Risk Factors for COVID-19: A Systematic Mapping Study

Paula A. ROSERO ①, Juan S. REALPE ①, Charic D. FARINANGO ①, David S. RESTREPO ①, Ricardo SALAZAR-CABRERA ① and Diego M. LOPEZ ①

① Telematics Department, University of Cauca, Popayan, Colombia

Abstract. Introduction: COVID-19 has affected people in several countries around the world. They experience respiratory symptoms that can be mild, moderate, or severe. Several reviews that characterize the risk factors of COVID-19 have been performed, but most address only risk factors associated with medical conditions, ignoring environmental and sociodemographic-socioeconomic factors. Objective: This study aims at characterizing different risk factors in the published literature that influence contagion by COVID-19. Methods: The review consists of three stages, including a systematic mapping with studies found in the Scopus database, an analysis of results, and finally the identification of relevant COVID-19 risk factors. Results: A map of studies is provided considering two main groups: the type of research and context. Most studies consider risk factors associated with medical conditions, while research on other factors is scarce. Conclusions: Medical conditions such as diabetes, obesity, cardiovascular disease, hypertension, and factors such as age and sex, appear to be the ones that increase the risk of contracting COVID-19. Further research is needed on environmental, sociodemographic, and socioeconomic risk factors.

Keywords. Covid-19, environment, health, risk factors, sociodemographic, socioeconomic

Introduction

COVID-19 is the disease caused by a new coronavirus called SARS-CoV-2. It was first learned on December 31, 2019, following a report of a cluster of ‘viral pneumonia’ cases in Wuhan, People’s Republic of China [1]. It was characterized as a Public Health Emergency of International Concern (PHEIC) by the World Health Organization (WHO) on January 30, 2020 [2]. On March 11, 2020, the WHO characterized COVID-19 as a pandemic [3].

Most infected people have respiratory symptoms which can be mild, moderate, or severe. Other symptoms vary from person to person depending on different factors. It has also been reported that older people and those with underlying medical conditions are more likely to present severe afflictions [4].

Although possible symptoms have been described and the vaccines have reduced the mortality rate, the long-term effects are still unknown. Furthermore, due to COVID-19’s novelty, long-term studies are not possible. For this reason, and because of the population at elevated risk, COVID-19 is still a serious public health issue.

① Corresponding Author. Paula Rosero, University of Cauca, Popayan, Colombia; E-mail: parosero@unicauca.edu.co
Confirmed testing rates for COVID-19 provide critical information to understand its full impact and identify ways to decrease morbidity and mortality. Nevertheless, a lack of testing, especially in low/middle-income countries, impacts the detection of its prevalence. In addition, hospitalization rates offer an indirect measure of local outbreaks and mortality. Still, these rates are generally affected by sociodemographic and socioeconomic factors such as insufficient funds or limited knowledge about COVID-19.

All the previous factors add up to the necessity of identifying risk factors that aid healthcare entities in making decisions to reduce COVID-19 cases.

The motivation of this study is to characterize the distinct types of risk factors present in the published literature that influence contagion by COVID-19. Accordingly, the authors performed the following stages: a) a systematic mapping, b) an analysis of results, and c) identifying relevant factors.

1. Related Work

Due to the profound impact that the COVID-19 pandemic has caused worldwide, many studies in the literature characterize risk factors, most of which address only one type of risk factor. We present some of the reviews that have been made.

Reviews [5-18] mention both Human (HM) and Sociodemographic-Socioeconomic (SD-SE) risk factors. Common factors include age, sex, obesity, hypertension, and diabetes.

Environmental (EV) risk factors are referenced in reviews [19-23]. They mention multiple types of factors where air contamination stands out. Unlike HM risk factors, the reporting of these types factors is not extensive.

HM risk factors alone are found more frequently like in [8], [19], [20], [24-44]. The most frequent HM factors are hypertension, diabetes, obesity and cardiovascular diseases.

The SD-SE risk factors found in [8], [25], [33], [35], [38], [39], [41], [42], [45-48] show that age and sex have been cataloged as risk factors that may increase the complications when diagnosed with COVID-19.

