

## Scientific information in times of pandemic analysis of indexing time of journal articles with controlled vocabulary in PubMed

Giovana Deliberati Maimone<sup>1</sup>  <http://orcid.org/0000-0002-4150-8084>

Thamyres Vieira dos Santos<sup>2</sup>  <https://orcid.org/0000-0003-3644-8600>

<sup>1</sup>University of São Paulo, São Paulo, SP, Brazil / e-mail: [gdmaimone@usp.br](mailto:gdmaimone@usp.br)

<sup>2</sup>University of São Paulo, São Paulo, SP, Brazil / e-mail: [thamyres.vieira@hotmail.com](mailto:thamyres.vieira@hotmail.com)

### ABSTRACT

**Introduction/Objective:** It presents a theme linked to the analysis and verification of delays in indexing scientific journal articles with the terms of the controlled vocabulary Medical Subject Headings (MeSH) in the PubMed search engine. **Methodology:** Includes bibliographic research for thematic contextualization, and is exploratory and descriptive with practical applicability on subjects related to the COVID-19 pandemic. A search strategy was built on PubMed with 16 terms related to the theme in the MeSH Terms and Text Word fields. The study was published in 2021 in the Portuguese language. The metadata of the 85 publications were exported for analysis in a spreadsheet under the aspects of entry of the publication in PubMed, time until indexing with the MeSH terms and category of the publication. **Results:** About 89% of the publications, considering the sample of 62 items with MeSH Terms, had a delay in indexing of at least 15 days; and about 11% collapsed in 15 to 135 days. **Conclusion:** In order for researchers to be able to retrieve the most scientific content on COVID-19, it is essential that they are constructed of searches that contemplate the use of the MeSH controlled vocabulary term combined with the use of the nomenclature variations of the theme in other fields, such as Text Word, since publications may appear in the search engine, but have not yet been indexed, in order to retrieve a greater number of scientific literature already published.

### KEYWORDS

Scientific and technical information. Indexing. Information retrieval.

## Informação científica em tempos de pandemia análise do tempo de indexação de artigos de periódicos com vocabulário controlado MeSH no PubMed

### RESUMO

**Introdução/Objetivo:** Apresenta temática vinculada à análise e verificação de atrasos na indexação de artigos de periódicos científicos com os termos do vocabulário controlado Medical Subject Headings (MeSH) no buscador PubMed. **Metodologia:** Contempla a pesquisa bibliográfica para contextualização temática, e é exploratória e descritiva com aplicabilidade prática sobre assuntos relacionados à pandemia de COVID-19. Foi construída estratégia de busca no PubMed com 16 termos relacionados à temática nos campos MeSH Terms e Text Word. Para o estudo foram selecionadas publicações de 2021 no idioma português. Os metadados das 85 publicações foram exportados para análise em planilha sob os aspectos de entrada da publicação no PubMed, tempo até a indexação com os termos MeSH e categoria da publicação. **Resultados:** Cerca de 89% das publicações, considerando a amostra de 62 itens com indexação MeSH, tiveram um atraso na indexação de pelo menos 15 dias; e cerca de 11% demoraram

de 15 a 135 dias. **Conclusão:** Para que os pesquisadores consigam recuperar os mais recentes conteúdos científicos sobre a COVID-19, é fundamental que sejam construídas estratégias de busca que contemplem a utilização do termo do vocabulário controlado MeSH aliado à utilização das variações de nomenclatura da temática em demais campos, como o Text Word, já que publicações podem constar no buscador, mas ainda não terem sido indexadas, a fim de recuperar maior número de literatura científica já publicada.

#### **PALAVRAS-CHAVE**

Informação científica e tecnológica. Indexação. Recuperação da informação.



**JITA:** IC. Index languages, processes and schemes.

## 1 INTRODUCTION

Information is a valuable asset for organizations in contemporary society, being used for strategic decision making and as input for the development of best practices. Similarly, scientific information is also fundamental to the development of science and its social impact policies. In the context of the pandemic of COVID-19, this role has become even more evident not only in academia, but in society in general: scientific information is positioned at the heart of the discussions and is an element for the development of research aimed directly at establishing medical, political and social conducts to face the disease.

COVID-19, a disease caused by the SARS-CoV-2 virus (or the new coronavirus) emerged in late 2019 in China and rapidly expanded around the world the following year, demanding multiple efforts from health professionals and the scientific community. Leivas-Cepas, Romero-Rodriguez, and Barroso Sevillano (2020) state that in the face of the emergence of the pandemic as a dissonant social element, considering its impact on social life, scientific research has the central role of answering questions such as "why" and "how."

Scientific information in health produces knowledge and guides the actions of professionals through scientific communication initiatives, i.e., those aimed at the academic community. Targino (2000, p. 1) states that "[...] thinking about the relevance of science demands recognizing the importance of scientific information, scientific knowledge, the scientific community, and, consequently, scientific communication. Formal initiatives for communicating scientific information consist of sources such as databases, repositories, platforms, and servers; while informal initiatives consist of blogs, websites, or academic-scientific social networks. However, not all initiatives are assertive in ensuring the retrieval of quality information that meets the needs of the researcher. Sepúlveda-Vildósola, Mejía-Aranguré, Barrera-Cruz, Fuentes-Morales and Rodríguez-Zeron (2020) discuss about the race of scientists to understand the behavior of the virus causing COVID-19 and develop appropriate interventions for the prevention and treatment of the disease. In this sense, quality science cannot be disregarded, and decisions about the pandemic should not only be made based on the practice of health professionals, but also through established national policies and based on scientific knowledge (publications in formal channels such as scientific journals).

The specialized health and multidisciplinary scientific databases are traditionally used by researchers looking for certified scientific information, considering that they have already gone through the peer review process and belong to scientific journals that have also been approved by a series of quality criteria to be indexed in these information sources.

However, when faced with an emergency situation such as a pandemic, researchers resort to various sources in order to access the most appropriate and safe information in the shortest time, whether hosted by scientific journals or repositories. Another source of information consulted are the servers that store preliminary results of research conducted, such as preprints<sup>1</sup>, although in these cases it is known that they are data that cannot be used for clinical decision making, considering that they have not undergone a prior evaluation process, i.e., they have not been evaluated and peer-reviewed, as happens with publications that belong to scientific journals.

