

COVID-19 Times: Impact on Information Generation and Data Sharing

Eden Rodríguez-Flores¹, Saúl Gómez-Manzo², Jaime Marcial-Quino³,
Roberto Arreguín-Espinosa¹, Eduardo Rodríguez-Bustamante^{1,4*}

¹Departamento de Química de Biomacromoléculas, Instituto de Química, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

²Laboratorio de Bioquímica Genética, Instituto Nacional de Pediatría, Secretaría de Salud, Ciudad de México, Mexico

³Investigador Independiente en Área de Salud, Ciudad de México, Mexico

⁴Departamento de Bioingeniería, Escuela de Ingeniería y Ciencias, Tecnológico de Monterrey, Campus Ciudad de México, Monterrey, Mexico

Email: *erodriguezb@gmail.com

How to cite this paper: Rodríguez-Flores, E., Gómez-Manzo, S., Marcial-Quino, J., Arreguín-Espinosa, R. and Rodríguez-Bustamante, E. (2022) COVID-19 Times: Impact on Information Generation and Data Sharing. *Advances in Internet of Things*, 12, 65-87.

<https://doi.org/10.4236/ait.2022.123005>

Received: May 11, 2022

Accepted: July 9, 2022

Published: July 12, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

At the beginning of 2020, human activities were interrupted by a new virus, identified as SARS-CoV-2, which causes COVID-19 disease. The scientific area was no exception: for a certain period, researchers around the world were forced to leave their laboratories and work remotely. There was a global necessity for finding alternatives focused on generating knowledge and publishing data, so repositories of scientific information, such as databases, represented strong support. In the specific case of life sciences, different strategies allowed rapid compilation of data and its sharing worldwide. Therefore, in this work, the impact of the SARS-CoV-2 pandemic on the amount of peer-reviewed and published papers during COVID-19 times was analyzed along with the role of databases. Our results pointed out that an increase in the number of papers belonging to different knowledge fields took place, with the medical field being the most significant. On the other hand, the complete genome of the new virus was sequenced, and repositories were created with sufficient data for monitoring, preventing, and controlling its dissemination. This was the case for the generation of vaccines in addition to potential candidates for drugs against COVID-19. However, although in 2021, vaccines allowed us to gradually return to our activities, databases and the generation of other repositories remain a key point for facing new strains and adapting to a new reality. Finally, this paper discusses joint efforts to tackle the obstacles of the pandemic, not only from a medical but also from the point of view regarding the fight against misinformation.

Keywords

COVID-19, Laboratories Closure, Data Availability, Information Repositories, Research Papers, Infodemics, Vaccines, Variants and Reopening

1. Introduction: COVID-19 Times and Their Effect Brought upon Research

The Coronavirus family is known to affect animals, causing respiratory affections in humans that range from common cold symptoms to severe infections [1]. The year 2020 marked a dramatic breakpoint in humanity, revolutionizing aspects of our lifestyle such as health, education, economics, politics, and, especially, people interactions. This worldwide situation was due to a new virus named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the main agent cause of the infectious disease COVID-19.

According to WHO (1), SARS-CoV-2 is transmitted primarily through droplets of saliva or nose discharges when infected people cough or sneeze. In order to prevent spreading, physical distance along with the use of face masks and the frequent habit of hand washing, were highly recommended by health and medical institutions, as well as by governments from different countries, giving birth to a new reality. Despite these measures, until December 31, 2020, globally, the number of deaths caused by SARS-CoV-2 exceeded one million and a half, and more than 75 million people were infected [1]. While during 2021, a total number of 290 million infected persons and 5.45 million deaths were recorded until December 31 [2], the present year stills shows increasing tendency of these numbers. Outbreaks in countries where the infection had already decreased, indicate that the disease does not seem to give up, a situation reflected by the second, third, and fourth waves of the pandemic. In this sense, the development and application of vaccines represent a new hope for humanity in its desire to go back to the lifestyle prior to COVID-19, but at the same time, the appearance of new strains (variants) of SARS-CoV-2, once again, generates a new global challenge for the control of this virus.

Universities, institutes, and other research centers were no exception with respect to the impact of COVID-19, and scientists across the globe were compelled to leave their laboratories and work from home. In 2020, in addition to many questions that arose among the scientific community, mainly focused on how to help during this crisis, an important issue was the lack of certainty on where research could be conducted. The scientific community has been on the front line since the beginning of the pandemic; faced with a new virus, it required the use of all available tools to deal with the global situation. As a response to this matter, attention was drawn to dry labs, leaving wet labs in temporary standby status. Dry labs are spaces designed to work with dry stored material, electronics, and/or large instruments with few pipe services [3], designed for making simula-

tions or data analysis with the aid of computers. On the other hand, wet labs are characterized by the role of liquid solutions in carrying out various kinds of experiments.

A survey conducted by Korbel and Stegle [4], and directed to the scientific community from Canada, France, Italy, Germany, Spain, the UK, and the USA, showed a loss of between 1 to 6 months of work due to laboratory closures, with a notable difference between wet and dry labs. While wet laboratory researchers reported a 73% loss in productivity, the decrease in dry laboratories was only 31% [4]. Due to frequent interruptions, wet labs experienced a larger productivity decrease, in contrast to dry labs, which were able to continue their work with remote access from domestic or institutional servers, and the use of provisional or permanent software licenses. Likewise, with education and distance working, researchers that adapted their workplace in their houses were forced to respond to the same challenges faced at home, for example, low connectivity, the number of computer equipment, or sharing space with family members. The abrupt slowdown of research activities led to the search for new strategies; nevertheless, despite the technological innovations developed during the past decades, there are few tools to carry out scientific work remotely. In this scenario, databases represent solid support, and in some cases, the sole source, for this task, as will be described in the following sections.

2. Relevance of Databases and Scientific Repositories: The Specific Case of Biological Databases

Scientists conduct thorough searches before, during, and after doing research, in order to design experimental strategies, formulate hypotheses, attain significant results and derive robust analysis from them. Access to academic electronic databases, defined as organized collections of related data that allow retrieval of scientific information, has a crucial role in generating further knowledge. This is a common practice whether an extraordinary situation such as the one lived throughout 2020 and 2021, is taking place or not. However, as stated above, a major fraction of the scientific community was compelled to change its paradigm for doing research, relying on the experiments so far achieved or on digital information repositories. Internet or online research has played a fast-developing and highly transformative role in how communication, information, and networking technologies influence the conduction of studies and projects [5]. During the COVID-19 pandemic, this form of doing research or keeping up with it represented the only alternative for many scientists.

At present, virtually all research projects, within the life science fields, rely to some extent on the use of computers and specialized software, which have become essential parts of the toolkit for the biological area [6]. New technologies have allowed the accumulation of biological data at unprecedented rates. High throughput data is derived from nucleic acid and protein sequencing, omics, flow cytometry, gene expression analysis, molecular screening, phylogenetics,

among others, rendering vast collections of life sciences information called biological databases [7]. Biological databases integrate massive amounts of omics-derived information (genomics, transcriptomics, proteomics, secretomics, metabolomics, and glycomics), turning them into indispensable elements for wet and dry lab scientists, in their efforts to share, retrieve, use, analyze, and correlate these data.

