ABSTRACT

Objective: to analyze the global development of scientific publications and build topics covered in research on the new coronavirus/COVID-19. Methods: review of abstracts of publications on the new coronavirus/COVID-19 with search in the Embase and Web of Science databases, on November 3, 2020. Data were analyzed by modeling the research topics using Latent Dirichlet Allocation (LDA) method. Results: a total of 31,769 publications were analyzed, involving researchers from 182 countries. The studies were published in 5,336 different journals. The analysis identified 20 main research topics on the new coronavirus/COVID-19, with more than half (54.5%) covering topics of impacts and challenges resulting from the pandemic (11.3%), immune response (9.6%), potential therapies (9.0%), epidemiology (7.8%), mental health (6.3%), recommendations for assistance with other health needs (5.3%), and clinical complications. The theme of telemedicine and telehealth was present in 4.2% of publications. Conclusions: scientific publications were distributed among 20 main themes, characterizing their importance and relevance in understanding the disease in question to outline strategies for control, prevention, treatment and attention to diseases/implications resulting from the pandemic. Studies on primary care measures/primary health care were identified as the main knowledge gap.

Keywords: Coronavirus infections. Pandemics. COVID-19. Review.

INTRODUCTION

Viral infections have been devastating over the years. In 2003, the severe epidemic of acute respiratory syndrome (SARS) affected about 8,000 people, killed 780 and caused a huge social and economic crisis. In 2006, a new H5N1 avian influenza and in 2009 a new swine-origin H1N1 influenza pandemic caused global anxiety(1). Situations such as polio (2014), Ebola in West Africa (2014), Zika (2016), and Ebola in the Democratic Republic of Congo (2019) were considered a public health emergency of international interest by the World Health Organization (WHO)(2).

However, none of the above took the current proportion named by the WHO as 2019-nCoV: coronavirus disease (COVID-19)(2), which in late December 2019 caused an outbreak of pneumonia of unknown etiology in Wuhan, Hubei, China, and spread rapidly across the country(3). The Chinese Center for Disease Control and Prevention identified a new beta-coronavirus called 2019-nCoV, now officially known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)(3), responsible for the current pandemic.

This was the third outbreak of zoonotic coronavirus in the first two decades of the 21st century, allowing human-to-human transmission and increasing global health concerns(4), as more than 46 million cases and more than 1.2 million deaths had been reported until November 3, 2020(5).

The pandemic requires actions on many fronts in the near future, through several strategies ranging from prevention, strengthening of early diagnosis, and treatment with effective drugs and vaccines capable of neutralizing the virus. In that sense, reason and science must be the guiding rule. However, as the pandemic increases and the virus becomes...
endemic, there is a need for a structured and well-coordinated approach in various sectors, such as prevention of contagion and complications of COVID-19 in special populations, such as pregnant women and chronically ill people; in the social sphere, with the increase in poverty and hunger; and in the economic sphere, forcing to rethink lifestyles(6). Thus, in order to tackle health issues, it is necessary and urgent to critically assess the available evidence.

COVID-19 has become a global crisis in terms of health, information, and is becoming a socioeconomic crisis of disproportionate extent(5). For this reason, public health authorities must continue to monitor the situation and constant investments must be made in research, considering that the more knowledge produced about this new virus and its associated outbreaks, the more effective will be the response capacity(2).

The urgency in combating the pandemic and the management of other problems/grievances resulting from it requires the search for the best scientific evidence in a summarized and updated manner, so as to assist in decision making. This need has generated a significant number of studies in the area and their publication in scientific journals from different countries, which have fostered the dissemination of research on the theme.

Due to the large volume of publications on the new coronavirus, the following question raises: what topics are most addressed and which are the possible knowledge gaps about the new coronavirus? In order to answer and highlight the most important aspects to be considered in studies on the topic, the objective of the present work was to analyze the global development of scientific publications and discuss the topics approached in extant research on the new coronavirus/COVID-19.

METHOD

Data set

A scoping review was carried out in the Embase and Web of Science databases. The choice of these bases is justified due to their global scope and because they allow the collection of a large volume of data in .csv format, enabling the analysis proposed by the present study.