The scientific community and its forms of publication have undergone readjustments since the beginning of the pandemic in order to evaluate and make studies available quickly, while ensuring quality. However, the indexing time of scientific journals in referential and

---

<sup>1</sup> Preprints are scientific publications, a model aligned to open science practices, made available on the web (as repositories for this type of publication) before going through the peer review process, which traditionally happens with articles submitted to scientific journals.

multidisciplinary databases is the target of studies and observations even before the pandemic, finding that many papers have a significant delay between the time of publication and their indexing in databases, i.e., there is scientific information published, but it is not available for retrieval by researchers through the construction of search strategies.

Given this problem, it is common to use alternative sources of information so that the timing of the discovery of scientific literature with possible novelties and its consequent implementation is not affected. These sources can be platforms or websites with a compilation of scientific information, in this case about the disease, initiatives from publishing groups, universities, or other research institutions.

This paper aims to discuss the role of information sources contained in scientific databases in the fight against COVID-19 and how possible delays in indexing time make them unproductive in the face of the need for rapid content to be made available to the scientific community; in addition to specifically verify how delays in indexing publications on COVID-19 in the PubMed search resource with the terms of the controlled vocabulary Medical Subject Headings (MeSH<sup>2</sup>) impact the development of policies to fight the disease.

In this context, the pandemic scenario assists, by necessity, the understanding of the services developed by specialized health information centers and also of the studies encompassed by the area of Information Science, highlighting the importance of fast and effective retrieval for researchers that directly impacts the development of science and technology, which in this particular context can accelerate results for learning about the disease and its thematic environments.

## 2 IMPACT OF THE COVID-19 PANDEMIC ON SCIENTIFIC PUBLICATIONS

| 4

COVID-19 altered the social dynamics not only for its serious impact on health, but also for its interference in the economic, political, and health perspectives. As a result, the gaze of a considerable part of society has turned to the academic-scientific community in search of answers.

Sepúlveda-Vildósola, Mejía-Aranguré, Barrera-Cruz, Fuentes-Morales, and Rodríguez-Zeron (2020) discuss that health professionals had to decide quickly about the best way to treat a previously unknown disease, considering its worldwide expansion and the consequent threat or effective collapse of health systems. Clinical research, evidence-based medicine, randomized studies, and systematic reviews are some of the tools that enable the establishment of health policy, and all information is the result of a process of scientific investigation that takes time. The organization of all these studies initially went through moments of instability, considering that at the beginning of the pandemic the experts were looking for adequate terms to describe the global scenario that was hitting us from that moment on.

This global emergency generated scientific research and articles, epidemiological and laboratory data, information from news organizations, multimedia documents, and even the need for terminological definition of how the disease and the virus would be known and at first the diverse denomination of both made it difficult to organize the data and information (SANTOS, 2020, p. 76).

---

<sup>2</sup> Controlled vocabulary used to index documents in PubMed. From it, the Brazilian Virtual Health Library (VHL) developed the Health Sciences Descriptors (DeCS), used to index and retrieve scientific publications in the information sources it is responsible for managing.

In addition to searching the formal sources consulted by the scientific community as national and international databases, it became evident the importance of advocating the open science movement and its initiatives. Researchers working on such research have made their preliminary results available on preprints servers, a practice that was already common in areas such as physics and mathematics (CALLAWAY, 2020), but which still generates discussions in the scientific community. Damasio (2018, p. 166) states that "in this context, the paradigm shift of anticipating peer review in open access has currently been one of the main discussions in scientific publishing and communication."

Since the beginning of the pandemic, several editorial groups and scientific journals have provided free access to disease-related publications. The manuscript review process by the editorial staff of scientific journals did their work in record time (CALLAWAY, 2020) with the aim of making scientific literature on the subject available in the shortest possible time. In this context, the amount of scientific literature available on the subject has grown since the onset of the disease, with studies attempting to address the main questions regarding prevention, diagnosis, and treatment (LEIVA-CEPAS; ROMERO-RODRIGUEZ; BARROSO-SEVILLANO, 2020, p. 1).

It is clear that the availability of literature on the disease is fundamental to the development of treatments and the dedication of several researchers worldwide must be truly accessible by their peers so that together they can advance the topic. These initiatives are beneficial for scientific development "[...] even with the difficulties of maintaining high quality standards of content" (CUNHA, 2020, p. 758). Although health care and mental health is a priority at this time (EINSEIN; AKHMANOVA, BEHRENS; WEIGEL, 2020), the editorial process proved efficient in disseminating research and researchers engaged in developing relevant publications on the topic.

The pandemic scenario in which the world finds itself imposes changes in social behavior that may or may not favor scientific advancement and the conditions of researchers for such development. The challenge of maintaining high quality standards in publications coupled with misinformation driven by political reasons and the influence of social media, may be a sequel with the potential to negatively affect the health (physical and mental) of millions of individuals (DEVANE, 2020).

Another point regarding the quality of research, which is intrinsically linked to the intense dissemination of information driven by the web is found in the level of retraction of articles during the pandemic and that have it as a topic addressed. Ioannidis (2020) comments on the case of a preprint on coronavirus that has a high altimeter indicator, which verifies the level of attention in web publications, but which was harshly criticized and withdrawn within a few days of its publication. In this case, although the journal has identified the error and removed the article, the reach of that publication was high and consequently the results and considerations of the paper were possibly internalized by researchers, i.e., although it has already been removed from the journal, its content will continue to reverberate among those who had the opportunity to access it and somehow can use it to inform new researches, clinical conducts, or health policies. Another case widely discussed on an international level was the retraction of scientific articles published in two prestigious medical journals, *The Lancet* and *The New England Journal of Medicine*, which brought data, later questioned and that presented flaws, about treatment for COVID-19. Although both publications were retracted by the journals, it is unquestionable how much their reach intensified debates in the scientific community and society, involving social and political issues amidst the chaos caused by the effects of the pandemic, potentially causing insecurity regarding further studies on the disease. The retractions, although they have always happened in the scientific community, in the current

context are a reflection of the intense productivity and engagement of researchers around the world.