Finally, reviews [20], [49-52] are some of the few ones that present the three types of risk factors, HM, SD-SE and EV, at the same time.

Considering the reviews found in the literature, this article aims to characterize the risk factors that affect the morbidity and mortality of COVID-19. We cover all three types of risk factors. Furthermore, this study differs from other reviews that include the three groups of factors without having geographical limitations or focusing on a specific risk factor. The results can be useful for health professionals and public health agencies to implement strategies in response to COVID-19. They can also be useful for any implementation based on technology that aims to support COVID-19 predictions and decisions.

2. Methodology

Our methodology consisted of three stages: 1) Systematic mapping, 2) Analysis of results, and 3) Identifying relevant factors. In particular, the systematic mapping stage was derived from the “Guidelines for performing Systematic Literature Reviews in Software Engineering” [53].
2.1. Systematic Mapping

Systematic mapping is a type of systematic review with a broad scope, designed to cover and provide an overview of a research area, classifying and counting contributions according to predefined categories. It is based on selecting, reading and classifying abstracts through 5 main phases, culminating with a results map.

2.1.1. Scope and Research Question

The research question led this study to discover works that described different risk factors for COVID-19. Therefore, having this goal in mind, the research question for the systematic mapping was: What risk factors are found in the COVID-19 literature?

2.1.2. Information Search

The problem at hand guided the selection of the main search terms. As a result, the chosen keywords generated the search string: "RISK FACTORS" AND "COVID". The next step allowed us to select the Scopus database, considering it has broad scientific research coverage. Finally, the query was applied to the title, abstract, and keywords. Specifically, the selected search string covered a larger number of studies. The search was conducted between August 9 and August 22, 2021, obtaining 1,786 studies. Most of them are presented in English.

2.1.3. Inclusion/Exclusion Criteria

The goal of this section was to define the inclusion/exclusion criteria to filter potentially relevant papers retrieved from the search. Table 1 shows the selected requirements.

<table>
<thead>
<tr>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC1: the paper should include at least one of the risk factors of interest. They are SD-SE, EV, and/or HM.</td>
<td>EC1: Papers were published before 2015 to construct a clear trend of publication</td>
</tr>
<tr>
<td>IC2: the paper is review-type. Other studies were analyzed allowing for a much broader vision of the subject</td>
<td>EC2: papers with a theoretical approximation to the problem</td>
</tr>
<tr>
<td>IC3: the paper has performed a study of the risk factors for COVID-19</td>
<td>EC3: papers do not mention risk factors for COVID-19</td>
</tr>
<tr>
<td>IC4: papers published as notes, editorials, or letters</td>
<td></td>
</tr>
</tbody>
</table>

The authors ensured that the reviewed papers mentioned the problem to be addressed. Furthermore, any duplicate papers were removed. Finally, no restriction was applied regarding the country of publication, country of study, results, or the language of publication.
2.1.4. Classification Scheme

During the initial review, we looked for ways to classify and order the information on the studies. Thus, considering the three identified groups of risk factors, we identified two main classes among the papers. These two classes grouped the type of review and the factors covered. It should be noted that the number of studies in the database increased daily. The classes were:

- "Type of research" It refers to the type of review that had been conducted. It was a result of scanning over the abstracts.
- "Research context" It refers to the types of factors found and their combinations.
  It was generated once the papers were organized and reviewed more deeply.

2.1.5. Data Extraction Strategy

To standardize the data extraction and reduce the bias of results and the informality of the process, we created a spreadsheet to follow during the examination. This template was adapted from [53]. First, authors in charge of examining the papers extracted the title of the document, authors, year of publication, type of research, research context and summary; this last information was compiled in a document containing the identified risk factors by each document. The resulting document is in Spanish and can be found in [54].

2.2. Analysis of Results

After analyzing the selected documents, we sorted and characterized each of them. The classification allowed us to identify the risk factors present in the papers. Extracting this knowledge gave us a sense of the importance of different risk factors for COVID-19 and the papers that have been made around them.