Abstracts of studies published in the years 2019 and 2020, in English, as articles or reviews were included in the present review in order to cover original research and reviews, as well as to limit the inclusion of letters, editorials, notes, among others. Studies that had the fields “author’s address” and “abstract” empty were excluded, and also those duplicated in the databases.

Data were downloaded in .csv format, including the following information: title of the publication; authors; journal name; year of publication; key words; author’s address; and abstract. Two researchers worked independently to download the data simultaneously. The linkage of databases was performed and data were analyzed using the Python language (version 3.7.5), via the Jupyter Notebook platform.

Data analysis

All articles that met the eligibility criteria were first descriptively analyzed to identify the countries with the largest number of researchers participating in the publications and the journals that most published on the topic. The number of researchers by country was identified using the author’s address field. The Pycountry (https://pypi.org/project/pycountry/) and Plotly (https://plotly.com/) libraries, available in the Python language (version 3.7.5), were used to this end.

Regarding the analysis of topics, the study went through the stages of pre-processing of data with the processes of cleaning, transformation and modeling of topics.

Pre-processing of the text was employed in order to organize, clean and standardize the
textual data to be consumed by the processing algorithm. The purpose of this procedure is to remove unnecessary content to reduce processing time and increase the performance of the models(7). The techniques used for the pre-processing of the data set were the removal of special characters, conversion of uppercase to lowercase letters, removal of words without semantic value (stopwords), removal of other unnecessary terms and application of tokenization, which breaks the text in words. These tasks were performed using the Python language (version 3.7.5) with support from the Natural Language Toolkit (NLTK) library (https://www.nltk.org/).

Topic modeling was used in order to find out which themes have most mobilized scientists. For this, the Latent Dirichlet Allocation (LDA) method was used. This is an algorithm that automatically detects topics in texts that belongs to the class of unsupervised methods. The LDA algorithm is based on the idea that documents are composed of a mixture of topics and seeks to discover latent (hidden) patterns to understand the relationships between documents and words, so that the words that occur in related documents are grouped into topics(8). The LDA method was implemented using the Python Machine Learning for Language Toolkit (MALLET) language pack (http://mallet.cs.umass.edu/).

Data were transformed into vectors for later use in the algorithms using the Bag of Words (BoW) attribute engineering method. The BoW model represents each text document as a numerical vector in which each dimension is a specific word in the corpus and the value corresponds to its frequency in the document(7).

The model was also adjusted to remove terms with less than 10 occurrences in all documents and terms that occurred in more than 60% of documents, reducing 36,672 to 8,239 terms. This practice increases the accuracy of the model and reduces computational effort, since unique or rare terms have a low semantic contribution to the model and recurring terms interfere in the interpretation of the context(7).

The optimal number of topics was defined based on the coherence score. To this end, models were generated with a number of topics ranging from 2 to 50. Figure 1 shows the variation of the coherence score according to the number of topics, showing that the greatest coherence (0.642) occurred with 33 topics, being this is the value adopted for this study. Thus, all sentences in the data set were classified into any of the 29 topics revealed by the LDA, according to the highest probability of belonging to each of the topics.

**Figure 1.** Variation of model coherence according to the number of topics (November 3, 2020)

For the identification of the themes approached in the research, the topics were manually labeled based on the combination of human judgment with reference to the literature on the theme of the research, so that those with similarities were pooled. Three topics that did not have word consistency were excluded from the labeling, leaving 20 research topics at the end of the analysis.

**RESULTS**

The initial research identified 57,389 publications, 34,498 from Embase and 22,888 from the Web of Science. There were 7,940 duplicate publications and 17,677 were excluded for not having the “author's address” and/or “abstract” fields filled. At the end, 31,769 abstracts published in 5,336 different journals were analyzed (Figure 2).
Figure 2. Flowchart of the study selection process (November 3, 2020)

Figure 3 shows the distribution of researchers according to the 182 countries involved in studies on the analyzed theme. The country with the largest number of researchers involved was the United States (19.6%), followed by China (10.6%), Italy (7.4%), the United Kingdom (5.9%), and India (5.7%). Brazil ranked in the 11th position, with 2.2% of research.

Figure 3. Distribution of the number of researchers according to the country (November 3, 2020)

Table 1 presents the 20 topics approached in scientific research built by the LDA. Considering the exclusion of three topics that could not be labeled due to the inconsistency of the word set, composed of 1,584 (4.9%) analyzed abstracts, the research topics of 30,185 articles were presented.