As mentioned by Torres-Salinas (2020), we are facing the largest concentration of scientific resources to solve a specific problem, surpassing even precedents such as the Manhattan project and the Apollo mission. There is much more scientific literature than researchers can keep up with and, consequently, they find it difficult to manage the focus of their studies. In view of this, Torres-Salinas (2020) proposes an analysis of the amount of scientific production on COVID-19 and its rate of growth in databases, search services and in repositories. According to the study conducted by the author in April 2020, the Web of Science base had 764 publications, Scopus 1568, Dimensions 9435, PubMed 4291 and repositories 4075. However, the amount of publications available for consultation by researchers is not in its entirety full articles or other publications with a high level of scientific evidence, according to a study conducted by Sepúlveda-Vildósola, Mejía-Aranguré, Barrera-Cruz, Fuentes-Morales, and Rodríguez-Zeron (2020). More than half of the scientific production available in the Web of Science and Scopus databases on April 30, 2020 corresponded to publications in the following categories: editorials, letters, comments, opinions, news, proceedings, book chapters, or questionnaires. Another issue that must be pointed out in face of the availability of these contents in the information sources is the common indexing of journals and consequently of articles in more than one database or repository, i.e., some of the contents found in a base x can also be found in a base y, causing duplicities in the contents found between the two sources, i.e., if the researcher consults these two sources, this fact may lead him to evaluate or verify more than once the same publication in distinct information sources.

Thus, the scientific community is faced with a considerable amount of literature that is available in various sources of information, but not all of this informational volume actually adds to the development of new practices, research, and conduct in health. The researcher, when counting on the search resources and navigation tools offered by the databases, many of them signed by the institutions to which they belong, has the possibility of organizing himself in such a way that this avalanche of contents goes through filters according to his interest and area of activity, also considering the identification of the relevance of the publication by the researcher; besides the possibility of using bibliographic managers, such as Mendeley or EndNote, so that he can follow the new papers that are published daily, using these tools to eliminate the duplicities mentioned and avoiding these discomforts throughout the research process.

| 6

### 3 DELAY IN THE INDEXING TIME OF THE PUBLICATIONS IN THE DATABASES

Scientific communication in a digital environment is a facilitator in the process of making literature available in a more "horizontalized" way, in quotation marks because it is directed to the academic community, but comprehensive since its immediate access is possible through institutional subscriptions, for example, or through deposit in academic-scientific social networks such as ResearchGate or Academia.edu, repositories or direct contact with authors.

Wersig (1993) states that knowledge has become more important than ever, and technologies aim to reduce its complexity. In this context, for the scientific community to have access to published literature, it is fundamental that it be in some way retrievable and available in electronic information sources such as databases. Thus, the prior indexing of these contents is assumed; however, the indexing of scientific journals involves intense discussion about quality, metrics, and author productivity, in addition to the indexing policy adopted by institutions.

A journal indexed in a database has scientific relevance and will have greater prestige since, generally, its impact is more evident, precisely because it has undergone selection criteria that others - not indexed in databases or search engines - were not submitted to.

In the scientific community, as Balhara (2012) states, indexed journals are considered of higher quality than those that are not indexed, but the question remains as to what type of indexing would be valid, that is, how to compare the quality of articles from journals that are indexed in different sources.

Indexing is seen as an indicator of the quality of scientific journals mainly due to the rigorous and judicious process to which journals are submitted to be incorporated into collections. There are a series of requirements and standards that must be followed so that all journals have a similar level of "quality", although this concept is not very clear to many members of the scientific community.

Consequently, researchers are encouraged to publish in those journals that belong to recognized national or international database collections, i.e. that have this level of "quality" and that automatically carry more weight when researchers are submitted to productivity evaluations or need to achieve goals in their research departments. At the institutional level we can mention, for example, the case of university rankings: to calculate a university's scientific productivity score, some of them consider only the publications that are indexed in the Web of Science<sup>3</sup> database. Clearly this is a criterion established by the ranking organizers, but it automatically disregards all scientific production that is not indexed in the selected information source.

Besides the discussions that involve the indexing of the journals in the databases and their quality, another point discussed in the scientific community is the time it takes for the journals to be indexed in the databases, not the time for the adequacy of requirements and eligibility for their aggregation to the collection, but the time for the journal articles to be available there and consequently be retrievable by peers.

PubMed, considered a reference service for searching scientific articles and maintained by the National Library of Medicine (NLM) in the United States, has MeSH controlled vocabulary to perform subject indexing of journal articles that belong to the collections indexed in Medline, a database that contains articles from several journals worldwide in the Health Sciences and Biomedical areas. Lancaster (2004, p. 3) points out that "the aim, of course, when searching a database, is to find documents that are useful for satisfying an information need, and to avoid retrieving useless items". Thus, controlled vocabularies are used for assertive retrieval of bibliographic information.

Documents can be "represented" in several ways, one of them is through the attribution of subjects to the material to be later recovered, an activity called indexing. According to the Brazilian Association of Technical Standards (ABNT) standard NBR 12676 indexing can be defined as:

The act of identifying and describing the content of a document with terms that represent its subject matter and constitute an indexing language (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p.2).

In this context and very important for the present work is the observation that Lancaster (2004) makes about the insertion of data from other information fields, such as titles, authors, sources, etc. for the same document, which the referred author calls descriptive cataloging. In

---

<sup>3</sup> Institutional subscription access: <https://webofknowledge.com>

this sense, when the indexing is done by controlled vocabulary in databases such as PubMed, there is an entry time of documents in the system, as we will see later; then when free descriptors are used and the search is performed by more fields than just subject by controlled vocabulary terms, there is the possibility of finding a larger set of documents. However, the retrieval of more material, provided by the use of free terms and in other fields of information, does not mean that these have precision, but the revocation will certainly be greater. The use of controlled vocabulary in the search directly impacts the assertiveness of content retrieval in information systems.

A controlled vocabulary is essentially a list of authorized terms. In general, the indexer can only assign terms to a document that are on the list adopted by the institution for which he works. Commonly, however, a controlled vocabulary is more than just a list. It generally includes a form of semantic structure. This structure is intended, in particular, to:

1. control synonyms (...);
2. differentiate homographs (...);
3. bring together or link terms whose meanings are more closely related to each other (...) (LANCASTER, 2004, p. 19).