Table 1 summarizes the types of biological databases based on several criteria. According to Zou *et al.* [8], different aspects apply in the classification of biological databases, depending on 1) their diverse aims, 2) the several types of data encompassed, or 3) the curatorial levels and methods. Likewise, Bhatt *et al.* [7] mentioned four major criteria used for their classification: 1) data source, 2) data types, 3) database design, and 4) composite databases. Due to their significance, each year, since 1999, the *Nucleic Acids Research* journal reviews new biological databases, eliminates discontinued URLs, and updates an existing compilation of them. This resource can be freely consulted at <https://www.oxfordjournals.org/nar/database/c/>, rendering a current total of 1637 databases for the early 2021 issue [9] and 1645 for the following one [10], divided into 15 categories shown in the following graph (**Figure 1**). The higher number of databases are encompassed in fields focused on genomes, followed by data regarding sequences (nucleotides, RNA, and proteins). A smaller number of

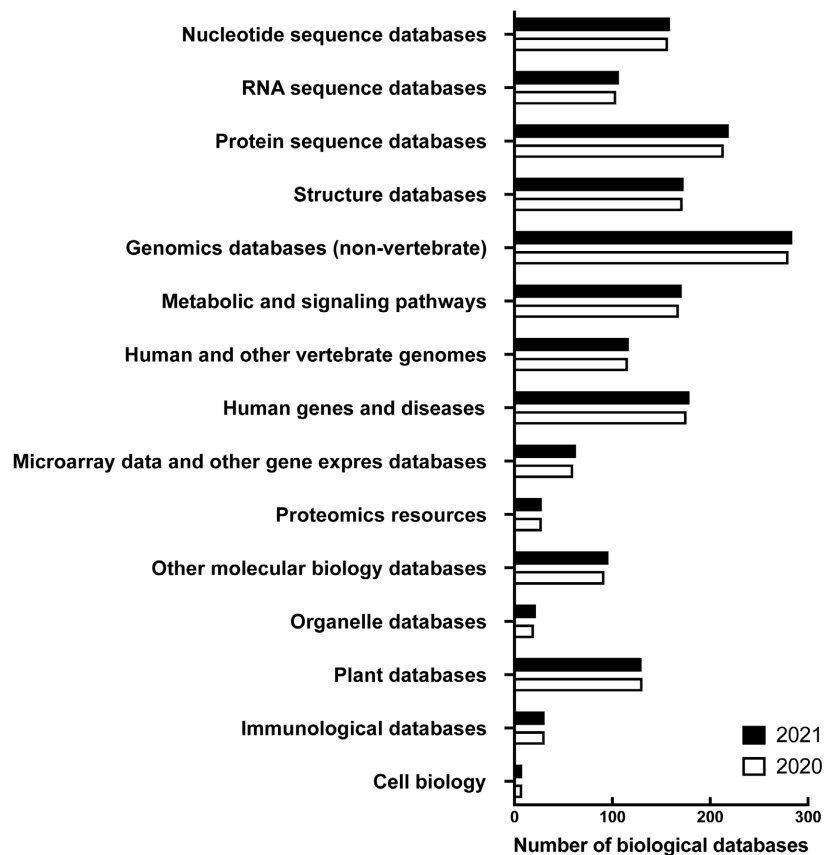


Figure 1. Number of databases based on their biological type according to Rigden and Fernández [9] [10].

Table 1. Types of biological databases based on different aspects.

Criterion according to	
Bhatt <i>et al.</i> [7]	Zou <i>et al.</i> [8]
Source of data:	Purpose of data coverage:
Primary (information derived from experiments)	Comprehensive (different types of data from numerous species)
Secondary (databases derived from primary)	Specialized (specific types of data from specific organism)
Type of data:	Type of data covered:
Sequence data	DNA
Structure	RNA
Literature	Protein
Pathway	Expression
Chemical	Pathway
Enzyme	Disease
Disease	Nomenclature
Domain	Literature
Structural classification of protein	Standard and oncology
Genome	
Database design:	Curation level:
Object-oriented (information is characterized in the form of objects)	Primary (raw data as archival repository)
Relational (comprehensive tables of data)	Secondary or derivative (curated data as added value)
Composite databases:	Curation method:
Combination of primary databases sources, making the query more efficient	Expert-curated Community curated (collectively and collaboratively)

databases (less than 50) is observed for categories within topics such as cell biology, immunology, organelles and proteomics resources, pointing out new areas for research opportunities.

3. Research Papers Published before and during COVID-19 Times: What Do Literature Databases Show Us?

Editors from the *eLife* scientific journal showed their solidarity towards the scientific community by modifying the guidelines for the acceptance of the articles, such as reducing the number of additional experiments, extending authors' revision times, and other strategies to facilitate the publication of research under the "new reality" [11]. A common practice observed during the sanitary

emergency was the request for publishing preprints in order to accelerate information availability and data sharing. Open access repositories such as bioRxiv and medRxiv, respectively focused on life and health sciences, attracted many authors in their aim to communicate research findings, especially dealing with SARS-CoV-2. Nevertheless, as questioned by Chirico *et al.* [12], this raises concerns, mainly about the publishing quality and validity.

The aim of our study centered on analyzing the impact of lockdown, *i.e.*, research centers, caused by SARS-CoV-2, on the production of scientific material. Therefore, to find out a possible effect of the COVID-19 pandemic on annual publications, we conducted a thorough search on the amount of peer-reviewed material per year, from 2015 to 2020. For this task, three bibliographic databases with their respective search engines were employed: 1) PubMed Central (<https://www.ncbi.nlm.nih.gov/pmc/>), 2) SciVerse Scopus (<https://www.elsevier.com/es-mx/solutions/scopus>) and 3) Pro-Quest (<https://www.proquest.com/>).

First, we looked into the overall quantity of published papers, followed by a more specific analysis based on categorizing the research according to its scientific field. In order to validate the findings from the overall published scientific literature as well as the papers categorized by topic, results derived from our search were subjected to statistical validation. For this task, analysis of variance (ANOVA), followed by a post hoc test (Tukey) only when means were considered significantly different, was performed using R software [13]. A notable difference in ciphers was observed between the scores rendered by Scopus and PubMed, compared to Pro-Quest, probably accountable to their correspondent search algorithm. Therefore, the raw data were normalized by calculating the annual increment rate, as the ratio between the publications from the current year compared to the papers from the prior. In addition, with this modification, it was possible to assess the normality of the data, by a Shapiro Wilk test, in order to use ANOVA and Tukey statistical tests. The shift was expressed as n -fold, employing the following relation: $TN_{year_n}/TN_{year_{n-1}}$, in which, TN_{year_n} represents the number of articles for the year of study, while $TN_{year_{n-1}}$ is the peer-reviewed material from the previous one. The following example is given in order to illustrate the latter: average hits for overall published material corresponding to the years 2019 and 2020 were in round numbers, 1,836,000 and 2,089,000, respectively. Hence, by dividing TN_{2020} by TN_{2019} , a 1.14-fold was calculated.

Continuous growth in the amount of scientific literature has been registered during the past years (**Figure 2(a)**); nevertheless, a linear tendency was only observed from 2015 to 2018. The year corresponding to the worldwide spread of SARS-CoV-2 (2020), exhibited a notable difference with a 14% annual increment in peer-reviewed material. When our analysis was focused on the specific scientific field that the literature belonged to (**Figure 2(b)**), the two major topics that caught our attention during the past five years (2015-2020) have centered