2.3. Identification of Relevant Factors

After structuring the data from the studies selected, we identified relevant factors, which are the most frequent factors mentioned in the studies. Initially, from the factors found in the review, we established a standard writing format so that similar terms were written equally. Then, the factors that were mentioned more than once were entered into the template. Finally, a function was applied to count the times that each factor was determined as a risk factor in all the studies. A table was created for each type of risk factor. Their combinations were not considered because the purpose was to establish how many times a factor was classified as a risk factor.

3. Results

From the 1,786 returned documents, 564 were related to the objective of the work and met the inclusion and exclusion criteria. Table 2 shows the derived classification scheme.
Table 2. Classification scheme

<table>
<thead>
<tr>
<th>Classification “Research context”</th>
<th>Classification “Type of research”</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Environmental factors.” These factors involve environmental variations</td>
<td>“Review and experimentation” Reviews that perform their own studies</td>
</tr>
<tr>
<td>“Human factors.” These factors are related to people’s health conditions</td>
<td>“Review” Documents that reach the goal of a review but did not perform their own studies</td>
</tr>
<tr>
<td>“Sociodemographic and Socioeconomic factors.” These factors involve the characterization of people</td>
<td>“Non-formal review” The document’s abstract did not indicate if the study considered databases or the number of documents that they examined</td>
</tr>
<tr>
<td>“Environmental and Human factors”</td>
<td>“Sociodemographic and Socioeconomic, and Environmental factors”</td>
</tr>
<tr>
<td>“Sociodemographic and Socioeconomic, and Human factors”</td>
<td>“Sociodemographic and Socioeconomic, Environmental and Human factors”</td>
</tr>
</tbody>
</table>

The main result was the map of studies shown in Figure 1, which relates the classes and indicates the number of articles found for each of them with their proportions.

![Figure 1](image-url)

Figure 1. Bubble graph mapping and associating the type of research with the research context. The percentages are calculated per each axis.

3.1. Analysis

The map shows that the largest number of documents corresponds “review” type and only addresses HM risk factors. The second largest number of papers, also being of “review” type, study HM and SD-SE factors. On the opposite side, the “review and experimentation” documents have the least dominance in the mapping.
The papers that address EV risk factors have the least representation among all combinations. In addition, among all categories of “type of research,” very few results were obtained for documents that addressed the category of EV and HM risk factors.

“Non-formal review” type has lower dominance in the mapping, and in terms of the research context, HM risk factors were the most prevalent for this category.

Finally, it should be noted that the risk factors with the most presence in the documents were comorbidities such as diabetes, hypertension, obesity, and cardiovascular diseases, as well as factors such as age and sex.

3.2. Relevant Risk Factors

The risk factors for COVID-19 were classified into the three initially proposed groups; HM risk factors, EV risk factors, and SD-SE risk factors.

Human risk factors. Table 3 shows that HM risk factors have the greatest presence in the studies, where diseases related to mental and physical health were found. Some diseases with the greatest presence were diabetes, obesity, cardiovascular and hypertension. Some documents that cover a large number of these factors were: [8], [9], [37], [41], [42], [44], [50], [55-64].

