Institute for Scientific Information

Research Professional News

May 2023

Insights: Coronaviruses

A 21st-century research phenomenon

*An **Institute for Scientific Information** report, published in partnership with **Research Professional News***

Clarivate™

About the Institute for Scientific Information

The Institute for Scientific Information™ at Clarivate has pioneered the organisation of the world's research information for more than half a century. Today it remains committed to promoting integrity in research while improving the retrieval, interpretation and utility of scientific information. It maintains the knowledge corpus upon which the Web of Science™ index and related information and analytical content and services are built. It disseminates that knowledge externally through events, conferences and publications, while conducting primary research to sustain, extend and improve the knowledge base.

For more information, please visit the ISI website.

www.clarivate.com/isi

About Research Professional News

Research Professional News™ is an award-winning, editorially independent news service, published by Clarivate. Our international team of specialist journalists provides exclusive, timely and trusted information focused on research policy, research funding and higher education around the world.

Our news provides invaluable insight, delivered through flexible technology to everyone in your institution. That means sharper awareness of the policy and funding environment in which you are operating, putting you in a position to make better decisions about where to invest, what funding to aim for, who to hire and what to do next.

Research Professional News takes an objective, independent view of events, bringing you insights informed by a deep network of policy contacts around the world.

To find out more and read our latest headlines, visit the Research Professional News website.

www.researchprofessionalnews.com

About this paper

The 21st century has witnessed three separate coronavirus pandemics, with Covid-19 causing the most significant global health crisis since the Spanish Flu pandemic a century ago. In this paper, attention is drawn to the phenomenal increase in coronavirus research this century and how this compares to other contemporary pandemics. Covid-19 research has been unprecedented in its magnitude and truly global reach. Research is dominated by the US and China, but localised effects have driven some countries to greatly invest in research. Topics within coronavirus research are wide-ranging and highlight important societal effects such as mental health, education and public awareness.

DOI number: https://dx.doi.org/10.14322/ISI.insight.3

About the author

Ross Potter is a lead data scientist at the Institute for Scientific Information.

He has extensive research experience within academia, including Nasa-related postdoctoral positions at the Lunar and Planetary Institute, Houston, Texas, and Brown University, Providence, Rhode Island.

ORCiD: 0000-0002-1338-5910

Web of Science ResearcherID: R-359 0-2019

ISI Insights: Report highlights

- Although present prior to 2000, research on coronaviruses has expanded rapidly and significantly in the 21st century due to three separate coronavirus pandemics.
- Zika and H1N1 virus research offers an important contemporary comparison for benchmarking coronavirus research.
- Covid-19 research has been unprecedented, dwarfing other (corona)virus research by orders of magnitude, illustrating the pandemic's truly global effects.
- The US and China are the main research hubs in terms of scientific publications and patents, although some countries have high shares of research due to localised pandemic effects (e.g. Saudi Arabia and Middle East Respiratory Syndrome).
- More than two-thirds of patents for coronavirus vaccines and treatments have been filed since 2019.
- Clinical trials featuring experimental coronavirus vaccines and antivirals have shifted from North America and Europe to the Asia-Pacific region, although the US still leads the way.
- Topic modelling reveals distinct and wide-ranging areas of coronavirus research—in particular, more societal aspects of pandemics, including mental health, education and media, directly related to UN Sustainable Development Goals.
- Topic-modelling clusters represent pillars of pandemic research that are vital to fully understanding any virus, informing and caring for the public and successfully managing future pandemics.

Bringing coronaviruses to public attention

The coronavirus disease Covid-19, caused by the severe acute respiratory syndrome coronavirus 2 (Sars-Cov-2), has led to the most significant global health crisis since the Spanish Flu pandemic a century ago. Over 750 million cases have been confirmed globally, including more than 6.5 million deaths. Many countries have now removed virus-related restrictions. However, new mass outbreaks cannot be ruled out, as witnessed in China following the removal of its zero-Covid policies.

Covid-19 may have brought coronaviruses to the forefront of public attention, but four coronaviruses are endemic in humans, though effects are commonly mild to moderate. This century has seen two other coronavirus pandemics: Severe Acute Respiratory Syndrome (Sars: 2002-04) and Middle East Respiratory Syndrome (Mers: 2012 onwards).