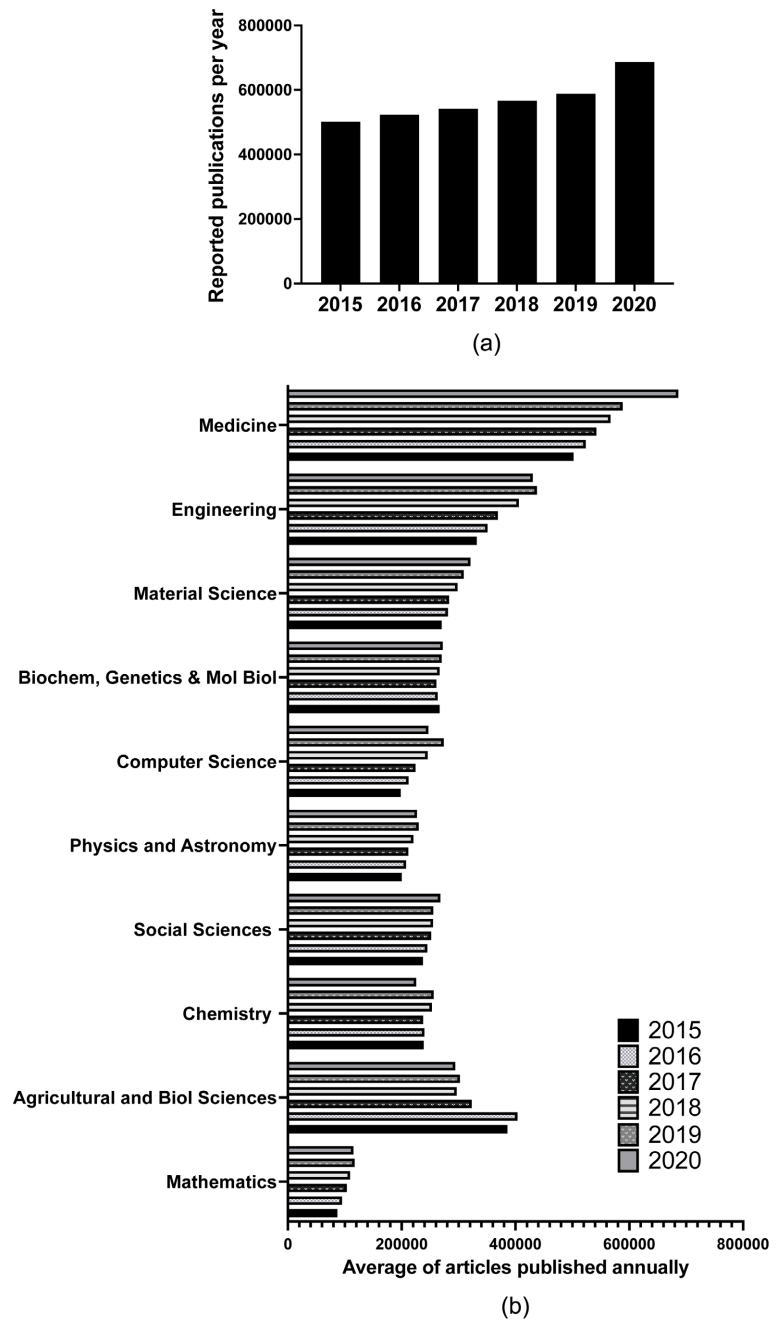


Figure 2. Bars represent the average of data retrieved from search engines PubMed, Scopus, and Pro-Quest. (a) Number of overall published material during the past years. (b) Number of publications corresponding to the years 2015 to 2020, categorized by the most frequent scientific topics.

on areas concerning medicine and engineering. Although there has been an expansion in publications, with respect to years before the pandemics, except for agricultural and biological sciences, we confirmed that the real increment in 2020 was clearly observed in the medical field. Areas concerning social sciences also exhibited a higher growth in the quantity of published material during 2020, compared to previous years, while material sciences continued its positive ten-

agency. On the other hand, research dealing with engineering, physics and astronomy, computer science, chemistry, and mathematics exhibited, as well, continuous growth in the number of published papers until 2019. In the past year, all these areas suffered a negative impact of varying degrees. Topics related to biochemistry, genetics, and molecular biology did not show a clear tendency, apparently remaining unchanged during the years studied. Finally, as mentioned above, papers focused on agricultural and biological sciences contrasted with the other disciplines. While 2016 marked a notable interest in this area, after this year, a pronounced decrease was observed in the number of publications regarding this matter.

ANOVA pointed out that no difference in annual means could be attributed to published material belonging to specific research areas, except for the case of medicine. Therefore, only significant results derived from the statistical analysis will be subsequently described, *i.e.*, peer-reviewed literature not restricted by topic and the medical field. However, all data derived from our search can be accessed upon request. An average shift value, ranging from 1.01 to 1.05-fold, was calculated for the years before the global spread of SARS-CoV-2 (2015-2019) for all publication types. Nevertheless, this finding notably contrasted with the 1.14-fold increase in the number of papers between 2019 and 2020 (**Figure 3(a)**). The research focused on medicine also showed a marked increment in published material for the same time period (2020/2019), exhibiting a mean shift of 1.16, compared to prior years (**Figure 3(b)**). Based on the post hoc statistical tests, it was concluded that the sanitary emergency lived throughout 2020, the year

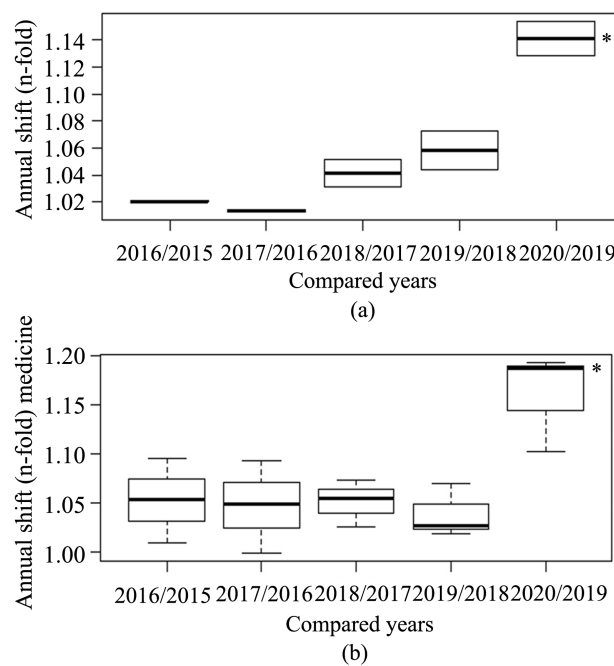


Figure 3. Box and whiskers plots showing annual shifts calculated as the ratio between the number of publications from the current year compared to the papers from the prior. (a) Analysis focused on overall published material. (b) Fold analysis for research papers belonging to the medical field.

marked by the COVID-19 pandemics, might be responsible for this outcome, *i.e.*, the tendency exhibited in the number of publications via peer-reviewing.

4. The Role of COVID-19 in the Increase of Published Material

Due to its impact on humankind, since its appearance, there has been a rapid interest in COVID-19 and the causative agent of this illness (SARS-CoV-2). A clear example of this was illustrated by the emergence of biological databases, such as LitCovid, which reached more than 85,000 publications on the subject by December the 31st, 2020 [14]. In the latest *Nucleic Acids Research* review issue [10], seven new databases inspired by the COVID-19 pandemic were added: COVID19db, Ensemble COVID-19 resource, ESC, SCoV2-MD, SCoVid, T-cell COVID-19 Atlas, and VarEPS. For the specific case of Elsevier's journals, submissions reached 58% more articles in 2020, compared to the year before. The increase focused on health and medicine titles as pointed out by Else [15], and our former statistical analysis. *The Lancet* medical journal reported on receiving an increase of 185% for publication requests in June 2020, with respect to the same period in the previous year; half of them centered on COVID-19. However, of more than 2000 submissions, only 5 accepted papers were cataloged as original research works [16]. Therefore, the next aim of our study centered on knowing how much attention was drawn to topics regarding SARS-CoV-2 from scientists, reviewers, and editors.

Using the same three search engines (PubMed Central, SciVerse Scopus, and Pro-Quest), we looked into the total number of articles released between January 1 and December 31, 2020, that contained the terms "COVID-19", "SARS-CoV-2", one or the other ("COVID-19" OR "SARS-CoV-2"), and both ("COVID-19" AND "SARS-CoV-2") (Figure 4). "COVID-19" was more popular than "SARS-CoV-2" as reflected by its usage order of magnitude higher, compared to the virus. In accordance with LitCovid

(<https://www.ncbi.nlm.nih.gov/research/coronavirus/>), regarding the amount of material focusing on this topic, it was found that no more than 4% of the published research was related to it. Thus, despite the compel for remote work brought by the quarantine, this outcome highlighted the great effort by researchers to publish their work related to different knowledge fields. The latter finding was reinforced by the 1.14-fold increase in publications calculated in this study when comparing the years 2019 and 2020. Therefore, in spite of the attention drawn by COVID-19 and SARS-CoV-2, the growth in the percentage of articles in 2020 (14%) cannot be entirely attributed to them. On the other hand, when the analysis was restricted to literature belonging to the medical field, as expected, the percentage of papers containing "COVID-19" or both ("COVID-19"/"SARS-CoV-2") was higher (8%). Therefore, unlike for the case of overall research material, specific focus on medicine reflects a marked influence, accountable for 50% on the positive shift (1.16-fold), shown in Figure 3(b), of