The use of controlled vocabularies by scientific information sources is an effective strategy for content retrieval, especially due to the specificities of the knowledge areas and their respective themes. Rodriguez (2016) points out that the search of scientific literature by simple terms causes the non-retrieval of relevant literature, since many articles may use different terminology to refer to the same topic, and also causes the retrieval of results with false positives, i.e., results that actually contain the search term inserted, but outside the ideal context and that do not satisfy the informational need of the researcher. On the other hand, it is pertinent that information sources continue to make available in their search interfaces the possibility of searching for simple terms, that is, terms that do not belong to the controlled vocabulary, so that the researcher also has the freedom to look for alternatives to find the contents he needs. As Araújo Júnior (2007, p. 41) states

[...] the procedure that seems more rational is the use of both possibilities, that is, both the use of controlled vocabulary when the situation requires it, and the use of free terms in order to seek alternatives for the thematic representation.

This autonomy gives freedom to the researcher to search for subjects that have a specific nomenclature and are little known outside the area of knowledge and that may not have been covered by the indexing instrument, that is, the controlled vocabulary; besides contributing to the optimization of the instrument by the team or institution responsible for its elaboration/maintenance.

If technology and the development of information networks have enabled the diffusion of knowledge, they have also boosted the direct publication of information at the source to the users of information retrieval systems. This generates a lack of standard, and consequently hinders search and retrieval in virtual environments. Users' information needs vary from one domain to another, from one group to another, according to the stage of knowledge, the nature of the users, their goals. Despite these variations, information needs to be reliable, current, properly indexed for retrieval and immediate access. The recovery is related to the storage media, and these to the treatment and organization of information (CARDOSO FILHO; SANTOS, 2012, p. 215).

And this possibility of searching and retrieving scientific documents by terms in fields without the use of MeSH controlled vocabulary is the target of studies in the area of Health Sciences, evidencing that this is a recurring concern of researchers that has been observed during research practice.

Jenuwine and Floyd (2004) conducted a study in which they constructed two search strategies, one using MeSH terms and the other using the Text Word field that retrieves publications by title, abstract and among others, with the objective of verifying how the two strategies performed in the retrieval of scientific articles on the subject of sleep; and concluded that each of the strategies retrieved unique results and that combining them provides greater retrieval of results, showing their potential for complementarity and highlighting that the effectiveness of the results by the Text Word field can be maximized if the writing of titles and abstracts by the authors of the articles is optimized and uses the terminologies of the study area consistently.

The consistent use of terminologies in scientific articles, whether in the authors' keywords, title or abstract, and their impact on the retrieval of scientific literature by fields that did not use controlled vocabulary goes back to discussions such as the one proposed by Brandau, Monteiro and Braile (2005) regarding the correct use of descriptors in publications, which should be encouraged by the journals and organizations responsible for them, as well as an awareness on the part of authors regarding the use of a structured vocabulary to optimize the retrieval of their own works in search systems.

While Minguet, Boogerd, Lopes, Salgado, Correr and Fernandez-Llimos (2014) conducted a study that, although not focused on a comparison between the use of controlled vocabulary in PubMed and the use of other fields for retrieval of scientific literature, aimed to identify a gap of MeSH terms for publications in the area of Pharmacology and found an imbalance in the vocabulary coverage of this area compared to the areas of Nursing and Dentistry. Although these are not conclusions and hypotheses raised by the authors of the study, the combined use of other fields of research in conducting the search could be used, considering the proper analysis of pertinence of the results, to find scientific publications that were not retrieved due to lower coverage of MeSH terms in the area studied.

In turn, Utagawa, Gambarato, and Pereira (2018), conducted a study regarding the use of descriptors in scientific articles and found that the use of MeSH controlled vocabulary in scientific research provided fewer search results, but were assertive and pertinent in relation to the desired theme, while articles retrieved exclusively through other fields, without the use of controlled terms, were less suited to the researched theme.

Although the researcher's broad search possibilities in information sources is interesting, the impact of the agile indexing process and the use of controlled vocabulary to retrieve scientific literature in search tools effectively is evident. In this sense, Rodriguez (2014, p. 322) discusses the notable delay between the entry of the paper in PubMed and its indexing with MeSH terms, that is, the article is hosted in PubMed, but as it has not yet received the indexing it does not become retrievable through the use of controlled vocabulary, a fact that excludes the most recent articles from a search conducted through MeSH, causing loss in the retrieval of relevant articles by researchers.

Silva and Fujita (2004, p. 136) state that "[...] according to Gardin's theoretical line, indexing is seen as a document representation operation with the pragmatic purpose of information retrieval". Considering the context of the pandemic of COVID-19, the dimension of the loss of scientific information by researchers and research groups that use only MeSH functionality, without free search, is amplified and may have as a consequence delays in the development of new conducts and treatments. What also potentiates this discussion is the fact that PubMed, as a search service, is considered a reference in the area, in addition to being free,

although not all publications available there are open access, and consequently it is more used by several researchers and institutions. This delay in indexing has been the target of study and proof by researchers in the last ten years (RODRIGUÉZ, 2014; RODRIGUEZ, 2016; TORT; TARGINO; AMARAL, 2012; MINGUET; SALGADO; BOOGERD, FERNANDEZ-LLIMOS, 2015; IRWIN; RACKHAM, 2017) fact that does not isolate the pandemic situation experienced.

The study conducted by the researchers Minguet, Salgado, Boogerd and Fernandez-Llimos (2015) verified the delay in the indexation of publications in the top ten journals in the area of Pharmacology, using five years as the period for analysis. The authors Irwin and Rackham (2017), on the other hand, analyzed the indexing time of publications of journals in the Health Sciences area according to their Impact Factor (IF) and area and also found a significant delay between the availability of publications in PubMed and their indexing with MeSH terms, even concluding that the delay affects the retrieval of articles and consequently can decrease the probability of citations. The authors contacted the NLM in order to understand a little more about the reasons that directly affect the indexing time; the arguments presented were the increase in the number of journals and articles available in PubMed, government budget issues, and a smaller staff that directly implies less time to process the records (MINGUET; SALGADO; BOOGERD, FERNANDEZ-LLIMOS, 2015). Tort, Targino and Amaral (2012) also verify the delay between the publication of an article and its indexing in scientific information sources, however, this is a study aimed at verifying the impact of this delay on the IF of scientific journals.