Coronaviruses form the Coronaviridae family of RNA viruses. This distinguishes them from other viruses such as Zika (Flaviviridae family) and H1N1 (aka 'swine flu', a strain of which caused the Spanish Flu; Orthomyxoviridae family), which have also produced 21st-century pandemics.

These viruses initially crossed from animals to humans, with subsequent rapid transmission through populations. As the world becomes increasingly globalised and urbanised, bringing humans and animals ever closer together, pandemic frequency is likely to increase. Assembling scientific expertise, innovation and technology is vital to fully understanding each virus and its variants, producing vaccines and informing the public to lessen

> **"As the world becomes increasingly globalised, pandemic frequency is likely to increase."**

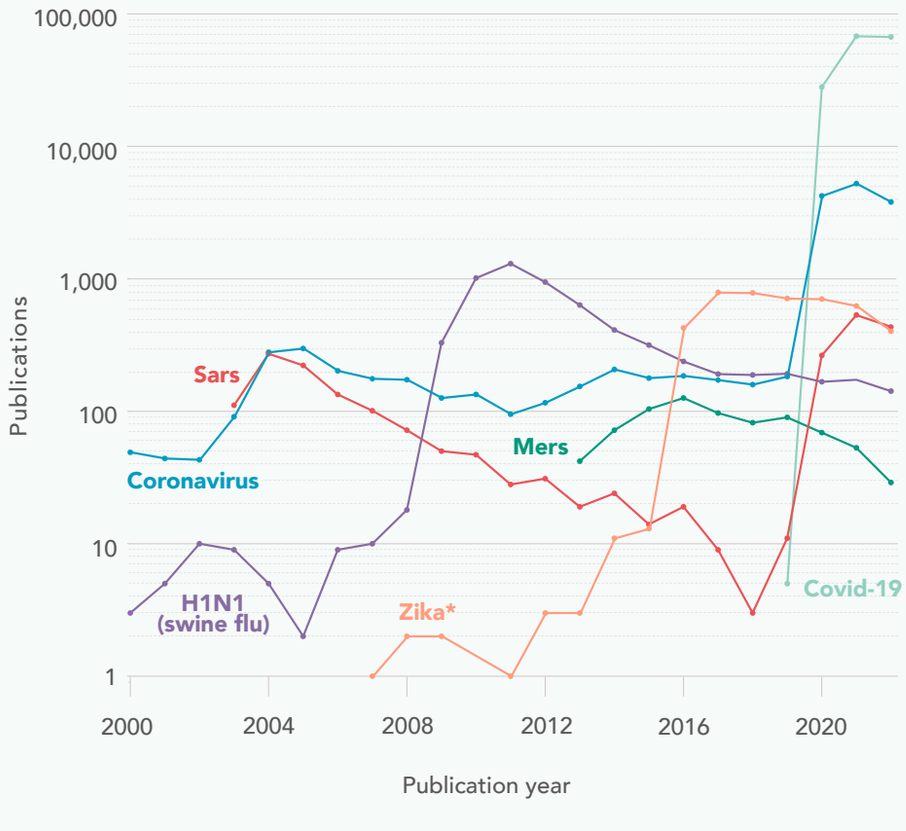
the crippling effects of future pandemics. Such work falls under the UN's 2030 Agenda for Sustainable Development, particularly the UN Sustainable Development Goals concerning good health and wellbeing (SDG 3) and sustainable cities and communities (SDG 11). All SDGs are interconnected in their shared goal of fostering a sustainable future for all life on earth. The Covid-19 pandemic has stalled or reversed progress in health matters such as child health (SDG 3) and disproportionately affected those living in the poorest conditions (e.g. slums; SDG 11).

In this ISI Insights paper, an overview of coronavirus research in the 21st century is provided to chart research evolution through trends and changes across the three coronavirus pandemics (Sars, Mers, Covid-19), using Zika and H1N1 research as comparators. To pinpoint the most relevant coronavirus documents, our Web of Science analysis focused on articles containing specific coronavirus-related terms within their titles and abstracts, published from 2000 to 2022. This provided a dataset of ~180,000 articles, with an additional ~4,000 articles retrieved for Zika and ~6,000 for H1N1, through a similar retrieval process.

