indicators of the Interdisciplinarity of Journals: Diversity, Centrality, and Citations. *Journal of Informetrics* 5(1):87–100. Available at **https://doi.org/10.1016/j.joi.2010.09.002**. Accessed 24 May 2021.

Lisée C, Larivière V, Archambault É. 2008. Conference Proceedings as a Source of Scientific Information: A Bibliometric Analysis. *Journal of the American Society for Information Science and Technology* 59(11):1776–84. Available at **https://doi.org/10.1002/asi.20888**. Accessed 28 June 2021.

Luukkonen T, Tijssen RJW, Persson O, Sivertsen G. 1993. The Measurement of International Scientific Collaboration. *Scientometrics* 28:15–36. Available at **https://doi.org/10.1007/BF02016282**. Accessed 24 May 2021.

Lyons E, Colglazier EW, Wagner CS, Börner K, Dooley DM, Mote Jr. CD, Roco MC. 2016. How Collaborating in International Science Helps America. *Science & Diplomacy* 5(2).

Marschke G, Nunez A, Weinberg BA, Yu H. 2018. Last Place? The Intersection of Ethnicity, Gender, and Race in Biomedical Authorship. *AEA Papers and Proceedings* 108:222–27. Available at **www.aeaweb.org/articles?id=10.1257/pandp.20181111**. Accessed 28 January 2021.

Myers KR, Tham WY, Yin Y, Cohodes N, Thursby JG, Thursby MC, Schiffer P, Walsh JT, Lakhani KR, Wang D. 2020. Unequal Effects of the COVID-19 Pandemic on Scientists. *Nature Human Behavior* 4:880–83. Available at **https://www.nature.com/articles/s41562-020-0921-y**. Accessed 24 May 2021.

Narin F, Hamilton K. 1996. Bibliometric Performance Measures. *Scientometrics* 36(3):293–310.

Narin F, Stevens K, Whitlow ES. 1991. Scientific Co-Operation in Europe and the Citation of Multinationally Authored Papers. *Scientometrics* 21:313–23. Available at **https://link.springer.com/article/10.1007%2FBF02093973**. Accessed 30 June 2021.

National Academy of Sciences, Engineering, and Medicine (NASEM). 2021. *Measuring Convergence in Science and Engineering: Proceedings of a Workshop.* Washington, DC: National Academies Press. Available at: **https://doi.org/10.17226/26040**. Accessed 24 May 2021.

National Center for Science and Engineering Statistics (NCSES). 2021. *Survey of Doctorate Recipients* (SDR). Alexandria, VA: National Science Foundation. Available at **https://nsf.gov/statistics/srvydoctoratework/#sd**. Accessed 14 June 2021.

National Research Council (NRC). 2014. *Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond.* Washington DC: National Academies Press. Available at **https://doi.org/10.17226/18722**. Accessed 28 May 2021.

National Science Board (NSB), National Science Foundation. 2010. *Science and Engineering Indicators 2010*. NSB-10-01. Alexandria, VA. Available at **https://www.heri.ucla.edu/PDFs/NSB.pdf**.

National Science Board (NSB), National Science Foundation. 2016. *Science and Engineering Indicators 2016*. NSB-2016-1. Alexandria, VA. Available at **https://www.nsf.gov/nsb/publications/2016/nsb20161.pdf**.

National Science Board (NSB), National Science Foundation. 2018. *Science and Engineering Indicators 2018.* NSB-2018-1. Alexandria, VA. Available at **https://www.nsf.gov/statistics/2018/nsb20181/**.

National Science Board (NSB), National Science Foundation. 2020. *Vision 2030*. NSB-2020-15. Alexandria, VA. Available at **https://www.nsf.gov/nsb/publications/2020/nsb202015.pdf**. Accessed 24 May 2021.

National Science Board (NSB), National Science Foundation. 2021a. Academic Science and Engineering Article Output per 1,000 Science, Engineering, and Health Doctorate Holders in Academia. *Science and Engineering Indicators: **State Indicators***. Alexandria, VA. **https://ncses.nsf.gov/indicators/states/indicator/academic-se-articles-per-1000-seh-doctorate-holders-in-academia**. Accessed on 1 July 2021.