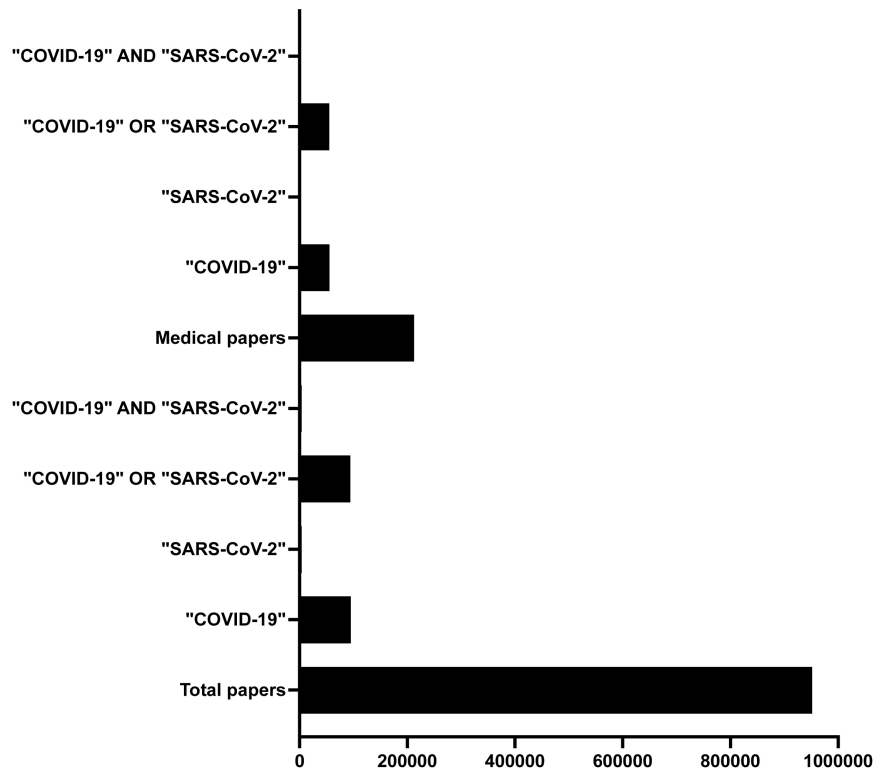


Figure 4. Overall and medical related peer reviewed research papers encompassing terms related to COVID-19 and/or SARS-CoV-2.

reviewed and published literature related to the illness and the causative viral agent of it.

5. Databases as Allies during COVID-19 Times

Without any doubt, the year 2020 represented a major breakpoint in life as we knew it, with COVID-19 as the major driving force. Research and literature derived from it, as products of human activity, could not escape the influence of the pandemic, as pointed out by this work and the publications from the authors here referred to. Despite its effects, the quarantine and the closure of many research centers, exhibited no negative effects in the number of publications related to overall topics. Possible explanations could be that previously held data were “recovered” during COVID-19 times and subsequently released, or alternatively, other data-gathering strategies might have been utilized. This situation took place in our research center (*Instituto de Química*, U.N.A.M.): Dr. Jorge Peón, the former head of our institution, let us know of the increase in published material during 2020 even though laboratories were not fully active (personal communication). Perhaps the generation of data relied on activities that did not require experiments in a wet lab or rather, depended on the acquisition of information stored in databases. Thus, the data deposited (with continuous updating) in these repositories serve as a reference for in-depth analysis of topics of scientific interest in general. As a reference to this, one of the authors of this

manuscript is currently conducting a phylogenetic study, working from home and collecting gene sequences stored and freely shared in different databases.

Clearly, in the medical field significant attention was directed to topics related to COVID-19. Outstanding efforts from healthcare personnel worldwide have been crucial for fighting against SARS-CoV-2 and attending patients infected with the aforementioned virus. At the same time, staff involved in research on COVID-19 from different centers (hospitals, clinics, universities, pharmaceutical industries, government, among others), focused primarily on one goal: knowing more about the causative agent of the disease. Chinese and Australian scientists combined efforts to determine the genome of the new virus: metagenomic samples of bronchoalveolar lavage of patients allowed the identification of a new RNA virus from the Coronaviridae family, rendering the first genome sequence by January 2020 [17]. These authors deposited the first sequence of the “Wuhan seafood market pneumonia virus isolate” in GenBank (accession number MN908947); nevertheless, in March 2020, this genomic sequence was replaced by MN908947.3 and identified since then, as “Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)”.

Immediate data sharing of the complete genome of SARS-CoV-2 and published material derived from research focused on it, opened routes for its diagnosis, as well as in determining possible targets for the development of drugs and vaccines. Within dry labs, intensive research using bioinformatics tools together with molecular modeling programs, became extraordinary allies in the battle against COVID-19, by generating scientific repositories of an incredible quantity of knowledge about the characteristics of the virus, essentially its pathogenesis, spread, mechanism of cellular infection, and evolution. Databases can be useful tools to analyze topics such as proteins, genes, genomes, metabolism, and mutations, without the requirement to work in a wet lab. Based on the similarity in the genome of SARS-CoV and SARS-CoV-2, the following proteins: Spike (S), Envelope (E), Membrane (M), and Nucleocapsid (N) were proposed for the design of rapid tests (based on antigens and/or antibodies), although not as specific as those performed using molecular techniques (RT-PCR and RT-qPCR) [18].

6. The Year 2021: Vaccines, Reopening, and Variants

Despite the numerous efforts on many battlefields against COVID-19, this respiratory disease remains the world’s leading health problem. Nucleocapsid protein (N) and Spike protein (E) are considered to be highly immunogenic and have been studied using bioinformatics to find potentially immunodominant regions [19]. The year 2021 has been marked by the advent of hope in the form of vaccines, but at the same time, of uncertainty brought about by more contagious and/or severe variants, as a result of viral evolution. The efficiency of vaccines against different variants has been a topic of interest in recent months, especially after November 2021 with the emergence of Omicron. Until January 2022, 334

vaccines for COVID-19 were reported, of which only 17 are currently being offered to the population [2]. By January 2022, 61.1% of the world population has received at least one dose [20], and since mid-2021, labs have been arduously working on variant-specific boosters.

After contrasting vaccination campaigns around the globe, reopening became a hot topic, mainly, due to its economic and social impacts. Businesses were compelled to implement strategies that guaranteed the health and safety of their customers, primarily focused on limiting full capacities. During the past year, a gradual and partial return to everyday activities represented a worldwide issue. Research centers were no exceptions: their staff was coming back to work but under restricted operating conditions.

In view of this scenario, once again we searched for the scientific production achieved during the period comprehended between January 1 and December 31, 2021, as described above. To our surprise, the Pro-Quest engine yielded results that contrasted greatly with our former findings, which we presumably account for a modification in its searching algorithm. Therefore, for the analysis corresponding to the year 2021, it was decided to employ hits only derived from PubMed and Scopus and to compare them to the data from 2020, excluding Pro-Quest ciphers, as well. Statistical validation was performed by a Student's t-test with the aid of R software [13]. After assessing the normality of the data, the means of the different years were compared in order to find a significant difference. A two-sided test was used first, focused on validating the difference between media as the alternative hypothesis. Next, only significant cases were further subjected to a one-sided test in order to know if the mean of scientific production from 2019 to 2020 was greater than the attained from 2020 to 2021.

A very slight increment was observed for 2021 in the total number of publications (**Figure 5(a)**); nevertheless, when the search was directed to specific scientific fields, only medicine and areas regarding biochemistry, genetics, and molecular biology, exhibited appreciable differences between the studied years. Published material in the medical field showed a 1.11-fold shift from 2020 to 2021, while topics belonging to biochemistry, genetics, and molecular biology decreased by a 0.6 factor (**Figure 5(b)**).

Differences were found only for an overall number of publications and the medical-related field, with averages significantly greater for material published during 2020 (**Figure 6**). Despite the result observed for papers belonging to biochemistry, genetics, and molecular biology, the t-test showed that the differences were not statistically significant, which might be accountable to the high values calculated in their standard deviations. Once again, this analysis reinforced our previous findings towards pointing out the role of COVID-19 in the generation of information.

Nowadays, the tracking down of SARS-CoV-2 mutations, as they arise, is feasible due to genomic surveillance. Scientists are monitoring genetic changes over time, on a scale never seen before; in the last year, more than 360,000 SARS-CoV-2 genomes have been sequenced and stored in GISAID [21]. Many