<table>
<thead>
<tr>
<th>Human Risk Factor</th>
<th>Count</th>
<th>Human Risk Factor</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes</td>
<td>107</td>
<td>Tuberculosis</td>
<td>6</td>
</tr>
<tr>
<td>Obesity</td>
<td>102</td>
<td>Coronary</td>
<td>6</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>90</td>
<td>EPOC</td>
<td>6</td>
</tr>
<tr>
<td>Hypertension</td>
<td>82</td>
<td>Genetic</td>
<td>5</td>
</tr>
<tr>
<td>Chronic disease</td>
<td>30</td>
<td>Cytokine storm</td>
<td>5</td>
</tr>
<tr>
<td>Cancer</td>
<td>26</td>
<td>Depression</td>
<td>4</td>
</tr>
<tr>
<td>Smoking</td>
<td>24</td>
<td>Alcoholism</td>
<td>4</td>
</tr>
<tr>
<td>Pulmonary</td>
<td>24</td>
<td>Sleep</td>
<td>3</td>
</tr>
<tr>
<td>Renal</td>
<td>24</td>
<td>Physical inactivity</td>
<td>3</td>
</tr>
<tr>
<td>Fatty liver</td>
<td>23</td>
<td>Anxiety</td>
<td>2</td>
</tr>
<tr>
<td>Respiratory</td>
<td>21</td>
<td>Epilepsy</td>
<td>2</td>
</tr>
<tr>
<td>Mental</td>
<td>18</td>
<td>Malnutrition</td>
<td>2</td>
</tr>
<tr>
<td>Cerebrovascular</td>
<td>16</td>
<td>Sclerosis</td>
<td>2</td>
</tr>
<tr>
<td>Asthma</td>
<td>15</td>
<td>Stress</td>
<td>1</td>
</tr>
<tr>
<td>HIV</td>
<td>10</td>
<td>Dementia</td>
<td>1</td>
</tr>
<tr>
<td>Inflammatory</td>
<td>10</td>
<td>Alzheimer</td>
<td>1</td>
</tr>
<tr>
<td>Blood</td>
<td>9</td>
<td>Addiction</td>
<td>1</td>
</tr>
<tr>
<td>Vitamins</td>
<td>9</td>
<td>Diet</td>
<td>1</td>
</tr>
<tr>
<td>Heart disease</td>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Environmental risk factors. Table 4 shows the EV risk factors, which are the ones that have the lowest representation among the studies. The EV factors with the greatest presence were the environment, air quality and pollution. Some documents that cover at least two of these three factors were: [21], [22], [50], [65-67].
Table 4. Number of times that Environmental risk factors were mentioned in the studies

<table>
<thead>
<tr>
<th>Environmental risk factors</th>
<th>Count</th>
<th>Environmental risk factors</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment</td>
<td>16</td>
<td>Particulate Matter</td>
<td>4</td>
</tr>
<tr>
<td>Air</td>
<td>14</td>
<td>Surface</td>
<td>3</td>
</tr>
<tr>
<td>Contamination</td>
<td>11</td>
<td>Humidity</td>
<td>3</td>
</tr>
<tr>
<td>Aerosol</td>
<td>10</td>
<td>Wastewater</td>
<td>3</td>
</tr>
<tr>
<td>Weather</td>
<td>5</td>
<td>Meteorological Phenomena</td>
<td>2</td>
</tr>
<tr>
<td>Respiratory droplets</td>
<td>4</td>
<td>NO2</td>
<td>1</td>
</tr>
<tr>
<td>Temperature</td>
<td>4</td>
<td>Solar light</td>
<td>1</td>
</tr>
<tr>
<td>Radiation</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sociodemographic and Socioeconomic risk factors. Table 5 presents the found SD-SE risk factors. Age is the most mentioned factor in the documents, followed by sex and race. The documents covering at least two of the three factors were: [8], [35], [41], [44], [59], [64], [68-76].

Table 5. Number of times that Sociodemographic and Socioeconomic risk factors were mentioned in the documents

<table>
<thead>
<tr>
<th>Sociodemographic and Socioeconomic risk factors</th>
<th>Count</th>
<th>Sociodemographic and Socioeconomic risk factors</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>399</td>
<td>Education</td>
<td>5</td>
</tr>
<tr>
<td>Sex</td>
<td>81</td>
<td>Lifestyle</td>
<td>3</td>
</tr>
<tr>
<td>Race</td>
<td>27</td>
<td>Public health</td>
<td>2</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>16</td>
<td>Geographic</td>
<td>2</td>
</tr>
<tr>
<td>Economic</td>
<td>16</td>
<td>Transport</td>
<td>2</td>
</tr>
<tr>
<td>Job</td>
<td>14</td>
<td>Immigrants</td>
<td>1</td>
</tr>
<tr>
<td>Income</td>
<td>10</td>
<td>Social networks</td>
<td>1</td>
</tr>
<tr>
<td>Housing</td>
<td>5</td>
<td>Overcrowding</td>
<td>1</td>
</tr>
</tbody>
</table>

4. Discussion

It was found that most of the publications have been made from 2020, the year in which the disease was considered a pandemic. Figure 2 shows the number of studies published per type of risk factor’s group since the pandemic started.