National Science Board (NSB), National Science Foundation. 2021b. Academic Science and Engineering Article Output per $1 Million of Academic Science and Engineering R&D. *Science and Engineering Indicators: **State Indicators***. Alexandria, VA. **https://ncses.nsf.gov/indicators/states/indicator/academic-se-articles-per-1-million-in-academic-rd**. Accessed on 1 July 2021.

National Science Foundation (NSF). 2019. *Growing Convergence Research*. Available at **https://www.nsf.gov/pubs/2019/nsf19551/nsf19551.htm**. Accessed 24 May 2021.

Porter AL, Chubin DE. 1985. An Indicator of Cross-Disciplinary Research. *Scientometrics* 8:161–76. Available at: **https://doi.org/10.1007/BF02016934**. Accessed 24 May 2021.

Porter AL, Cohen A, Roessner JD, Perreault M. 2007. Measuring Researcher Interdisciplinarity. *Scientometrics* 72(1):117–47. Available at **https://akjournals.com/view/journals/11192/72/1/article-p117.xml**. Accessed 8 April 2021.

Royal Society. 2011. *Knowledge, Networks and Nations: Global Scientific Collaboration in the 21st Century*. RS Policy Document 03/11. London: Royal Society.

Science-Metrix. 2019. *Bibliometric Indicators for the SEI 2020. Technical Documentation*. Montreal, Canada: Science-Metrix. Available at **https://www.science-metrix.com/?q=en/publications/reports#/?q=en/publications/reports/bibliometric-indicators-for-the-sei-2020-technical-documentation**. Accessed 24 May 2021.

Science-Metrix. 2021a. *Bibliometric Indicators for the Science and Engineering Indicators 2022. Technical Documentation.* Available at **https://science-metrix.com/bibliometrics-indicators-for-the-science-and-engineering-indicators-2022-technical-documentation/**. Accessed 1 November 2021.

Science-Metrix. 2021b. *Special Tabulations of Elsevier Scopus Abstract and Citation Database.* Montreal, Canada: Science-Metrix.

Sugimoto CR, Larivière V. 2018. *Measuring Research: What Everyone Needs to Know*. United Kingdom: Oxford University Press.

Sugimoto CR, Robinson-Garcia N, Murray DS, Yegros-Yegros A, Costas R, Larivière V. 2017. Scientists Have Most Impact When They're Free to Move. *Nature* 550(7674):29–31. Available at **https://www.nature.com/news/scientists-have-most-impact-when-they-re-free-to-move-1.22730**. Accessed 24 May 2021.

Wagner CS. 2018. *The Collaborative Era in Science: Governing the Network.* London: Palgrave Macmillan.

Wagner CS, Roessner JD, Bobb K, Klein TJ, Boyack KW, Keyton J, Rafols I, Börner K. 2009. *Evaluating the Output of Interdisciplinary Scientific Research: A Review of the Literature.* Unpublished report to National Science Foundation.

Wagner CS, Roessner JD, Bobb K, Klein JT, Boyack KW, Keyton J, Rafols I, Börner K. 2011. Approaches to Understanding and Measuring Interdisciplinary Scientific Research (IDR): A Review of the Literature. *Journal of Informetrics* 5(1):14–26. Available at **https://www.sciencedirect.com/science/article/abs/pii/S1751157710000581**. Accessed 13 May 2021.

Wallin JA. 2005. Bibliometric Methods: Pitfalls and Possibilities. *Basic & Clinical Pharmacology & Toxicology* 97(5):261–75. Available at **https://doi.org/10.1111/j.1742-7843.2005.pto_139.x**. Accessed 24 May 2021.

Waltman L, van Eck NJ, Wouters P. 2013. Counting Publications and Citations: Is More Always Better? *Journal of Informetrics* 7(3):635–41. Available at **www.sciencedirect.com/science/article/abs/pii/S1751157713000357**. Accessed 28 January 2021.

Wang J. 2012. Citation Time Window Choice for Research Impact Evaluation. *Scientometrics* 94:851–72.

Wang J, Thijs B, Glänzel W. 2015. Interdisciplinarity and Impact: Distinct Effects of Variety, Balance, and Disparity. *PLOS ONE* 10(5):e0127298. Available at **https://doi.org/10.1371/journal.pone.0127298**. Accessed 24 May 2021.