More than half (54.5%) of the publications were included in the first seven topics of analysis. The most addressed topic was the impacts and challenges resulting from the pandemic (11.3%), followed important topics for the development of prevention, diagnosis and treatment mechanisms (Topics 2 and 3) and the epidemiological aspects involved (7.8%). As for the specific areas of health care, it was found that mental health (Topic 5) was the most addressed (6.3%). The topic of telemedicine and telehealth (Topic 11) also stood out, evidenced in 4.2% of the analyzed publications.
Modeling of research topics on the new coronavirus: application of Latent Dirichlet Allocation

Table 1. Topics addressed in research on the new coronavirus (November 3, 2020)

<table>
<thead>
<tr>
<th>LDA topic</th>
<th>N</th>
<th>%</th>
<th>Topic label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3421</td>
<td>11.3</td>
<td>Impacts and challenges</td>
</tr>
<tr>
<td>2</td>
<td>2892</td>
<td>9.6</td>
<td>Immune response</td>
</tr>
<tr>
<td>3</td>
<td>2730</td>
<td>9.0</td>
<td>Therapeutic potentials</td>
</tr>
<tr>
<td>4</td>
<td>2357</td>
<td>7.8</td>
<td>Epidemiology</td>
</tr>
<tr>
<td>5</td>
<td>1894</td>
<td>6.3</td>
<td>Mental health</td>
</tr>
<tr>
<td>6</td>
<td>1603</td>
<td>5.3</td>
<td>Recommendations for comprehensive health care in times of COVID-19</td>
</tr>
<tr>
<td>7</td>
<td>1570</td>
<td>5.2</td>
<td>Clinical complications of COVID-19</td>
</tr>
<tr>
<td>8</td>
<td>1346</td>
<td>4.5</td>
<td>Clinical manifestations</td>
</tr>
<tr>
<td>9</td>
<td>1304</td>
<td>4.3</td>
<td>Safety of healthcare workers and patients</td>
</tr>
<tr>
<td>10</td>
<td>1300</td>
<td>4.3</td>
<td>Pulmonary evaluation</td>
</tr>
<tr>
<td>11</td>
<td>1270</td>
<td>4.2</td>
<td>Use of telemedicine and telehealth</td>
</tr>
<tr>
<td>12</td>
<td>1185</td>
<td>3.9</td>
<td>Training of health professionals</td>
</tr>
<tr>
<td>13</td>
<td>1186</td>
<td>3.9</td>
<td>Hospitalization and intensive care</td>
</tr>
<tr>
<td>14</td>
<td>1139</td>
<td>3.8</td>
<td>Associated comorbidities</td>
</tr>
<tr>
<td>15</td>
<td>1061</td>
<td>3.5</td>
<td>Mechanism of virus action</td>
</tr>
<tr>
<td>16</td>
<td>1048</td>
<td>3.5</td>
<td>Vaccine development</td>
</tr>
<tr>
<td>17</td>
<td>924</td>
<td>3.1</td>
<td>Attention to the health of pregnant women</td>
</tr>
<tr>
<td>18</td>
<td>836</td>
<td>2.8</td>
<td>Government strategies to control virus transmission</td>
</tr>
<tr>
<td>19</td>
<td>627</td>
<td>2.1</td>
<td>Attention to children’s health</td>
</tr>
<tr>
<td>20</td>
<td>492</td>
<td>1.6</td>
<td>Population awareness</td>
</tr>
</tbody>
</table>

Box 1 presents a descriptive synthesis of the topics approached in the studies in each of the 20 topics analyzed.