Coronavirus research grew rapidly as pandemics began

The extent of coronavirus research since 2000 can be illustrated using specific article title terms (Figure 1). Such an approach may not return all coronavirus research articles but identifies relevant articles with a high degree of precision. Research on coronavirus(es) as a general term was steady, though at low volumes, in the early 2000s. An increase in papers occurred during the Sars pandemic (2002-04) before stabilising at one to several hundred papers per year. The onset of Covid-19 saw a significant spike, taking it to over 5,000 publications in 2021. Similarly, research on H1N1 was present in the early 2000s before peaking in 2011 at ~1,300 papers following the H1N1 pandemic (2009-10); output has since dropped by an order of magnitude.

Figure 1



Yearly publication of research articles containing specific title words relating to coronavirus (as a general term), Covid-19, Middle East Respiratory Syndrome (Mers), Severe Acute Respiratory Syndrome (Sars), Zika and H1N1.

**Zika had no papers in 2010*

From its onset in 2002, the Sars pandemic spurred a new line of research, peaking in

> **“Covid-19 research has been produced on an entirely different scale, reflecting the pandemic’s impact.”**

2004 when 272 papers were published. Thereafter, the data show a steady decline to just three papers in 2018, before a significant increase as Covid-19 began, with 533 papers published in 2021. Mers similarly generated a new line of research, emerging in 2013 and peaking in 2016, when 124 papers were published on the topic; output has recently declined though the virus continues to circulate.

For comparison, the Zika virus emerged in 2007 for the first time in 25 years due to an outbreak on the Micronesian island of Yap. Subsequent outbreaks on other Pacific islands and within Asia saw a handful of papers published before 2015, when the virus emerged in Brazil and grew into a pandemic. As the virus spread throughout the Americas from 2015 to 2016, research increased dramatically, peaking at 789 papers in 2017; output had almost halved by 2022.

Covid-19 research, however, has been produced on an entirely different scale, increasing from only a handful of papers in 2019 to nearly 28,000 papers in 2020 and then to almost 68,000 in 2021, with that level maintained in 2022. This reflects the truly global extent and impact of the Covid-19 pandemic relative to the other viruses.

Table 1

Country	Coronavirus	Covid-19	H1N1	Mers	Sars	Zika
Argentina	0.45	0.59	1.00	0.00	0.27	0.89
Australia	2.70	3.74	4.77	2.62	2.14	2.50
Brazil	2.55	3.54	2.41	0.13	1.90	24.34
Canada	3.41	4.35	5.73	5.10	5.00	4.51
China	19.02	11.52	14.43	15.58	24.52	9.92
France	2.91	3.34	4.01	2.88	3.26	7.31
Germany	4.78	4.92	3.90	7.20	4.28	4.36
India	4.62	6.17	4.63	1.05	3.54	3.24
Indonesia	0.64	1.15	0.22	0.13	0.48	0.40
Italy	5.50	7.66	3.67	0.92	3.91	2.93
Japan	4.72	2.66	5.69	3.27	5.54	1.59
Mexico	0.69	1.36	2.20	0.00	0.41	2.35
Russia	1.47	1.09	1.19	0.52	0.31	0.51
Saudi Arabia	3.61	2.95	1.09	27.36	1.19	1.03
South Africa	1.02	1.42	0.35	0.39	0.71	0.8
South Korea	3.64	2.05	4.89	20.42	2.24	1.59
Turkey	3.41	3.85	2.20	0.26	0.95	0.42
UK	7.02	9.41	6.57	6.54	5.30	8.29
US	29.17	25.72	26.87	34.69	33.8	46.7
Colombia	0.32	0.63	0.25	0.13	0.48	3.58
Iran	2.86	2.46	1.20	1.18	0.99	0.31
Singapore	1.65	0.95	1.87	1.05	3.81	2.99
Taiwan	1.89	1.11	2.12	1.57	5.44	1.18
UAE	0.71	0.76	0.06	3.53	0.07	0.11

Percentage share of total global research articles (2000-22) for each of five viruses: Covid-19, Mers, Sars, H1N1 and Zika, as well as all (general) coronavirus research.