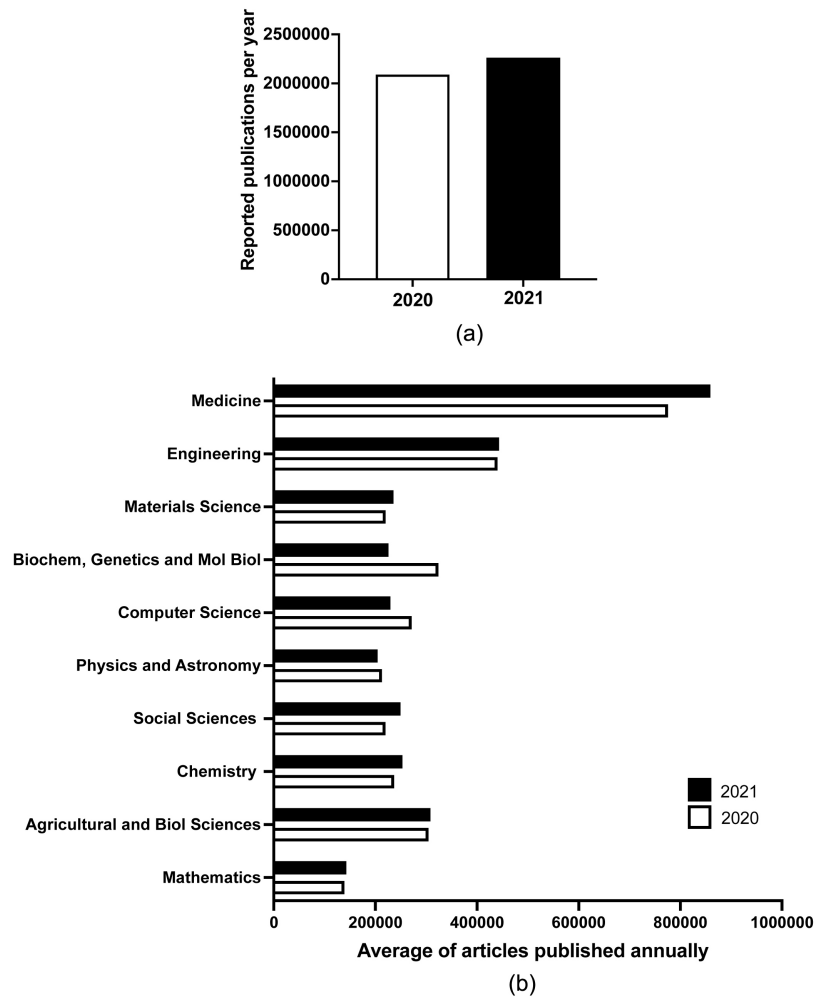


Figure 5. Bars represent the average of data retrieved from search engines PubMed and Scopus. (a) Number of overall published material during the past couple of years. (b) Number of publications corresponding to the years 2020 to 2021, categorized by the most frequent scientific topics.

variants have been detected, but only those that have represented a public health problem have been studied. Therefore, only the following variants related to outbreaks and high viral loads will be described. In September 2020, a SARS-CoV-2 variant was detected in England and was called B117 (alfa) [22]. Moreover, in October, the variant 501Y.V2 (beta) was detected in South Africa [23]. A little later, in December, the P.1 (gamma) strain emerged in Brazil [24], and the B.1.617.2 (Delta) led to an increase in the number of cases in India [25]. Finally, until June 2021, Deng *et al.* detected a new variant in California, which was called epsilon (B.1.427/429). In a recent study, it was found that this variant was able to cause more severe disease and higher viral loads [26]. Currently, the variant Omicron (B.1.1.529) has spread worldwide, impacting considerably due to its severe contagiousness and mutations on the Spike (S) protein receptor-binding domain (RBD); therefore, further investigation on the effectiveness of vaccines against this variant is needed [27].

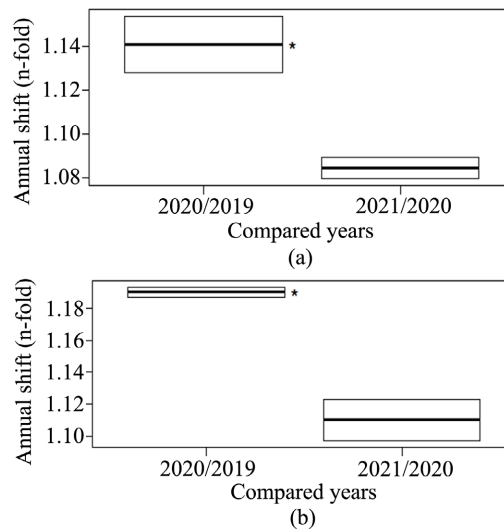


Figure 6. Box and whiskers plots showing annual shifts calculated as the ratio between the number of publications from the current year compared to the papers from the prior. (a) Analysis focused on overall published material. (b) Fold analysis for research papers belonging to the medical field.

Sequence analysis of SARS-Co-V-2 variants has demonstrated the emergence of mutation in the Spike protein, which represents the main target for vaccine generation. At present, a key issue relies on whether COVID-19 vaccines will be able to protect against infectious diseases caused by new and upcoming variants. The efficiency of vaccines against known variants has been demonstrated [28]; still, the development of immunotherapies effective to combat these variants is urgently needed [29]. Certainly, the presence of the new strains has been an arduous challenge to face; nevertheless, the utility of databases has allowed multidisciplinary work (doctors, chemists, biologists, mathematicians, engineers, computer scientists, etc.) to face this adversity. As pointed out by Zhou and Wang [30], research focused on the design of new vaccines requires hard endeavors within laboratories (wet and dry) and clinical trials, which in turn rely on the genomic monitoring of viruses for variant detection.

7. A Deeper Look into Published Material: Subtopics Analysis from the Specific Knowledge Fields Formerly Describes

Years 2019, 2020, and 2021 were selected for this section, and our search was based on the subcategories included in Scopus for each of the scientific fields previously analyzed. As described above, data were collected from PubMed and Scopus, limiting the query to the specific subtopic. The comparison between years was expressed in n-fold, as calculated before, and the statistical validation was accomplished by a Student's t-test using R software [13]. Despite the number of subareas analyzed which can be consulted in the supplementary material, significant differences were extremely limited and only related to the medical field (Figure 7). Within this scientific area, “endocrinology” and “virology” ex-

hibited a considerable shift during 2020 and decreased their number of publications for the next year. Other subtopics belonging to knowledge fields different from medicine were not statistically significant.

As stated before, the increase in medical publications was expected due to the health situation lived during the past two years. Except for the case of virology, in which increment is accountable for the sanitary situation due to COVID-19, the differences found within other subfields of knowledge cannot be explained by our analysis. In this matter, we reviewed hits for 2019 as the year that started the health emergency of COVID-19 and the subsequent years of its worldwide spreading (2020 and 2021) to determine the variation of topics of medical interest. The search was performed on PubMed and Scopus using the term “Medicine” and then separated by the publication year. The subtopics found in these first queries were grouped in the three databases using the following commands: (TITLE-ABS-KEY (medicine)) AND (subtopic) AND (LIMIT-TO (PUBYEAR, year)). Results derived from our quest are presented in **Figure 8**.

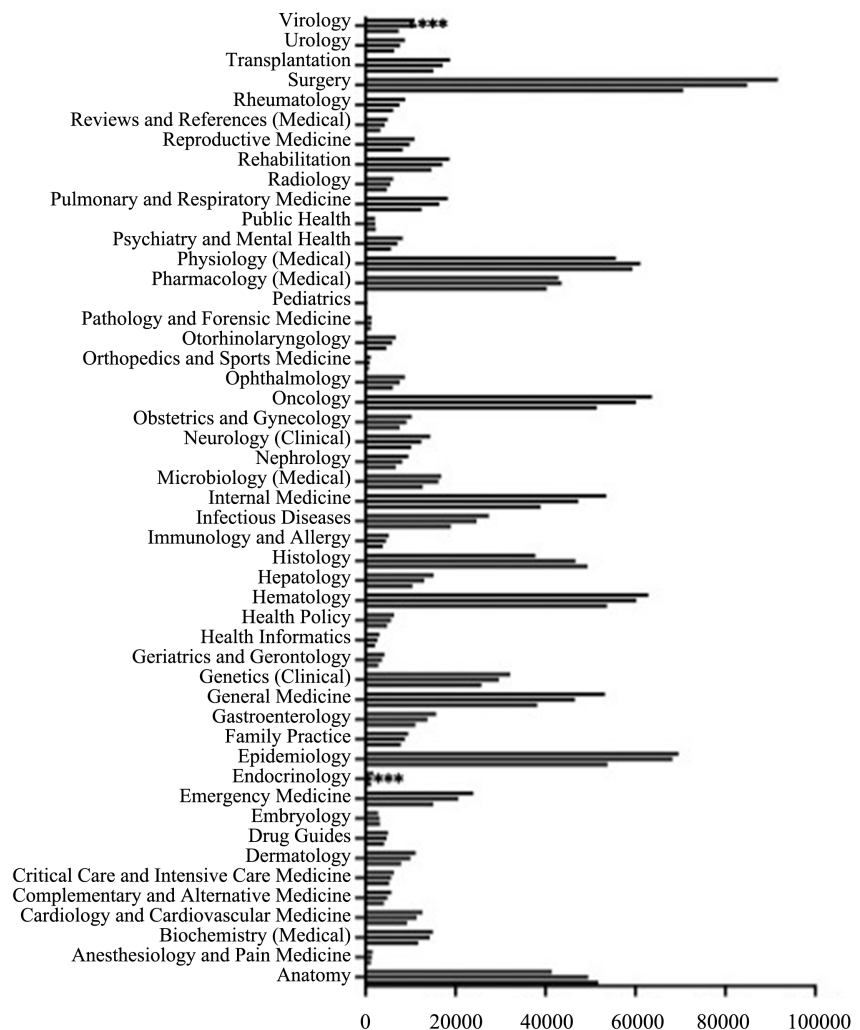


Figure 7. Bars represent the average of data retrieved from search engines PubMed and Scopus, corresponding to subtopics belonging to the medical area.