Box 1. Descriptive synthesis of the topics (November 3, 2020)

<table>
<thead>
<tr>
<th>Topic</th>
<th>Includes studies dealing with:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Impacts and challenges, in the short and long term, in the systems and different sectors, namely, health, socioeconomic, environmental, specific health/disease conditions/situations (for example: hemodialysis, mental health, emergency situations) caused by the new coronavirus/COVID-19 and/or changes resulting from the pandemic.</td>
</tr>
<tr>
<td>2</td>
<td>Analysis of the immune response to the new coronavirus for disease control and treatment, and drug and vaccine development.</td>
</tr>
<tr>
<td>3</td>
<td>Development and testing of drugs for prevention and treatment.</td>
</tr>
<tr>
<td>4</td>
<td>Epidemiological aspects of the emergence of the new coronavirus, scenario of dissemination/growth, incidence, transmission and mortality rates, modeling of possible scenarios.</td>
</tr>
<tr>
<td>5</td>
<td>Psychological changes and development of stress, anxiety and/or depression in the population and in health professionals; implications for patients with pre-existing mental health problems/diseases.</td>
</tr>
<tr>
<td>6</td>
<td>Clinical guidelines for health care in acute or chronic situations during the pandemic, such as cancer, urgent/emergency care, surgical procedures.</td>
</tr>
<tr>
<td>7</td>
<td>Complications in organs (kidney, liver, heart...) and systems in patients with or without previous pathologies.</td>
</tr>
<tr>
<td>8</td>
<td>Symptoms and clinical manifestations in people with the new coronavirus.</td>
</tr>
<tr>
<td>9</td>
<td>Need and quality of personal protective equipment in different health sectors and establishment of protocols for emergency and elective surgeries in patients not contaminated by the new coronavirus, for the protection and safety of health professionals and patients.</td>
</tr>
<tr>
<td>10</td>
<td>Evaluation of the characteristics and findings of chest imaging exams in patients with COVID-19.</td>
</tr>
<tr>
<td>11</td>
<td>Use of telemedicine and telehealth in different scenarios/conditions as a means of ensuring access to health care and preventing contamination</td>
</tr>
<tr>
<td>12</td>
<td>Training and distance learning programs for health professionals in a pandemic period, mainly for medicine and nursing.</td>
</tr>
<tr>
<td>13</td>
<td>Challenges and need of hospitals and intensive care units; clinical guidelines for intensive care; characteristics and prognosis of hospitalized and/or intensive care patients.</td>
</tr>
<tr>
<td>14</td>
<td>Analysis of prevalence, outcomes, risk and severity of COVID-19 in patients with associated comorbidities.</td>
</tr>
<tr>
<td>15</td>
<td>Knowledge about the mechanisms of action of the new coronavirus and its relationship with clinical characteristics and development of prevention and treatment therapies.</td>
</tr>
</tbody>
</table>

To be continued…
DISCUSSION

The search for scientific publications about the new coronavirus and the modeling of research topics showed a high number of publications at a global level, with greater participation of researchers of the United States and China, which are two main world economic powers today. China was the country where the pandemic started and the United States represents the country with the highest number of infected and dead people, factors that may be motivating investments in research.

Publications were distributed in journals from different areas, which indicates the concern of different fields to understand and find solutions to the problems arising from the pandemic in different sectors.

Twenty main research topics were identified. Most refer to impacts and challenges, considering that COVID-19 has emerged as a threat to global health, and each nation has faced unique challenges. The pandemic led to the need to restructure health systems to provide timely care for people with COVID-19, affecting, on the other hand, patients who require assistance for other conditions, generating delays in care and treatment, as well as the need for reorganization of services to serve different specialties\(^9\)–\(^{13}\).

The pandemic has also caused intense economic impacts due to the need for emergency resources and a drop in tax collection\(^{14}\), in addition to revealing the close interconnections between social/economic inequality, indicating that policies that promote social solidarity may also improve the health of the population and economic performance for better coping with future emergency situations\(^9\)–\(^{13}\).

The approach of themes that involve impacts and challenges in the face of the pandemic expresses the concern to seek normality for the unprecedented deconstruction of society. Global epidemics are not new, but the COVID-19 epidemic happens to take place in an increasingly complex, technological and connected world, in which risk events mean different and unexpected ways of coping\(^{16}\).

Studies on the new coronavirus have also relevantly addressed the epidemiological aspects of the pandemic, since its emergence to global spread. They have sought to understand the differences and similarities in transmission, incidence and mortality rates, so as to assist in the development of coping strategies, such as government (Topic 18) and population awareness (Topic 20).

In this topic (Topic 4), the importance of developing predictive models for infectious diseases is emphasized, enabling early alertness and significantly improving the capacity for surveillance, control and prevention of outbreaks, particularly in developing regions\(^{17}\).