Virus research output demonstrates international, as well as national, effects and concern

Data on authors' addresses provide insight into the countries conducting virus-related research. Table 1 shows the global percentage share of research articles for the G20 countries and five comparators across the five viruses, plus 'coronavirus' in general terms. The US ranks as the most productive country for each virus, with its overall share of articles consistently around 30 per cent, although almost 50 per cent of Zika articles were published by the US. As the US is the largest global research producer, this finding is not unexpected—but it does demonstrate that US output regarding these viruses is roughly in line with its overall world share of articles.

China (including Hong Kong and Macau) also produces a significant share of research output for all viruses, particularly Sars (whose origin was in China), where its share is more than a quarter. China's output agrees with its global status as the second most productive nation during the period analysed and is, again, roughly in line with its overall world share of research output (~17 per cent). G7 countries (excluding the US) such as the UK and Germany have produced ~4-9 per cent of research on each virus.

This pattern is reflected in patent filings for coronavirus vaccines and treatments, with the US and China seeing far and away the most filings (6,349 in the US and 5,149 in China), followed by the EU's European Patent Office (947) and South Korea (814). More than two-thirds of patents for coronavirus vaccines and treatments have been filed since 2019, with 8,598 filed in 2020 and another 2,127 filed in 2021.

Outside China and the G7, countries generally have smaller (<5 per cent) shares of virus research. However, there are important exceptions where country-specific or regional epidemics provided an imperative to conduct or be involved in research.

Mers was first reported in Saudi Arabia and has, as of January 2023, claimed the lives of 854 Saudis, over 90 per cent of all deaths. Hence, Saudi Arabia accounts for a significant share of global research (~27 per cent) into the virus, which has been particularly prevalent on the Arabian Peninsula (the United Arab Emirates has a ~3.5 per cent share of research, a larger proportion than Japan and Australia). There was also a significant outbreak in South Korea, which helps explain its ~20 per cent of published Mers research. Mers has been relatively isolated, with 27 countries confirming cases, though it continues to kill.

> **"Important exceptions show where country-specific epidemics provided a research imperative."**

Sars spread quickly throughout Asia after appearing in China; Taiwan and Singapore account for ~5 per cent and ~4 per cent of Sars research, respectively (both countries recorded more than 30 deaths). The virus did spread to other regions, most notably Canada, where a significant outbreak occurred in Toronto. However, the Canadian share of Sars research (~5 per cent) is similar to that of other viruses. Sars was also fairly isolated, with 8,096 people in 29 countries contracting the virus and 774 deaths in 2003—the height of the pandemic. However, since 2004 there have not been any known cases of Sars reported globally.

For the Zika virus, Brazil's global share of research is ~24 per cent, making Brazil the second-biggest contributor globally, behind the US. This demonstrates the Americas' focus on the virus. Colombia, a more emerging research economy, accounts for ~3.6 per cent of Zika research (its global share for the other viruses is <1 per cent). Argentina, however, which was less affected by Zika, accounts for only ~1 per cent of research (comparable to its H1N1 output), while Mexico, which was more affected, contributes only ~2.5 per cent of global research (also comparable to its H1N1 output and roughly double its Covid-19 share). To date, a total of 89 countries and territories

have reported evidence of the Zika infection. During the pandemic, 51 deaths were reported, with a mean case fatality rate of 0.02 per cent in the Americas (*Cardona-Ospina et al., 2019, International Journal of Infectious Diseases*). The H1N1 pandemic was more widespread, with cases confirmed in 74 countries by June 2009, having been first identified in April 2009. However, cases were eventually confirmed in at least 171 countries and territories. In the US alone (where the virus was first discovered), ~12,500 deaths were recorded from April 2009 to April 2010. This greater global reach is

> **"The UK has suffered the greatest number of deaths in Europe in absolute terms."**

reflected in the wider and larger research distribution for many G20 countries, with Australia, Canada, France, India and South Korea among those having a share of >3.5 per cent. Covid-19 has reached even more countries and territories than H1N1, with only North Korea and Turkmenistan reporting zero confirmed cases to the World Health Organization. The US and China have lower output shares, relative to almost all the other viruses, signifying less research dominance. Research output is spread between most of the G20 countries, with India (~6 per cent), Turkey (~4 per cent), the UK (~9.5 per cent), Italy (~7.5 per cent) and Indonesia (though only ~1 per cent) having their greatest contribution of any virus. Italy was the initial epicentre of the virus in Europe, and the UK has suffered the greatest number of deaths in Europe (in absolute terms). Iran, with a global research share of ~2.5 per cent, was also an early epicentre of the virus. This is reflected in its share, which is twice that of its Mers and Sars contributions.