Figure 2. Number of studies per type of context published in each quarter since the pandemic started.
Most of the studies around COVID-19 risk factors were published in the first two quarters of 2021, reflecting the great effort made in 2020 to understand the new disease. This temporality, however, also show the low amount of research on risk factors in recent quarters. Furthermore, this decreasing trend contributes to the lack of exploration of other risk factors, such as the EV type.

From the included studies, we determined three groups of risk factors: HM, SD-SE, and EV. Some studies mention elements belonging to more than one group, so possible combinations between them were also considered. Most identified risk factors belonged to the HM type, followed by SD-SE. Few studies have identified EV risk factors, which the recent appearance of the disease can explain.

As stated before, the most frequent factors belong to the HM class. The possible reason is that it contains clinical factors, the most common ways to diagnose a disease. In general, for the characterized HM risk factors, it was determined that they could be classified into two large groups; factors related to mental health such as depression, stress, and sleep disorders, among others; and factors that refer to the diseases present in people who have been infected with COVID-19. On the other hand, EV risk factors were the least studied and, among the outstanding factors are the environment, air quality and pollution. Finally, for SD-SE factors, we identified factors such as: age, sex, race, ethnicity, economic income, and type of work.

Regarding the practical and research implications, this review makes it possible to explore the risk factors of COVID-19, which would allow an in-depth study and thus help health organizations in their decision-making process. In addition, this study identifies factors that generally increase vulnerability to other diseases. Moreover, the study allows us to identify gaps for future research, such as that a few documents mention the three types of risk factors.

In addition, thanks to the map of the studies, it is recommended that novel studies focus their research on type “review and experimentation,” in conjunction with EV and SD-SE risk factors. It would be beneficial to consider combining the mentioned factors with other factors, since there is little research around them.

5. Strengths and Limitations

This study includes all the risk factors found for COVID-19, regardless of the study type. It helps readers to know the factors that influence the presence of the disease. However, since only review articles were considered for the systematic mapping, other factors were likely excluded. While Scopus is one of the largest indexing digital libraries for health literature, this study did not include other digital databases. Also, the initial filtering of papers was focused on the title and abstract, which could lead to excluding relevant papers. Finally, other classification criteria might be appropriate for extracting further study information.

6. Conclusions

This study presents the systematic mapping of risk factors for COVID-19. Human risk factors are the ones that most influence the presence of the disease, highlighting diseases such as diabetes, obesity, hypertension, and cardiovascular diseases. In addition, most of the studies have considered these factors since the start of the pandemic. Regarding EV
factors, very little research is found in the literature, even though risk factors such as air, pollution and the environment are evident. As to SD-SE factors, the risk factors most closely related to the disease are age, sex, race and ethnicity. More research on EV and SD-SE risk factors is required to define what other factors can influence COVID-19. It is expected that this study will help to formulate future work around risk factors for COVID-19. In particular, this mapping lays the groundwork for the future research project entitled "Multidimensional machine learning model to determine a COVID-19 vulnerability index" on which the authors of this article are currently working. The objective is to characterize the risk factors that explain the vulnerability of COVID-19 and to build and evaluate a machine learning model that allows the calculation of a COVID-19 vulnerability index considering the most relevant variables, mainly those available from public datasets.

References


Sanchez-Ramirez DC, Mackey D. Underlying respiratory diseases, specifically COPD, and smoking are associated with severe COVID-19 outcomes: A systematic review and meta-analysis. Respir Med. 2020 September 1;171:106096.


[70] SciELO - Brazil - Extreme phenotypes approach to investigate host genetics and COVID-19 outcomes. Accessible at https://www.scielo.br/j/gmb/a/1hWg8yCqggCqzTDXjkyDW5C/?lang=en (accessed 17 March 2022)