Thus, it is pertinent that an analysis and evaluation of the indexing time of scientific publications on the pandemic of COVID-19 in PubMed with MeSH terms be carried out, considering that it is an open access search engine and that it is recognized in the search for scientific literature by researchers in the area of Health Sciences, and delays in indexing have a direct and negative impact on the retrieval/use of content by the academic-scientific community.

## 4 METHODS

Considering the incidence of scientific literature on the theme and also the relevance and recognition in the academic community, the PubMed search service was chosen with the objective of verifying the delay in indexing publications with MeSH controlled vocabulary terms.

Based on the choice of PubMed for analysis, the terms were decided to perform test searches using different variations that encompass the scientific literature related to the COVID-19 pandemic, to ensure that the search and consequently the study would assertively recover material on the subject: wuhan coronavirus; covid-19; 2019 ncov; 2019-ncov; 2019ncov; new coronavirus; SARS-CoV-2; SARS CoV 2; 2019 novel coronavirus disease; 2019 novel coronavirus infection; 2019-nCoV disease; 2019-nCoV infection; coronavirus disease 2019; coronavirus disease-19; coronavirus infections and coronavirus infectious.

Variations of the terms in different languages were not used, only in English, because the database selected for the study indexes the articles with the metadata in the English language. Thus, the use of the terms retrieves papers in different languages, including those published by Brazilian journals indexed in PubMed.

With the chosen terms, the search strategy (Chart 1) was built considering the availability of fields and the respective specificities of the information source, but without changing the previously mentioned search terms. The search expression and its special

characters such as parentheses, brackets and over spacing are automatically inserted by PubMed when the strategy is built, so that it varies according to the system or mechanism used.

**Chart 1.** Search strategy and fields used in PubMed.

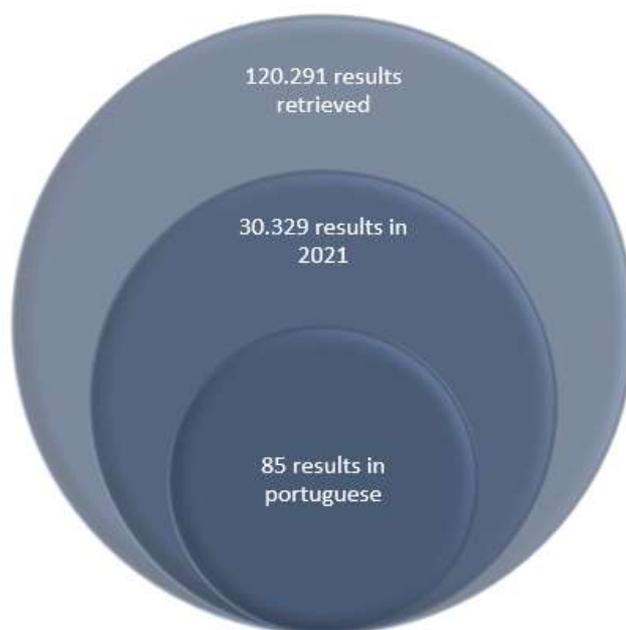
SERVIÇO	SEARCH STRATEGY	FIELDS USED
PUBMED	("coronavirus infections"[MeSH Terms]) OR (((((((((((("wuhan coronavirus"[Text Word]) OR ("covid 19"[Text Word])) OR ("2019 ncov"[Text Word])) OR ("2019ncov"[Text Word])) OR ("new coronavirus"[Text Word])) OR ("sars cov 2"[Text Word])) OR ("2019 novel coronavirus disease"[Text Word])) OR ("2019 novel coronavirus infection"[Text Word])) OR ("2019 ncov disease"[Text Word])) OR ("2019 ncov infection"[Text Word])) OR ("coronavirus disease 2019"[Text Word])) OR ("coronavirus disease 19"[Text Word])) OR ("coronavirus infections"[Text Word])) OR ("coronavirus infectious"[Text Word]))	MeSH Terms; Text Word

Source: elaborated by the authors, survey data (2021).

In the PubMed search engine, the advanced search modality was used, first building the search strategy by the term that appears in the MeSH vocabulary, in this case, coronavirus infections, in the "MeSH Terms" field; after this, all other terms and also the MeSH term were inserted into a strategy built based on the "Text Word" field, which retrieves information that appears in the title, abstract, and other fields, with the objective of also mapping the papers that have not yet been indexed with MeSH terms; and finally, the two search strategies were consolidated into a single one, always using OR as a Boolean operator, with the objective of retrieving the materials that contain all the search terms previously selected, broadening the results found.

The search was performed on March 15, 2021. Before the total number of results retrieved by the search strategy (120,291), we filtered by the year 2021 (30,329) and then by the Portuguese language (85), with the aim of making a more recent cut in the study and contextualized to the literature on the disease produced in Portuguese this year, regardless of the country (Figure 1). As directed by the National Library of Medicine (2021) in the PubMed user guide, the filter by language restricts the search results to publications written in the selected language. However, it is noted that the sample filtered by the Portuguese language presents the metadata of the publications in English, a characteristic of PubMed, and some of them have bilingual versions in English and Portuguese, which at first may generate the perception that the filter does not retrieve literature produced in Portuguese. Filtering was also performed to make data analysis feasible. No additional filters were performed beyond those mentioned, i.e., the number of results indicates the total number of records retrieved by the search strategy regardless of the publication category (whether article, editorial, notes, etc.).

**Figure 1.** Search clipping with the publications selected for the study.



Source: elaborated by the authors, survey data (2021).

Thus, the metadata of the 85 publications were exported for analysis in a spreadsheet under the following aspects: publication entry in PubMed, time to indexing with MeSH terms, and publication category. To identify the meaning of the elements used for metadata description, the guidance made available by the National Library of Medicine (2019) was used. To verify the entry of the publication in PubMed, the field "Entrez Date (EDAT)" was queried, the field "MeSH Date (MHDA)" to verify the date that the publication was indexed with MeSH controlled vocabulary terms, and the field "Publication Type (PT)" to verify the publication category. If the contents of the two date fields, EDAT and MHDA, in the same publication are the same, that is, they have the same date, it means that the publication has not yet been indexed. The three metadata fields of the set of 85 publications were analyzed manually.

