An Overview of COVID-19 Data Requirement for Business and Economics Research

Gour Gobinda Goswami
North South University, Bangladesh
Email: gour.goswami@northsouth.edu

KH Habibur Rahman
North South University, Bangladesh
Email: habibur.rahman02@northsouth.edu

Khaled Md. Saifullah
North South University, Bangladesh
Email: Khaled.saifullah01@northsouth.edu

ABSTRACT

Mainstream research on COVID-19 has an apparent bias toward clinical studies dominated by medical science, health science, or public health. But, considering the overall dimension of COVID-19 and its linkage with the recessionary impact on society's business and economic aspects, non-medical data are increasingly important nowadays to examine effects on society, environment, politics, economics, and business conditions of countries. As a result, a broad review of data on the overall impact of COVID-19 on all the different aspects is needed so that policymakers can find an appropriate policy-mix to deal with this pandemic. This article attempts to provide a thorough review of existing datasets in this area. We have identified and categorized five types of COVID-19 and related data: international, country-level, sub-national, community-level, and closely related data. The primary purpose of this review article is to identify and validate the most legitimate data sources available on the COVID-19 pandemic to guide future researchers and policy makers on using more accurate data in policy development and research. As an anecdote, the paper has also identified issues related to the data used in business and economic analysis on COVID-19. Utilizing the archival method, we

have compiled all the articles and the data. The data covered in this research is valid up to the middle of 2020.

Keywords: COVID-19; Coronvirus; Data; Socio-economic Data; Demographic Data; Environmental Data, Governance Data; Political Data; Data requirement; COVID-19 Research; International Data; Country level data; Sub-national Data; County Level Data; SARS-CoV-2; COVID-19 Data Review

1. Introduction

At the recent outbreak of the COVID-19 pandemic, the world witnessed many deaths daily, covering almost 227 countries and territories starting from December 31, 2019 until today. According to WHO, as of 18 December 2022, the world has witnessed over 649 million confirmed cases and 6.6 million deaths. Every day the number of cases and demises increased without any semblance of coming to a halt. At present, COVID-19 is considered the most dangerous pandemic right after the Spanish Flu, an even deadlier contagion in 1918, and the Flu was claimed to take the lives of 50 million people worldwide. The first case of COVID-19 was discovered in China's Wuhan province at the end of December 2019. Soon, the Chinese authority started quarantining the affected regions to contain the virus. However, it was beyond any comprehension that by March 2020 it would be declared a global pandemic. So far, the death toll crossed hundreds of thousands, and millions of people have been infected. Nations such as Italy, France, Spain, the United States, Russia, India, Brazil, the United Kingdom, Mexico, and many others suffered considerably.

With no success at finding a proper COVID-19 vaccine, many countries locked down their major cities and states to avoid further contamination**.** Parallel to the global healthcare situation, the world was hit by a massive economic depression. Due to the lockdown and subsequent fall in international trade, many countries had to shut down or downsize their industries and sectors, which caused unemployment and a deceleration in economic growth. According to a prediction by the World Trade Organization, global trade would likely shrink between 13% and 32% in 2020 due to COVID-related economic disruption. Sectors like manufacturing, art-entertainment, publishing, retail, tourism, technology, transport, and many others suffered enormously. Many countries announced their national fiscal responses and economic reforms to combat the prevalent situation and the upcoming probable

recession. Despite various attempts in many countries, unemployment was on the rise. The International Labor Organization (ILO) predicted that by the second quarter of 2020 there would be a 6.7% loss of job hours worldwide. The situation was catastrophic in developing and underdeveloped countries where technology has not advanced sufficiently. The economic downturn would worsen in the future if COVID-19 persisted further.

