SIDEBAR

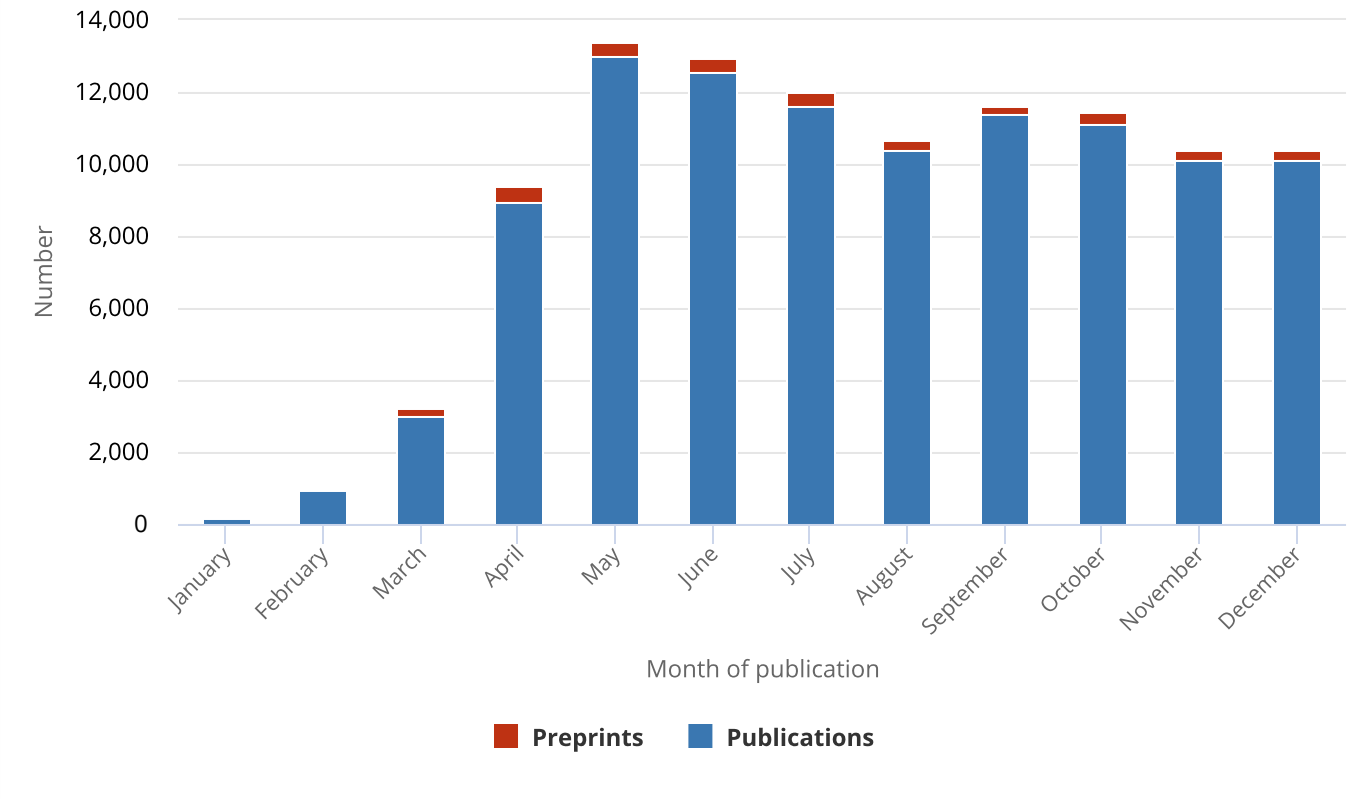
Coronavirus Publication Output and International Collaboration

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.

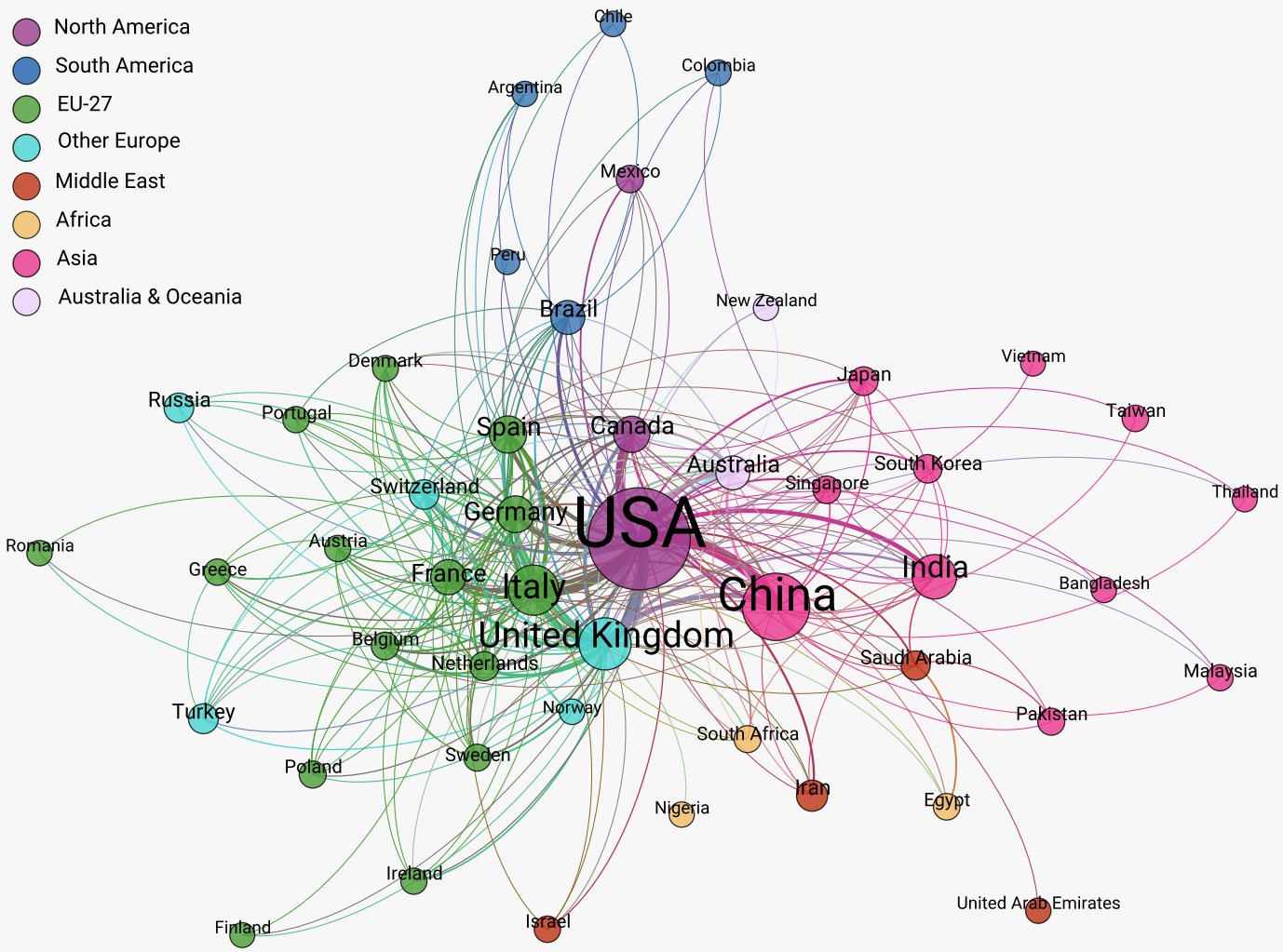
Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed May 2021.

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.