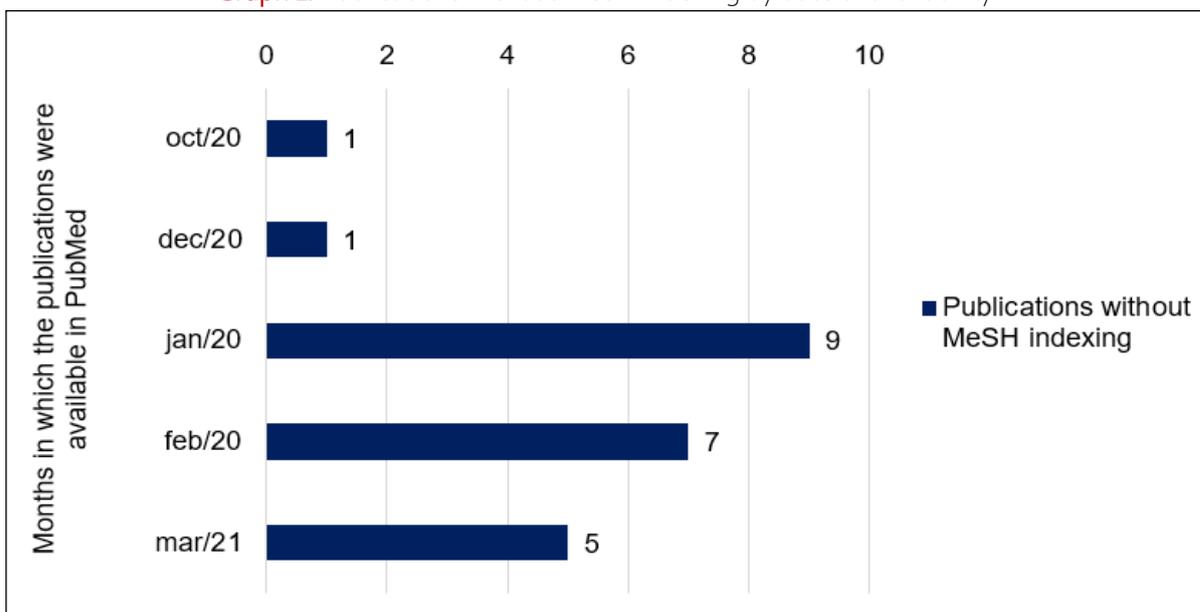
| 12

## 5 RESULTS

Considering the total of 85 publications, 23 of them were not indexed with MeSH terms, representing 27%. Chart 1 shows the months in which the publications without indexing were available in PubMed, especially the first three months of 2021. Although the filter performed initially was only for the year 2021, i.e., for publications that were included in issues of this year in the respective journals, some publications were made available online a few months before the official publication, called ahead of print, so the study results mention the last months of 2020.

Among the publications without MeSH indexing, the following publication categories are present, following the same nomenclature adopted in the PubMed metadata: *journal article*, *letter*, *case report*, and *erratum*.

Graph 1. Publications without MeSH indexing by date of availability.



Source: elaborated by the authors, survey data (2021).

Among the publications with MeSH indexing, which represent 73% of the total publications analyzed in the study, the following publication categories are present, also following the same nomenclature adopted in the PubMed metadata: *journal article, case report, commentary, letter, editorial, and erratum.*

This universe of 62 publications with MeSH indexing, which together took 872 days to receive indexing, took on average 14 days between the date of entry in PubMed, that is, when the paper was available for consultation, and the date of indexing with controlled vocabulary terms.

Table 1. Delay period between paper publication and indexing.

DELAY PERIOD BETWEEN PAPER PUBLICATION AND INDEXING	PUBLICATIONS
Up to 1 week (1 to 7 days)	40
1 to 2 weeks (8 to 14 days)	15
3 to 4 weeks (15 to 30 days)	0
5 to 6 weeks (31 to 45 days)	0
7 to 8 weeks (46 to 60 days)	3
9 to 10 weeks (61 to 75 days)	1
11 to 12 weeks (76 to 90 days)	1
13 to 14 weeks (91 to 105 days)	0
15 to 16 weeks (106 to 120 days)	0
17 and 18 weeks (121 to 135 days)	2

Source: elaborated by the authors, survey data (2021).

When analyzing the number of days between the date of publication or availability of the paper and the date of indexing with MeSH terms (Table 1), we see that about 89% of the publications, considering the sample of 62 items, had a delay of at least 15 days; and about 11% had a delay of 15 to 135 days.

If we consider that 40 publications (more than 50%), out of the 62, had a delay of up to one week for indexing we can say that there is a significant delay, considering the urgency of the case. However, with the need for ephemeral research and advancement, the faster the access, the faster the verification of its effectiveness in real terms

## 6 FINAL CONSIDERATIONS

Scientific information, which has always been a fundamental input for the development of science, had its role highlighted in the face of the health crisis caused by the coronavirus, which began in 2019. The scientific community, which was dedicated to the most diverse research inserted in different areas of knowledge, directed its efforts, when possible and applicable, to somehow study and reflect about the still present moment and be able to point future paths.

The number of publications about COVID-19 has grown exponentially, and the most diverse efforts resulted in studies that could be used to guide actions to confront the disease.

In this context, the changes in scientific communication resulting from this global medical emergency come into discussion, such as the increased use of preprints platforms, a faster publication process (including study design, peer review, layout, and availability), and free access to publications on the subject even by restricted access journals.

The speed in this process culminated in the need to verify the indexing time of scientific articles entered in PubMed, considering its visibility as an information source in the Health area, with terms from the MeSH controlled vocabulary. Thus, it was possible to verify the delay time in this indexing and its possible impact on scientific information retrieval by researchers who use this search resource.

The sample of publications used in the present study shows that there is a significant delay between the publication and indexing of scientific articles with MeSH controlled vocabulary terms in PubMed.