Epidemiological information helps to define, contain, prevent and successfully treat the impact of infectious diseases. The use of epidemiology and its framework makes it possible to support hypotheses that provide understanding to contain or eliminate the disease\(^{18}\). The challenge of epidemiology is great because government responses or public health measures to manage the pandemic may be ineffective as problems occur in various sectors of society and are manifested in unknown ways\(^{16}\).

Another important issue highlighted in the topics of this study is that, to date, there is no specific antiviral treatment recommended for COVID-19 and the therapeutic strategies to deal with the infection are only supportive\(^{19}\). Many attempts are being made to develop vaccines against SARS-CoV-2 (Topic 16). Thus, due to the strong genetic similarities with the SARS-CoV epidemic in 2003, there are investigations on the reuse of approved drugs, such as antivirals, antimalarials and immunomodulators\(^{20}\).

Understanding the immune response is
essential for the development of potential therapies. Thus, these themes were widely addressed in 2 of the 20 topics analyzed in the present study, emphasizing their importance and relevance for the control, treatment and prevention of COVID-19.

In addition to efforts to understand the epidemiology, clinical characteristics, transmission and management of COVID-19, the theme of mental health was the most common health problem addressed in the publications. In the pandemic period, people need to deal with issues such as social isolation, detachment, death, xenophobia and other situations that have generated psychological consequences such as stress, anxiety, fear and depression in the general population and in different age groups and social classes\(^{(21,22)}\), as well as in vulnerable groups and/or with pre-existing mental health problems\(^{(23,24)}\). Still, there is considerable concern in the studies regarding psychological disorders in health professionals, especially those who work on the front line to combat COVID-19\(^{(25)}\).

As a way to overcome some challenges in accessing health care, ensuring continuity of care and reducing contact between patients and health professionals, the topic of telemedicine and telehealth was another high point of the publications analyzed. These strategies have been widely used in different health specialties with discussions about the quality and effectiveness of this type of assistance\(^{(26–28)}\).

Within the scope of COVID-19, telemedicine emerged as a tool for screening the population, so that remote monitoring has specific operational resources that are suitable for this condition, especially asynchronous communication, which can be used to collect data about the pandemic and obtain real-time clinical feedback\(^{(29)}\).

In addition to the topics discussed, the analysis using the LDA method allowed to identify other topics approached in the researches about COVID-19, including the development of clinical guidelines for care in elective or urgent/emergency situations, personal protective equipment, government strategies and awareness of population to control the transmission, clinical aspects, and health care for pregnant women and children. This ample scope of research becomes dynamic and essential for the understanding of many aspects affected by COVID-19.

Despite the wide variety of themes, there was a gap in relation to the topic of primary health care (PHC). This topic has been researched, but not so intensely as to appear as a dominant topic in the analysis by LDA, which leads to question whether the evidence produced so far is sufficient to support actions in this field.

Even though health systems in many countries are not structured through PHC, strengthening this level of care is the most effective policy response, especially in low and middle income countries. This would increase the protection in health emergencies and contribute to achieve universal health coverage and promote health and well-being\(^{(30)}\). This role of PHC has been evident during the SARS-CoV-2019 pandemic, in which strong PHC systems have contributed a more quick and effective response in some countries, achieving better health outcomes for their populations\(^{(31)}\).

Even more challenging may be to develop research involving issues not considered a priority for the escalation of contagion of the virus. Beyond learning about therapeutic strategies, social and environmental research can be important to demonstrate the adaptation or lack of it in the populations’ way of survival as the pandemic increases, in the sense of understanding the social dynamics of groups, as in favelas, in the countryside, in small cities, facing all stages of the pandemic, even those not yet detected or discovered.

Although attempts have been made to develop a comprehensive search strategy, it is possible that some relevant studies have been lost. Also, materials and research that were not published in journals indexed in the Web of Science and Embase databases were not included, and they could have provided some additional information. Thus, this represents a limitation of the study.

**CONCLUSION**

The present study analyzed the global development of scientific publications and build topics approached in research on the new coronavirus/COVID-19. There were many
publications in different journals, carried out by researchers from 182 countries. The main themes that have been addressed were identified, characterizing their importance and relevance in understanding the disease in question to outline strategies for control, prevention, treatment and attention to the health problems/implications resulting from the pandemic.