Wang LL, Lo K, Chandrasekhar Y, Reas R, Yang J, Burdick D, Eide D, Funk K, Katsis Y, Kinney R, Li Y, Liu Z, Merrill W, Mooney P, Murdick D, Rishi D, Sheehan J, Shen Z, Stilson B, Wade A, Wang K, Wang NXR, Wilhelm C, Xie B, Raymond D, Weld DS, Etzioni O, Kohlmeier S. 2020. *CORD-19: The COVID-19 Open Research Dataset.* Available at **https://arxiv.org/abs/2004.10706**. Accessed 12 May 2021.

Wang Q, Schneider JW. 2020. Consistency and Validity of Interdisciplinarity Measures. *Quantitative Science Studies* 1(1): 239–63. Available at **https://doi.org/10.1162/qss_a_00011**. Accessed 24 May 2021.

Woolston C. 2020. Pandemic Darkens Postdocs' Work and Career Hopes. *Nature* 585:309–12. Available at **https://www.nature.com/articles/d41586-020-02548-2**. Accessed 12 May 2021.

World Bank. 2021. *World Bank Country and Lending Groups.* Available at **https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups**. Accessed 12 August 2021.

Xie Q, Freeman RB. 2018. *Bigger Than You Thought: China's Contribution to Scientific Publications.* Working Paper 24829. Cambridge, MA: National Bureau of Economic Research. Available at **http://www.nber.org/papers/w24829**. Accessed 24 May 2021.

Yin Y, Dong Y, Wang K, Wang D, Jones B. 2021. *Science as a Public Good: Public Use and Funding of Science.* Working Paper 28748. Cambridge, MA: National Bureau of Economic Research. Available at **https://www.nber.org/papers/w28748?utm_campaign=ntwh&utm_medium=email&utm_source=ntwg15**. Accessed 5 May 2021.

Notes

1 It is important to note that the biases, introduced from either inclusion of non-peer-reviewed articles or preferential inclusion of English-language journals, cannot be statistically measured nor 100% removed. The publication database contains administrative data that are collected for purposes of cataloging research output, as opposed to a statistical population survey in which errors could be specified and measured.

2 Publication output only includes those indexed in the Scopus database. The publication output discussion uses fractional counting, which credits coauthored publications according to the collaborating institutions or countries based on the proportion of their participating authors. Country assignments refer to the institutional address of authors, with partial credit given for each international coauthorship. As part of our data analysis, we employ filters on the raw Scopus S&E publication data to remove publications with questionable quality, which appear in what are sometimes called *predatory* journals (NSB ***Indicators 2018*: Bibliometric Data Filters sidebar**).

3 This report uses the World Bank (2021) country income classifications accessed in March 2021. The World Bank updates the classifications each year on 1 July. The World Bank income classifications are assigned using the gross national income per capita as measured in current U.S. dollars. This report uses the rankings. More information is available at **https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups**.

4 It is possible that the growth rates could be influenced by fractional counting. For example, the compound annual growth rate for France using whole counting is 1%. Publication output using whole counting is available in Table SPBS-17.

5 The proportion of output attributable to the large producers is consistent whether using fractional counting, as in Figure PBS-2 and Table PBS-1, or whole counting, as in Table SPBS-17. There is a slight difference between the United States and China when looking at the whole counting total production numbers. Using whole counting for 2020, the United States had 600,053 articles, while China had 742,431. A whole counting measure allocates one full count to each country with an author contributing to the article; in fractional counting, each country receives a proportion of the count based on the number of authors from that country. For example, if an article had four authors—with two from the United States, one from China, and one from Brazil—the fractional scores would be 2/4 for the United States, 1/4 for China, and 1/4 for Brazil. In this example, the difference between whole and fractional counting indicates that the United States had more authors on the example paper, compared to the number of authors in China or Brazil.

6 There is little difference between whole or fractional counting of publications for the large producing countries. Whole counting shows a difference for small countries with high collaboration rates because they only receive a fraction of a point for each article, while whole counting awards them a full point (Table SPBS-17 through Table SPBS-31).

7 European Commission, *Horizon Europe,* **https://ec.europa.eu/info/research-and-innovation/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en**.

8 The total international collaboration rate differs from the co-publications of individual countries. Individual country scores use whole counting in which each country, with an institutional address on the paper, receives one point. Therefore, the basis count for individual countries will be larger than the number of papers that have authors with institutional addresses from two or more countries, which is the basis for computing the total international collaboration rate.