Although this delay is up to 15 days with about 89% of the publications analyzed, it should not be disregarded due to the urgency of the theme and its access, since it directly impacts the decision of the paper being retrieved or not by a search strategy built only with MeSH terms. It is necessary to say that access to the document includes possibilities of quick knowledge to make decisions of essential importance to face the pandemic (treatments, vaccines, protective equipment, hospital monitoring equipment, among others).

In the literature it was possible to verify that the controlled vocabulary is the most efficient strategy for scientific information retrieval in information sources, but that it must also be allied, when pertinent and if well used, to the retrieval by free terms. For the verification and execution of the present research, free terms were used with the objective of retrieving all scientific production on the subject indexed in PubMed, with or without MeSH terms; however, this activity demands more time and articulation of the variations in the nomenclature of the theme sought by the researcher.

Thus, for researchers to be able to retrieve the most recent scientific content about COVID-19, it is essential that search strategies be built that contemplate the use of MeSH controlled vocabulary terms allied to the use of variations in the nomenclature of the theme in other fields, such as Text Word; guaranteeing that no essential content for the development of research, conduct, and policies to fight the pandemic is not retrieved at the time of the bibliographic survey. For this, it is necessary to guide researchers, at the time of the search, to use fields other than those involving the MeSH controlled vocabulary, since publications may appear in the search engine, but have not yet been indexed.

## CRediT

**RECOGNITIONS:** Not applicable.

**FINANCING:** Not applicable.

**INTEREST CONFLICTS:** The authors certify that they have no commercial or associative interest that represents a conflict of interest in relation to the manuscript.

**ETHICAL APPROVAL:** Not applicable.

**AVAILABILITY OF DATA AND MATERIAL:** Not applicable.

**AUTHORS 'CONTRIBUTIONS:** - Conceptualization: SANTOS, T. V.; MAIMONE, G.D.; Data Curation, Formal Analysis, Research, Methodology, Visualization, Writing – original draft: SANTOS, T. V.; Project Administration, Supervision, Validation, Writing – revision & editing: MAIMONE, G. D.

## REFERENCES

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR 12676:** Métodos para análise de documentos: determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro: ABNT, 1992.

ARAÚJO JÚNIOR, Rogério Henrique de. **Precisão no processo de busca e recuperação da informação.** Brasília, DF: Thesaurus, 2007.

BALHARA, Yatan Pal Singh. Indexed journal: what does it mean? **Lung India**, Mumbai, v. 29, n. 2, p. 193, apr./jun. 2012. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3354504/>. Access on: 20 dez. 2020.

BRANDAU, Ricardo; MONTEIRO, Rosangela; BRAILE, Domingo M. Importância do uso correto dos descritores nos artigos científicos. **Revista Brasileira de Cirurgia Cardiovascular**, São José do Rio Preto, v. 20, n. 1, p. VII-IX, jan./mar. 2005. Available at: [https://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0102-76382005000100004](https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-76382005000100004). Access on: 23 maio 2021.

CALLAWAY, Ewen. The COVID-19 crisis could permanently change scientific publishing. **Nature**, London, v. 582, p. 167-168, jun. 2020. Available at: <https://www.nature.com/articles/d41586-020-01520-4>. Access on: 11 jun. 2020.

CARDOSO FILHO, Jair Cunha; SANTOS, Márcia Mazo. Processos e temas selecionados. *In:* ALVARES, Lilian (org.). **Organização da informação e do conhecimento: conceitos, subsídios interdisciplinares e aplicações.** São Paulo: B4 Editora, 2012. p. 185-224.

CUNHA, Murilo Bastos da. Os impactos da COVID-19 nas áreas da Ciência da Informação. **Revista Ibero-Americana de Ciência da Informação**, Brasília, DF, v. 13, n. 3, p. 756-759, 2020. Available at: <https://periodicos.unb.br/index.php/RICI/article/view/33373>. Access on: 3 jan. 2021.

DAMÁSIO, Edilson. *Preprints* na comunicação científica: uma introdução. **Biblos: revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 32, n. 2, p. 155-168, jul./dez. 2018. Available at: <https://periodicos.furg.br/biblos/article/view/8635>. Access on: 31 mar. 2021.

DEVANE, C. Lindsay. The role of scientific publishing in the SARS-CoV-2 Pandemic. **Pharmacotherapy**, Carlisle, v. 40, n. 5, p. 376-378, may 2020. Available at: <https://pubmed.ncbi.nlm.nih.gov/32364280/>. Access on: 20 dez. 2020.

EISEIN, Michael B.; AKHMANOVA, Anna; BEHRENS, Timothy; WEIGEL, Detlef. Publishing in the time of COVID-19. **eLife**, Cambridge, v. 9, n. e57162, p. 1-3, mar. 2020. Available at: <https://pubmed.ncbi.nlm.nih.gov/32209226/>. Access on: 20 dez. 2020.

IOANNIDIS, John. P. A. Coronavirus disease 2019: the harms of exaggerated information and non-evidence-based measures. **European Journal of Clinical Investigation**, Oxford, v. 50, n. e13222, mar. 2020. Available at: <https://onlinelibrary.wiley.com/doi/full/10.1111/eci.13222>. Access on: 20 dez. 2020.

IRWIN, Adriane N.; RACKHAM, Daniel. Comparison of the time-to-indexing in PubMed between biomedical journals according to impact factor, discipline, and focus. **Research in Social and Administrative Pharmacy**, New York, v. 13, n. 2, p. 389-393, mar./apr. 2017. Available at: <https://pubmed.ncbi.nlm.nih.gov/27215603/>. Access on: 22 maio 2021.

JENUWINE, Elizabeth S.; FLOYD, Judith A. Comparison of Medical Subject Headings and text-word searches in MEDLINE to retrieve studies on sleep in healthy individuals. **Journal of the Medical Library Association**, Chicago, v. 92, n. 3, p. 349-353, jul. 2004. Available at: <https://pubmed.ncbi.nlm.nih.gov/15243641/>. Access on: 23 maio 2021.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2. ed. Brasília, DF: Briquet de Lemos, 2004.