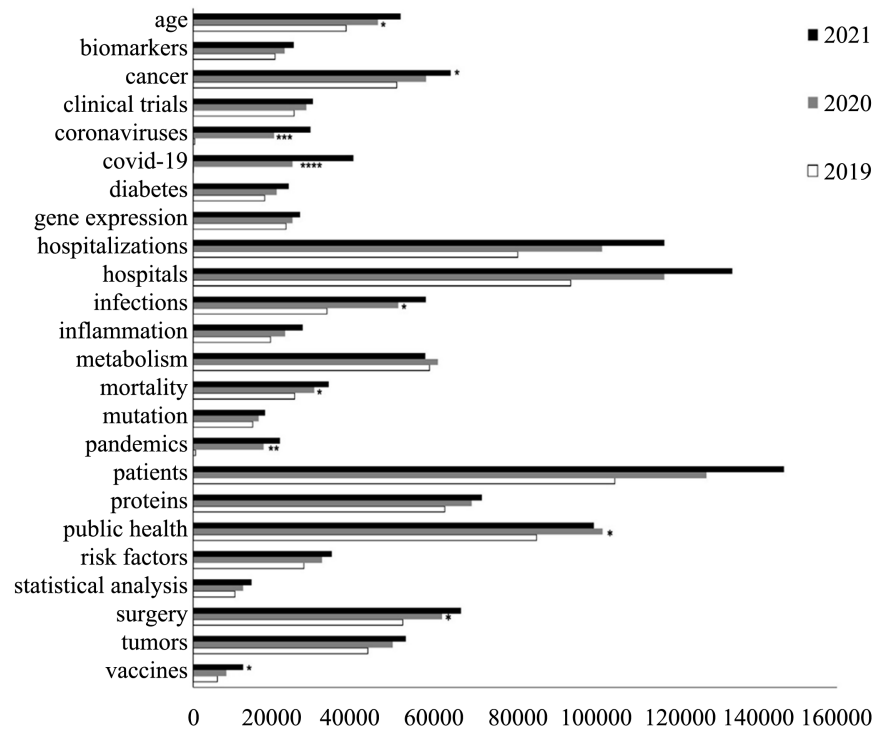


Figure 8. Bars represent the average of data retrieved from search engines PubMed and Scopus, corresponding to keywords belonging to the medical area.

Our findings showed that the following keywords: “age”, “infections”, “mortality”, “public health”, and “risk factors” exhibited a mild increase during 2020. On the other hand, keywords such as “coronaviruses”, “COVID-19”, and “pandemic” increased significantly during this same year, confirming again that much of the published material focused on SARS-CoV-2. Other important health issues such as “cancer” recovered importance after the year marked by the global pandemic, while “vaccines” attracted a lot of attention due to the health, economic, political, and social pressures lived throughout 2021. Once more, these results confirmed the effect of COVID-19 on the research carried out through 2020, and at the same time, for subsequent years, the diversification on other topics.

8. Final Comments: Information Sources, Availability, Data Sharing, and Joint Efforts

With the spread of the internet and information technologies, humanity is currently living in an era referred to as the “information age”. This time period is marked by complexity and change, challenging our capacity to create, store, retrieve, manage, and communicate data. In this scenario, the term “information literacy” is defined as “the set of integrated abilities encompassing the reflective discovery of information, the understanding of how information is produced and valued, and the use of information in creating new knowledge and participating ethically in communities of learning” [31]. According to UNESCO [32],

information literacy enables people to interpret and make informed judgments as users of information sources, as well as to become data producers. In a digital era, it is essential to have skills that allow the use of information technologies (computer-based data systems), in order to perform these tasks. Even though the concept has traditionally been associated with an academic environment, information literacy has a broader spectrum in an individual and a social context, reflected in good decision-making in education, health, and/or politics, among other areas.

The need to be informed is crucial in our modern society, a phenomenon observed by the increasing number of people around the world looking for data. In this context, the use of computers, and in past years, smaller gadgets such as smartphones or more recent smartwatches with an internet connection and a search engine have been fundamental. Parallel to this necessity lies the speed of acquiring information, so each time, user-friendly devices with larger and more powerful capacities are designed. Therefore, access to data has been facilitated and enhanced by technology in an inconceivable manner, focused on a wide variety of areas, for example, business, education, entertainment, fitness, research, shopping, and even, traveling.

In times of the information age, social media (blogs, podcasts, social networking, and streaming) exert a profound impact on people's perceptions. Unfortunately, sometimes this way of being informed might lack trustworthiness and be misleading. Regarding the present coronavirus sanitary emergency, "infodemic" arose, reflected by a daily exponential growth of information on COVID-19. A serious issue about this phenomenon consists of the dissemination of fake news, which in turn, influences population awareness of the virus and government response against it. Whether verified or not, data availability from traditional (radio, TV, and printed information) or social media, negatively influenced people's minds and their consciousness of SARS-CoV-2. Consequently, the rapid spread of news concerning COVID-19 infections and deaths, caused anxiety, panic, stigma, mistrust, and rumormongering among the general population [33]. In order to counterattack this global threat WHO, UN, UNICEF, UNESCO, UNAIDS, ITU, UN Global Pulse, and IFRC [34] published a joint statement, calling on the Member States and other stakeholders (media, social media platforms, researchers, technologists, civil society leaders, and influencers) to promote the spread of accurate data and combat against infodemics. In addition, social media platforms, such as Facebook, Reddit, Twitter, and YouTube, were compelled to adopt countermeasures versus mis- and disinformation regarding COVID-19, highlighting: 1) flagging, 2) fast checking by third parties, 3) less visibility or deletion of bots or offenders accounts, and 4) adding visible links or redirecting to reliable information [35]. In this regard, Tornberg *et al.* [36] assessed the dissemination of COVID-19-related data, *i.e.*, articles, across social media by alternative metrics. Still, there is more to overcome, as the vaccination phase against the COVID-19 pandemic has faced contradictory opinions with their respective misleading judgements.

Efficient generation of scientific information from the research and rapid sharing of data represented, and still does, an excellent alternative for attaining information literacy on this topic, labor supported by the existence of databases. Many international academic journals directed efforts for quick review and publication of scientific papers on COVID-19, and various websites have been created to assemble and disclose articles on this disease, thus, reducing public panic and providing real-time guidance [33]. The attention drawn by this disease was assessed via Scopus, analyzing the number of citations regarding this topic, as well as other impact indexes, such as “Field-Weighted Impact Factor” and “Field Weighted View Impact” (Figure 9). It was observed that 2020 marked a clear tendency in the interests of the scientific community, with COVID-19 as a hot topic. Figure 9 shows that the number of COVID-19 citations concentrated during 2020, the year in which the pandemic had a larger impact on global restrictions. In 2021 and 2022, scientists were able to work more relaxed, although the effect of SARS-CoV-2 continues to play a significant role in global