9 Countries contributing less than 1% of all internationally coauthored papers in 2020 are not included in the analysis.

10 Relative citation index is a citation-based measure of scientific influence. It is calculated from a country's share of all cited S&E publications divided by the other country's share of all cited S&E publications; an index less than 1.00 means a lower-than-expected tendency to cite the other country.

11 Publications include some peer-reviewed journal articles that receive citations as examples of poor or outdated research, but often the articles refuting the low-quality article receive many more citations (Wallin 2005).

12 The share of S&E articles in the top 1% of cited articles is computed by field because different fields of science have different rates of citation. Details provided in Science-Metrix (2021a).

13 Table SPBS-73 and Table SPBS-89 provide the data to compare the computer and information sciences publications in the top 1% for both journal articles and conference proceedings with the HCA scores with only journal articles. For the large producing countries, the HCA that include conference proceedings show the same trend but with slightly higher HCA scores as the HCA with only journal articles. This comparison is made for computer and information sciences because conference proceedings are close to 20% of the references in which proceedings are much less cited (Lisée, Larivière, and Archambault 2008).

14 The share of top 1% articles produced by a country is computed as follows: *Sx = HCAx / Ax*, where *Sx* is the share of output from country x in the top 1% most-cited articles; *HCAx* is the number of articles from country *x* that are among the top 1% of most-cited articles; and *Ax* is the total number of articles from country *x* with a relative citation (RC) score. Both *HCAx* and *Ax* are based on whole counting. The RC score is a normalized citation score assigned to a paper and used to rank articles into the top 1%. The RC score takes into consideration the citation behavior between fields and years of publication.

15 Computing the HCA score using fractionally counted publications gives a more exact view of performance for countries with relatively fewer articles.

Acknowledgments and Citation

Acknowledgments

The National Science Board extends its appreciation to the staff of the National Science Foundation and to the many others, too numerous to list individually, who contributed to the preparation of this report.

This report was produced under the leadership of Amy Burke, Program Director, Science and Engineering Indicators Program of the National Center for Science and Engineering Statistics (NCSES); Emilda B. Rivers, Director, NCSES; and Vipin Arora, Deputy Director, NCSES.

The report benefitted from extensive contributions from NCSES staff. Wan-Ying Chang provided advice on statistical issues. Carol Robbins and Karen White served in administrative roles. May Aydin, Catherine Corlies, and Rajinder Raut coordinated the report's publication process and managed the development of its digital platform. Christine Hamel and Tanya Gore conducted editorial and composition review.

SRI International, Center for Innovation Strategy and Policy, assisted with report preparation. Jessica Avery and Mia Thomas of SRI International provided especially large contributions to the report. August Gering at RTI International provided editing services. Michelle Goryn at SRI International provided editorial assistance. Staff at Penobscot Bay Media, LLC (PenBay Media), created the report site. The following persons and agencies reviewed this report:

Nicole Gingrich, National Institute of Standards and Technology

Vincent Larivière, L'Université de Montréal

Ya-Ling Lu, U.S. Food and Drug Administration

Francis Narin, CHI Research (retired)

Ismael Rafols, Leiden University

Caroline S. Wagner, The Ohio State University

National Science Foundation

Office of Science and Technology Policy

The National Science Board is especially grateful to the Committee on National Science and Engineering Policy for overseeing preparation of the volume and to the National Science Board Office, under the direction of John Veysey, which provided vital coordination throughout the project. Nadine Lymn led the outreach and dissemination efforts. Reba Bandyopadhyay and Portia Flowers served as Board Office Liaisons to the committee. Carol Robbins, Anne Emig, and May Aydin were the Executive Secretaries.

Citation

National Science Board, National Science Foundation. 2021. Publications Output: U.S. and International Comparisons. *Science and Engineering Indicators 2022*. NSB-2021-4. Alexandria, VA. Available at https://ncses.nsf.gov/pubs/nsb20214/.

Contact Us

Report Author

Karen White
Senior Analyst
Science and Engineering Indicators Program, NCSES
Tel: (703) 292-4344
E-mail: kewhite@nsf.gov

NCSES

National Center for Science and Engineering Statistics
Directorate for Social, Behavioral and Economic Sciences
National Science Foundation
2415 Eisenhower Avenue, Suite W14200
Alexandria, VA 22314
Tel: (703) 292-8780
FIRS: (800) 877-8339
TDD: (800) 281-8749
E-mail: ncsesweb@nsf.gov