More studies on the new coronavirus/COVID-19 are recommended, especially in less contemplated areas such as PHC.

MODELAGEM DE TÓPICOS DE PESQUISA SOBRE O NOVO CORONAVÍRUS: APLICAÇÃO DO LATENT DIRICHLET ALLOCATION

RESUMO

Objetivo: analisar o desenvolvimento global de publicações científicas e construir tópicos abordados em pesquisas sobre o novo coronavírus/COVID-19. Métodos: estudo de revisão de resumos de publicações sobre o novo coronavírus/COVID-19, com busca nas bases de dados Embase e Web of Science, em 03 de novembro de 2020. A análise dos dados se deu pela modelagem dos tópicos de pesquisa pelo método Latent Dirichlet Allocation (LDA). Resultados: analisaram-se 31.769 publicações, com envolvimento de pesquisadores de 182 países. Os estudos foram publicados em 5.336 periódicos diferentes. A análise identificou 20 tópicos principais de pesquisa sobre o novo coronavírus/COVID-19, sendo que mais da metade (54,5%) das publicações foram contempladas nos temas de impactos e desafios decorrentes da pandemia (11,3%), resposta imunológica (9,6%), potenciais terapêuticas (9,0%), epidemiologia (7,8%), saúde mental (6,3%), recomendações para assistência a outras necessidades de saúde (5,3%) e complicações clínicas. O tema da telemedicina e telessaúde foi evidenciado em 4,2% das publicações. Conclusões: as publicações científicas estavam distribuídas dentre 20 principais temas, caracterizando sua importância e relevância no entendimento do agravo em questão para traçar estratégias de controle, prevenção, tratamento e atenção aos agravos/implicações decorrentes da pandemia. Identificaram-se como principal lacuna de conhecimento estudos sobre cuidados primários/atenção primária à saúde.


MODELADO DE TÓPICOS DE INVESTIGACIÓN SOBRE EL NUEVO CORONAVÍRUS: APLICACIÓN DELLATENT DIRICHLET ALLOCATION

RESUMEN

Objetivo: analizar el desarrollo global de publicaciones científicas y construir tópicos tratados en investigaciones sobre el nuevo coronavirus/COVID-19. Métodos: estudio de revisión de resúmenes de publicaciones sobre el nuevo coronavirus/COVID-19, con busca en las bases de datos Embase Web of Science, el 03 de noviembre de 2020. El análisis de los datos se llevó a cabo por el modelo de tópicos de investigación por el método Latent Dirichlet Allocation (LDA). Resultados: se analizaron 31.769 publicaciones, con envolvimiento de investigadores de 182 países. Los estudios fueron publicados en 5.336 periódicos diferentes. El análisis identificó 20 tópicos principales de investigación sobre el nuevo coronavirus/COVID-19, más de la mitad (54,5%) de las publicaciones fue contemplada en los temas de impactos y desafíos resultantes de la pandemia (11,3%), respuesta inmunológica (9,6%), potenciales terapéuticos (9,0%), epidemiología (7,8%), salud mental (6,3%), recomendaciones para asistencia a otras necesidades de salud (5,3%) y complicaciones clínicas. El tema de la telemedicina y telesalud fue evidenciado en 4,2% de las publicaciones. Conclusiones: las publicaciones científicas estaban distribuidas entre 20 principales temas, caracterizando su importancia y relevancia en el entendimiento del agravio encuestón para esbozar estrategias de control, prevención, tratamiento y atención a los agravios/implicações resultantes de la pandemia. Se identificaron como principal lacuna de conocimiento estudios sobre cuidados primarios/atenção primária à saúde.


REFERENCES


Cinc Cuid Saude 2021;20:e56403
Vivências acerca da hospitalização: percepções de gestantes de alto risco

https://doi.org/10.4081/jphr.2020.1786

Corresponding author: Tatiane Baratieri. Rua Alameda Elio Antonio Dalla Vecchia, 838. Departamento de Enfermagem. CEP 85040-167. Bairro Vila Carli, Guarapuava/PR. E-mail: baratieri.tatiane@gmail.com

Submitted: 26/10/2020
Accepted: 19/12/2020

Cienc Cuid Saude 2021;20:e56403