Given the global reach of these pathogens, clinical trials for vaccine and treatment candidates were surprisingly Western-focused prior to the eruption of Sars-Cov-2 in 2019—from 2002 to 2018, there were 52 of these trials in the US, 22 in the UK, 11 in Canada and 10 in Germany. In 2020, when trials for Covid-19 treatments peaked, the US still hosted the most trials (1,130) but was followed closely by mainland China (1,004) and India (914).

This hive of activity has impacted the world of intellectual property as well—since 2005, a total of 17,443 families of patents for coronavirus vaccines and treatments have been filed worldwide.

Topic modelling illustrates wide-ranging research

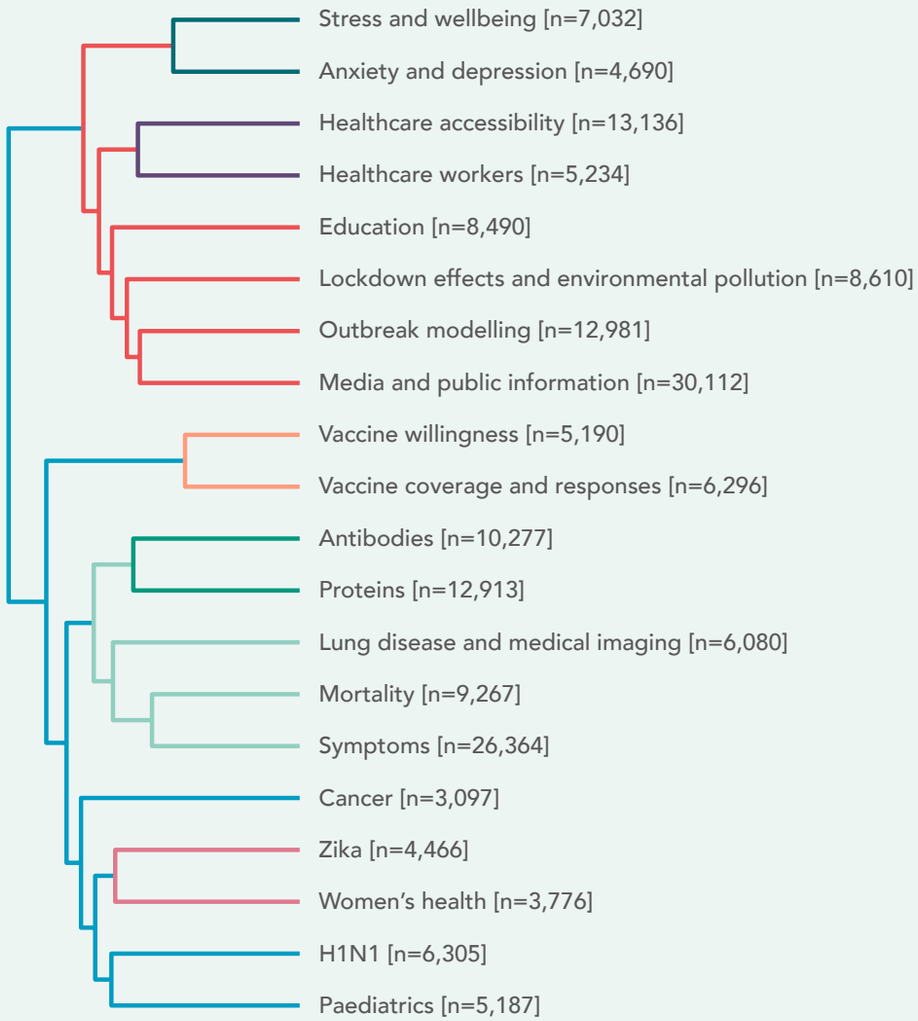
Covid-19 has wrought significant effects on public health systems and damage to economies (at all scales), not to mention its impact on mental health and wellbeing. To illustrate the range and prevalence of Covid-19 topics researched and their links, a topic model was produced based on words within all the coronavirus-relevant article titles, abstracts and keywords.