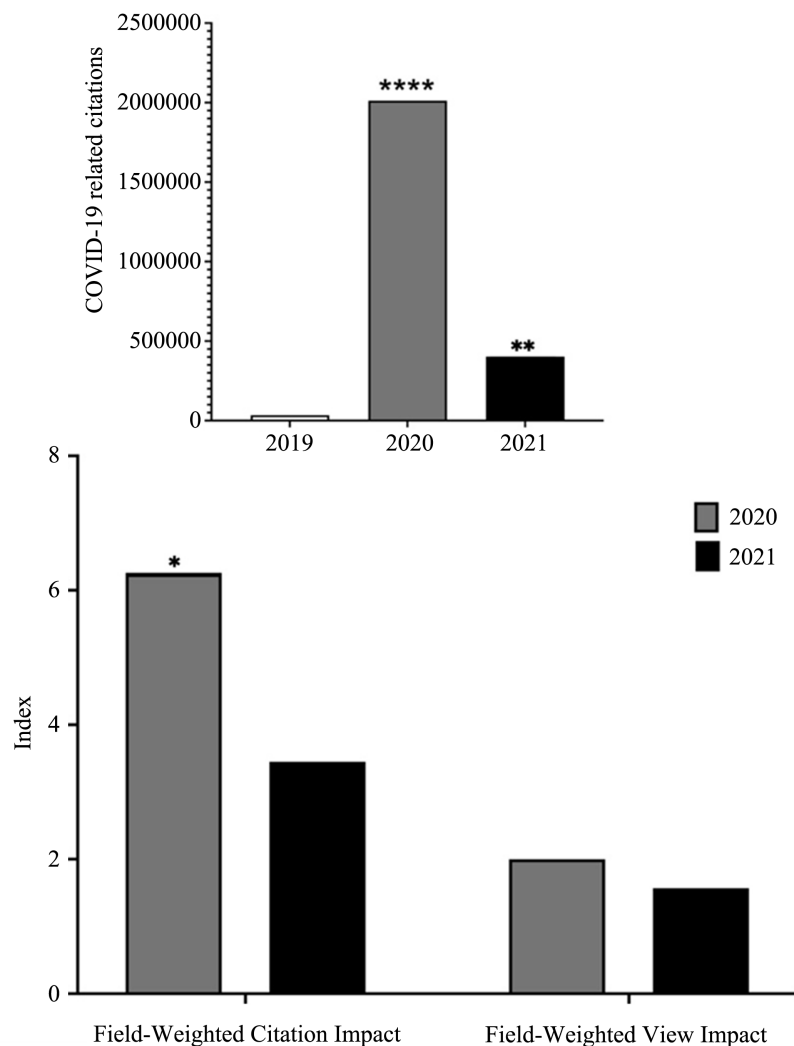


Figure 9. Published material impact regarding COVID-19 information.

research. Likewise, this same graph, demonstrates the preference in citations and reads towards this subject, with a respective n-fold for 2020 of 6.3 and 2.0, and for 2021 of 3.5 and 1.6. Both indexes reflect an increase in requests for COVID-19 information focused on data production. At the same time, the global demand for clinical data drove a joint collaboration from an open community (clinicians, researchers, and data scientists) to create the National COVID Cohort Collaborative (N3C) (<https://covid.cd2h.org/>) and improve public health.

Undoubtedly, COVID-19 revolutionized our lifestyles and continues to do so, as new regulations for fighting against its spreading are constantly being implemented. For example, negative COVID tests are required before letting travelers enter some countries, or in some cases, people are being asked to show vaccination certificates in order to get access to public places. In the case of Mexico, sanitary filters have been installed before accessing closed places such as cafeterias and restaurants, health centers and hospitals, gyms, shopping centers, supermarkets, schools and education centers, among others. Even though control measures tend to relax, wearing the face mask indoors continues to be mandatory. Therefore, human daily activities were and still are strongly impacted by SARS-CoV-2, as pointed out by Popouet's study in Africa focused in areas such as economy, food security and household income, financial balances, and unemployment [37].

As pointed out above, during the lockdown period, research centers and universities were compelled to close, and work had to be done remotely. Different strategies were adopted by the scientific community in order to prevent a pause in the generation of knowledge and sharing of data. In addition to the fight against COVID-19, another important issue was represented by infodemics, so another battlefield took place within the informatics area. In this work, we turned our attention to databases, as repositories of reliable information, on knowing the effect of pandemics on the amount of published material during COVID-19 times. It was found that, in spite of the global closure, a considerable increase in the number of publications from different scientific fields took place, with medicine being the most notable and significant one, especially, with topics regarding "COVID-19" and "SARS-CoV-2". The latter was reinforced by the subtopic analysis, in which notorious differences were observed in comparison to other knowledge areas. Even though other knowledge areas exhibited no significance in the statistical tests, the year 2020 might be considered a very productive period, suggesting the usefulness of information repositories, while laboratories were closed, for the generation of publications not directly related to the pandemic, as demonstrated by our review.

Since the beginning of the pandemics, general interests have focused on its effects on various fields. Publishing during COVID-19 times represented quite a challenge and the center of several research papers. An analogous study to this work was conducted by Sepúlveda-Vildósola *et al.* [38] during the first months of the worldwide closure (March-April, 2020), in which they addressed scientific material related to COVID-19, highlighting: 1) the number of new articles pub-

lished daily focused on specific fields of knowledge, 2) the journals with most papers dealing with the pandemics, 3) the languages used for writing those documents, 4) the countries of origin, and 5) the type of publication. Nevertheless, despite certain similarities, our review represents the first report in which the impact of COVID-19 was studied at a much broader time scale, addressing overall publications and specifically categorizing them by the area of knowledge they belonged to. This time interval included five years before the pandemics (2015-2019), the year marked by the spread of the disease around the globe (2020), and the subsequent one, characterized by the advent of hope in the form of vaccines. Certainly, due to the sanitary emergency, it could be expected that the medical field drew almost all the attention of the scientific and non-scientific community; with our study, this hypothesis was confirmed and statistically validated.

Databases have made it possible to collect and analyze data from more than 207 countries on 6 continents [20]. In addition, the international exchange has improved the monitoring, prevention and control of the spread of the virus. In 2020, the year marked by the beginning of the pandemic, data generation and data sharing were important allies in the fight against COVID-19, and during 2021, their support has continued to be crucial. To illustrate this, we conducted a search, using the same engines described before, focused on research papers containing the term “COVID-19”. Surprisingly, for 2021, the same amount of articles was published in comparison to the previous year. When the search was narrowed down by the inclusion of the term “vaccine”, a 3-fold increase in the number of publications was observed in 2021, compared to 2020. Nowadays, it is amazing how technology has contributed to speeding up, at unconceivable rates, the process of information generation, data sharing, and its accessibility. In COVID-19 times, an outstanding example is represented by the research and experimentation behind vaccine production; considering the traditional time for their development, 2020-2021 could be considered as a scientific record. Finally, new challenges concerning SARS-CoV-2 are arising, such as the presence of novel strains; nevertheless, the utility of databases has allowed inter- and multi-disciplinary collaboration in order to face it.

Acknowledgements

We are deeply grateful to Yavonne Zhang and Alice Yao for their guidance and support during the peer-review and publication processes of this work.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] WHO (World Health Organization) (2021) Coronavirus Disease (COVID-19). https://www.who.int/health-topics/coronavirus#tab=tab_1

- [2] WHO (World Health Organization) (2021) Coronavirus (COVID-19) Dashboard. <https://covid19.who.int>
- [3] WBDG (Whole Building Design Guide—National Institute of Building Sciences) (2019) Laboratory: Dry. <https://www.wbdg.org/space-types/laboratory-dry>
- [4] Korbel, J.O. and Stegle, O. (2020) Effects of the COVID-19 Pandemic on Life Scientists. *Genome Biology*, **21**, 113. <https://doi.org/10.1186/s13059-020-02031-1>
- [5] Tsatsou, P. (2016) Digital Technologies in the Research Process: Lessons from the Digital Research Community in the UK. *Computers in Human Behavior*, **61**, 597-608. <https://doi.org/10.1016/j.chb.2016.03.053>
- [6] Gauthier, J., Vincent, A.T., Charette, S.J. and Derome, N. (2019) A Brief History of Bioinformatics. *Briefings in Bioinformatics*, **20**, 1981-1996. <https://doi.org/10.1093/bib/bby063>
- [7] Bhatt, V.D., Patel, M. and Joshi, C.G. (2018) An Insight of Biological Databases Used in Bioinformatics. In: Wadhwa, G., Shanmughavel, P., Singh, A. and Bellare, J., Eds., *Current Trends in Bioinformatics: An Insight*, Springer, Singapore, 3-25. https://doi.org/10.1007/978-981-10-7483-7_1
- [8] Zou, D., Ma, L., Yu, J. and Zhang, Z. (2015) Biological Databases for Human Research. *Genomics Proteomics Bioinformatics*, **13**, 55-63. <https://doi.org/10.1016/j.gpb.2015.01.006>
- [9] Rigden, D.J. and Fernández, X.M. (2020) The 27th Annual Nucleic Acids Research Database Issue and Molecular Biology Database Collection. *Nucleic Acids Research*, **48**, D1-D8. <https://doi.org/10.1093/nar/gkz1161>
- [10] Rigden, D.J. and Fernández, X.M. (2022) The 2022 Nucleic Acids Research Database Issue and the Online Molecular Biology Database Collection. *Nucleic Acids Research*, **50**, D1-D10. <https://doi.org/10.1093/nar/gkab1195>
- [11] Eisen, M.B., Akhmanova, A., Behrens, T.E. and Weigel, D. (2020) Publishing in the Time of COVID-19. *Elife*, **9**, e57162. <https://doi.org/10.7554/eLife.57162>
- [12] Chirico, F., Teixeira da Silva, J.A. and Magnavita, N. (2020) Questionable Peer Review in the Publishing Pandemic during the Time of COVID-19: Implications for Policy Makers and Stakeholders. *Croatian Medical Journal*, **61**, 300-301. <https://doi.org/10.3325/cmj.2020.61.300>
- [13] R Core Team (2020) R: A Language and Environment for Statistical Computing. <https://www.r-project.org>
- [14] Chen, Q., Allot, A. and Lu, Z. (2021) LitCovid: An Open Database of COVID-19 Literature. *Nucleic Acids Research*, **49**, D1534-D1540. <https://doi.org/10.1093/nar/gkaa952>
- [15] Else, H. (2020) How a Torrent of COVID Science Changed Research Publishing—In Seven Charts. *Nature*, **588**, 553. <https://doi.org/10.1038/d41586-020-03564-y>
- [16] TLG Health. (2020) Publishing in the Time of COVID-19. *The Lancet Global Health*, **8**, e860. [https://doi.org/10.1016/S2214-109X\(20\)30260-6](https://doi.org/10.1016/S2214-109X(20)30260-6)
- [17] Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H., Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C. and Zhang, Y.Z. (2020) A New Coronavirus Associated with Human Respiratory Disease in China. *Nature*, **579**, 265-269. <https://doi.org/10.1038/s41586-020-2008-3>
- [18] Giovanetti, M., Benedetti, F., Campisi, G., Ciccozzi, A., Fabris, S., Ceccarelli, G., Tambone, V., Caruso, A., Angeletti, S., Zella, D. and Ciccozzi, M. (2021) Evolution