LEIVA-CEPAS, F.; ROMERO-RODRIGUÉZ, E.; BARROSO SEVILLANO, M. Las revistas científicas ante la pandemia por COVID-19. **Semergen**, Madrid, v. 46, n. S1, p. 1-2, agosto 2020. Available at: <https://www.elsevier.es/es-revista-medicina-familia-semergen-40-articulo-las-revistas-cientificas-ante-pandemia-S1138359320302045>. Access on: 20 dez. 2020.

MINGUET, Fernando; SALGADO, Teresa M.; BOOGERD, Lucienne van den; FERNANDEZ-LLIMOS, Fernando. Quality of pharmacy-specific Medical Subject Headings (MeSH) assignment in pharmacy journals indexed in MEDLINE. **Research in Social and Administrative Pharmacy**, New York, v. 11, n. 5, p. 686-695, sep./oct. 2015. Available at: <https://pubmed.ncbi.nlm.nih.gov/25498253/>. Access on: 22 maio 2021.

MINGUET, Fernando; BOOGERD, Lucienne van den; LOPES, Silvia C.; SALGADO, Teresa M.; CORRER, Cassyano J.; FERNANDEZ-LLIMOS, Fernando. Processo automático de identificação de termos MeSH em falta: caso prático de Farmácia. In: XI JORNADAS APDIS, 2014, Lisboa. **Anais [...]**. Lisboa: Associação Portuguesa de Documentação e Informação em Saúde, 2014. Available at: <https://apdis.pt/publicacoes/index.php/jornadas/article/view/49/41>. Access on: 23 maio 2021.

NATIONAL LIBRARY OF MEDICINE. **MEDLINE®/PubMed® data element (field) descriptions**. Bethesda: NLM, 2019. Available at: <https://www.nlm.nih.gov/bsd/mms/medlineelements.html#mhda>. Access on: 4 jan. 2021.

NATIONAL LIBRARY OF MEDICINE. **PubMed user guide**. Bethesda: NLM, 2019. Available at: <https://pubmed.ncbi.nlm.nih.gov/help/#filters-language>. Access on: 22 may. 2021.

RODRIGUEZ, Ryan W. Comparison of indexing times among articles from medical, nursing, and pharmacy journals. **American Journal of Health System Pharmacy**, Bethesda, v. 73, n. 8, p. 569-575, apr. 2016. Available at: <https://pubmed.ncbi.nlm.nih.gov/27045069>. Access on: 22 dez. 2020.

RODRIGUEZ, Ryan W. Delay in indexing articles published in major pharmacy practice journals. **American Journal of Health System Pharmacy**, Bethesda, v. 71, n. 4, p. 321-324, feb. 2014. Available at: <https://pubmed.ncbi.nlm.nih.gov/24481157/>. Access on: 16 dez. 2020.

SANTOS, Cibele Araújo Camargo Marques dos. A organização e representação do conhecimento na pandemia de COVID-19: contribuições e desafios. **Revista Fontes Documentais**, Aracaju, v. 3, n. esp., p. 75-85, 2020. Available at: <https://aplicacoes.ifs.edu.br/periodicos/index.php/fontesdocumentais/article/view/649/498>. Access on: 13 mar. 2020.

SEPÚLVEDA-VILDÓSOLA, Ana Carolina; MEJÍA-ARANGURÉ, Juan Manuel; BARRERA-CRUZ, Carolina; FUENTES-MORALES, Narxeli Alexandra; RODRIGUÉZ-ZERON, Carlos. Scientific Publications During the COVID-19 Pandemic. **Archives of Medical Research**, New York, v. 51, p. 349-354, july 2020. Available at: <https://pubmed.ncbi.nlm.nih.gov/32507366/>. Access on: 20 dez. 2020.

SILVA, Maria dos Remédios da; FUJITA, Mariângela Spotti Lopes. A prática de indexação: análise da evolução de tendências teóricas e metodológicas. **Transinformação**, Campinas, v. 16, n. 2, p. 133-161, maio/ago. 2004. Available at: <https://www.scielo.br/pdf/tinf/v16n2/03.pdf>. Access on: 01 abr. 2021.

TARGINO, Maria das Graças. Comunicação científica: uma revisão de seus elementos básicos. **Revista Informação & Sociedade: estudos**, João Pessoa, v. 10, n. 2, p. 37-85, 2000. Available at: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/326>. Access on: 23 dez. 2020.

TORRES-SALINAS, Daniel. Ritmo de crescimento diário de la producción científica sobre COVID-19. Análisis em bases de datos y repositorios em acceso abierto. **El profesional de la información**, Barcelona, v. 29, n. 2, p. 1-6, abr. 2020. Available at: <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/79259>. Access on: 20 dez. 2020.

TORT, Adriano B. L.; TARGINO, Zé H.; AMARAL, Olavo B. Rising publication delayx inflate Journal Impact Factors. **PLOS One**, San Francisco, v. 7, n. 12, e53374, dec. 2012. Available at: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0053374>. Access on: 15 dez. 2020.

UTAGAWA, Cláudia Yamada; GAMBARATO, Bruno Chaboli; PEREIRA, Vinicius Gomes. O uso de descritores em artigos científicos na área de educação em saúde. **Revista de Saúde Digital de Tecnologias Educacionais**, Fortaleza, v. 3, n. 1, p. 27-40, ago./dez. 2018.

Available at: [https://www.researchgate.net/profile/Bruno-Gambarato/publication/329964145\\_O\\_uso\\_de\\_descritores\\_em\\_artigos\\_cientificos\\_na\\_area\\_d\\_e\\_educacao\\_em\\_saude/links/5c2606ffa6fdccfc706d4c8c/O-uso-de-descritores-em-artigos-cientificos-na-area-de-educacao-em-saude.pdf](https://www.researchgate.net/profile/Bruno-Gambarato/publication/329964145_O_uso_de_descritores_em_artigos_cientificos_na_area_d_e_educacao_em_saude/links/5c2606ffa6fdccfc706d4c8c/O-uso-de-descritores-em-artigos-cientificos-na-area-de-educacao-em-saude.pdf). Access on: 23 maio 2021.

WERSIG, Gernot. Information Science: the study of postmodern knowledge usage. **Information Processing & Management**, Oxford, v. 29, n. 2, p. 229-239, 1993.



Article submitted in the similarity system

Submitted: 16/04/2021 – Accepted: 20/05/2021 – Published: 14/06/2021

---