The topic model focused on unigrams (i.e. single words) and considered those which, after removing common words, appeared in at least six documents but no more than 30 per cent of the entire

> **"Covid-19 has wrought significant effects on public health systems, economies and mental health."**

corpus. This set contained just over 60,000 words, allowing information to be distilled down to key areas. How often each word appeared in the title, abstract and keywords was used as a weight input to a non-negative matrix factorisation algorithm, which identified a specified number of topics, partitioned based on the frequency with which terms appear together (i.e. their similarity). Twenty topics were chosen based on the topic size and the coherency and granularity of the topic terms; topic labels were assigned indicatively.

Figure 2



Dendrogram showing the similarity between 20 topics identified among ~180,000 articles using a non-negative matrix factorisation algorithm. The number of articles in each topic is shown to the right of the topic label. Colours indicate topics that cluster together with relatively high degrees of similarity.

Figure 2 illustrates the resulting dendrogram. The branches of the dendrogram represent the different topics, which cluster together to form a larger family tree. Topics clustering together are more similar in their use of terminology than topics that are far apart in the dendrogram. Topics are divided into approximately eight clusters (discriminated by colour) and two main branches.

The top branch consists of three clusters. The top cluster covers mental health, demonstrating the negative effects of the coronavirus on society; the second cluster covers healthcare, namely workers (who dealt directly with Covid-19 patients) and accessibility (the rise of telehealth as well as services and surgeries—many of which were cancelled or postponed due to Covid-19). The third cluster illustrates the close links between the effects of lockdown, such as severe disruption to schooling and reductions in air pollution from lower traffic levels, as well as outbreak modelling and public information, demonstrating the accessibility and importance of coronavirus science in mass communication. The lower branch covers more clinical and preventative medicine topics. One cluster covers vaccines, which includes contrasting public perceptions towards the safety and efficacy of vaccines (willingness) as well as the extent of vaccine rollouts and responses. Another cluster focuses on immunity (antibodies and proteins); cancer has its own cluster, representing its pervasive effect on society. Diagnostics, symptoms and outcomes are also covered in one cluster.

The Zika virus and H1N1 have their own distinct topics, demonstrating that, relative to Covid-19, they are smaller though prevalent and unique topics. Given that the Zika virus can pass through the placenta, affecting foetuses, it is no surprise that the topic is clustered with women's health (where perinatal care is a major subtopic). H1N1 is related to paediatrics; research found that children were more likely to catch 'swine flu' than adults (*Cauchemez et al., 2009, New England Journal of Medicine*).

The threat of further pandemics remains, requiring specific foci for research efforts

Coronaviruses have been a significant 21st-century phenomenon but not the only cause of post-2000 pandemics. However, due to continued global urbanisation and connectivity, the threat of further pandemics remains. This analysis shows the tremendous scale and magnitude of research and patents, particularly around Covid-19, mirroring pandemics' far-reaching global effects.

Though the US and China dominate research and patents, other countries' output can be directly linked to localised virus effects, such as Mers research in Saudi Arabia and Zika in Brazil.

The topic model illustrates the wide-ranging issues that have been researched. As well as clinical and preventative medicine, these include the emergence of the more social aspects of pandemics: mental health, wellbeing, public information and crisis management. These also align with the UN's Sustainable Development Goals on good health and wellbeing (SDG 3) and sustainable cities and communities (SDG 11). Ultimately, the topic clusters represent pillars of pandemic research that are vital to fully understanding any virus, informing and caring for the public and successfully managing future pandemics; they should be the foci of future research efforts.

> **"The topic clusters represent pillars that should be the foci of future research efforts."**

About Clarivate data sources used in this report

This report features data from several Clarivate datasets, including:

- Citations data on published research from the Web of Science, a trusted independent global citation database allowing researchers, publishers and academicians to understand the flow of scientific information.
- Clinical trial data from Cortellis Clinical Trials Intelligence™, which helps life science companies to identify trial sites and investigators, design successful trial protocols, choose appropriate endpoints and support patient recruitment efforts.
- Patent filing data from the Derwent World Patents Index™, which enables R&D teams, intellectual property professionals and national patent offices worldwide to identify relevant patents and facilitate innovation.

Institute for Scientific Information

published in partnership with

Research Professional News

Clarivate™

DOI number: https://dx.doi.org/10.14322/ISI.insight.3