- Patterns of SARS-CoV-2: Snapshot on Its Genome Variants. *Biochemical and Biophysical Research Communications*, **538**, 88-91.
<https://doi.org/10.1016/j.bbrc.2020.10.102>
- [19] Dai, Y., Chen, H., Zhuang, S., Feng, X., Fang, Y., Tang, H., Dai, R., Tang, L., Liu, J., Ma, T. and Zhong, G. (2020) Immunodominant Regions Prediction of Nucleocapsid Protein for SARS-CoV-2 Early Diagnosis: A Bioinformatics and Immunoinformatics Study. *Pathogens and Global Health*, **114**, 463-470.
<https://doi.org/10.1080/20477724.2020.1838190>
- [20] Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., Macdonald, B., Dattani, S. and Roser, M. (2020) Coronavirus Pandemic (COVID-19). <https://ourworldindata.org/coronavirus>
- [21] Cyranoski, D. (2021) Alarming COVID Variants Show Vital Role of Genomic Surveillance. *Nature*, **589**, 337-338. <https://doi.org/10.1038/d41586-021-00065-4>
- [22] COVID-19 Genomics UK (COG-UK) (2020) An Integrated National Scale SARS-CoV-2 Genomic Surveillance Network. *The Lancet Microbe*, **1**, e99-e100.
[https://doi.org/10.1016/S2666-5247\(20\)30054-9](https://doi.org/10.1016/S2666-5247(20)30054-9)
- [23] Duong, D. (2021) What's Important to Know about the New COVID-19 Variants? *Canadian Medical Association Journal*, **193**, E141-E142.
<https://doi.org/10.1503/cmaj.1095915>
- [24] Jogalekar, M.P., Veerabathini, A. and Gangadaran, P. (2021) SARS-CoV-2 Variants: A Double-Edged Sword? *Experimental Biology and Medicine (Maywood)*, **246**, 1721-1726. <https://doi.org/10.1177/153537022111014146>
- [25] Otto, S.P., Day, T., Arino, J., Colijn, C., Dushoff, J., Li, M., Mechai, S., Van Domseelaar, G., Wu, J., Earn, D.J.D. and Ogden, N.H. (2021) The Origins and Potential Future of SARS-CoV-2 Variants of Concern in the Evolving COVID-19 Pandemic. *Current Biology*, **31**, R918-R929.
<https://doi.org/10.1016/j.cub.2021.06.049>
- [26] Carroll, T., Fox, D., van Doremalen, N., Ball, E., Morris, M.K. Sotomayor-Gonzalez, A., Servellita, V., Rustagi, A., Yinda, C.K., Fritts, L., Port, J.R., Ma, Z.M., Holbrook, M., Schulz, J., Blish, C.A., Hanson, C., Chiu, C.Y., Munster, V., Stanley, S. and Miller, C.J. (2021) The B.1.427/1.429 (Epsilon) SARS-CoV-2 Variants Are More Virulent than Ancestral B.1 (614G) in Syrian Hamsters. *PLOS Pathogens*, **18**, e1009914.
<https://doi.org/10.1371/journal.ppat.1009914>
- [27] Araf, Y., Akter, F., Tang, Y., Fatemi, R., Alam, S., Zheng, C. and Hossain, G. (2022) Omicron Variant of SARS-CoV-2: Genomics, Transmissibility, and Responses to Current COVID-19 Vaccines. *Journal of Medical Virology*, **94**, 1825-1832.
<https://doi.org/10.1002/jmv.27588>
- [28] Fontanet, A., Autran, B., Lina, B., Kieny, M.P., Karim, S.S.A. and Sridhar, D. (2021) SARS-CoV-2 Variants and Ending the COVID-19 Pandemic. *The Lancet*, **397**, 952-954. [https://doi.org/10.1016/S0140-6736\(21\)00370-6](https://doi.org/10.1016/S0140-6736(21)00370-6)
- [29] Callaway, E. and Ledford, H. (2021) How to Redesign COVID Vaccines So They Protect against Variants. *Nature*, **590**, 15-16.
<https://doi.org/10.1038/d41586-021-00241-6>
- [30] Zhou, W. and Wang, W. (2021) Fast-Spreading SARS-CoV-2 Variants: Challenges to and New Design Strategies of COVID-19 Vaccines. *Signal Transduction and Targeted Therapy*, **6**, 226. <https://doi.org/10.1038/s41392-021-00644-x>
- [31] ACRL (Association of College and Research Libraries) (2015) Framework for Information Literacy for Higher Education.
http://www.ala.org/acrl/sites/ala.org/acrl/files/content/issues/infolit/Framework_IL

[HE.pdf](#)

- [32] UNESCO (United Nations Educational Scientific and Cultural Organization) (2022) Information Literacy. <https://www.unesco.org/en/communication-information/media-information-literacy>
- [33] Song, P. and Karako, T. (2020) COVID-19: Real-Time Dissemination of Scientific Information to Fight a Public Health Emergency of International Concern. *BioScience Trends*, **14**, 1-2. <https://doi.org/10.5582/bst.2020.01056>
- [34] WHO (World Health Organization) (2020) Managing the COVID-19 Infodemic: Promoting Healthy Behaviours and Mitigating the Harm from Misinformation and Disinformation. <https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation>
- [35] Marin, L. (2020) Three Contextual Dimensions of Information on Social Media: Lessons Learned from the COVID-19 Infodemic. *Ethics and Information Technology*, **23**, 1-8. <https://doi.org/10.1007/s10676-020-09550-2>
- [36] Tornberg, H.N., Moezinia, C., Wei, C., Bernstein, S.A., Wei, C., Al-Beyati, R., Quan, T. and Diemert, D.J. (2021) Assessing the Dissemination of COVID-19 Articles across Social Media with Altmetric and PlumX Metrics: Correlational Study. *Journal of Medical Internet Research*, **23**, e21408. <https://doi.org/10.2196/21408>
- [37] Popouet, S.Y. (2021) COVID-19: Impact and Opportunities in Gabon. *Open Journal of Social Sciences*, **9**, 167-185. <https://doi.org/10.4236/jss.2021.98012>
- [38] Sepúlveda-Vildósola, A.C., Mejía-Aranguré, J.M., Barrera-Cruz, C., Fuentes-Morales, N.A. and Rodríguez-Zeron, C. (2020) Scientific Publications during the COVID-19 Pandemic. *Archives of Medical Research*, **51**, 349-354. <https://doi.org/10.1016/j.arcmed.2020.05.019>