Predictions and recovery plans are based exclusively on the availability of data. Sources with accurate and real-time data on COVID-19 must ensure proper policies to cope with the pandemic and economic turbulence. Nevertheless, it is a global challenge to ensure the accuracy of data. This is because there are discrepancies in the COVID-19 data of many countries. Many countries show that their cases and death rates are significantly lower compared to some countries where the death rates are incredibly high. The inconsistencies in data and information have baffled many researchers. Therefore, it has become imperative to examine data sources with fewer discrepancies. With this objective, this study is based on the review of data sources of all types starting from clinical to non-clinical data on environment, demography, business conditions, mobility, governance, political superstructure, societal values in existing organizations as well as across economic, cultural, and political systems which have a bearing on COVID-19 or other similar pandemics.

Our findings show that different sources use different methods in different countries. For instance, many countries use different approaches to counting deaths. Many countries do not conduct adequate testing, or update the data irregularly, which show a smaller number of confirmed cases and deaths. This paper has identified various data sources from different countries and analyzed the methods and types of data they provide, which can be used for multiple research purposes. For example, policymakers require past and present data on COVID-19 to forecast future conditions accurately. Only then can effective policy to cope with the situation be ensured. Many countries are considering reopening economic activity slowly, by easing the lockdown in some areas, by relying on existing data. In other words, the degree of imposing or lifting lockdowns depends on the intensity of transmission of the disease and numbers of infected cases, with various deciding factors as perceived by the local or central authority. For this reason, getting access to the COVID-19 database is essential.

In many cases, policymakers and researchers need to compare various countries/states/communities using time-series, panel, and cross-sectional data to identify the flaws and efficacies of the different existing policies. This paper concentrates on the various country-level, sub-national, and community-level sources of the COVID-19 database. It analyzes the content and focus of these data sources, which will ensure a better understanding of the available database by avoiding incongruities. With this end in view, the article is organized as follows: Section 2 includes a brief literature review; section 3 provides a broad classification of data sources; section 4 deals with the uses of data; section 5 presents the limitations of the data; section 6 provides recommendations for future research; and section 7 concludes the study with some policy suggestions.

2. Literature Review

2.1 Literature on past pandemic

Researchers use various data to comprehend the nature and effects of the COVID-19 virus. To understand the virus, it is necessary to consider past research on previous pandemics like SARS and MERS and understand the researchers' data presented in those papers. Considering such references of past pandemics will help gather relevant data for researchers to get a better picture of the future.

Al-Raddadi et al. (2020) use a dynamic approach to estimate MERS-CoV's devastating consequences over a period of 31 months after the virus was first identified in 2012, including the nation-wide MERS-CoV epidemic that surfaced in Saudi Arabia in 2014 up until January 2015. In their study, the researchers analyze the real-time polymerase chain reaction (RT-PCR) data from surveillance, as confirmed by the health ministry. The researchers use the necessary information of the confirmed cases—for instance, sex, age, occupation, hospitalization date, patients' duration of stay at the hospital, mortality, duration of the case identification, rate of secondary attack, and calculation of the basic reproduction number. The researchers also perform descriptive epidemiology on the demographic data and analyze the data to assess the fatality rate among health care workers and others. The significant finding is that the MERS-CoV epidemic caused a heavy toll on the health system of Saudi Arabia. The researchers acknowledged that enhancing MERS-CoV data sharing and understanding the virus's

epidemiology would help make effective policies and preventive measures to reduce the load on the health system.

Alburikan & Abuelizz (2020) identify the factors associated with MERS-CoV infection among suspected case infections with symptoms and signs of exposure to the virus. The authors use the data on both confirmed and suspected cases. For instance, demographic data such as age and sex, existing medical records of patients, and medications at the screening point were collected. To report the demographic data, descriptive statistics were used. The authors' findings show that the MERS-CoV virus was more common among males with diabetes and advanced age.

In a study by Altamimi & Ahmed (2020), the authors aimed to find the correlation between climate factors and MERS-CoV. They aggregated monthly MERS-CoV case data from the Riyadh region between 01 November 2012 and 31 December 2018. Throughout the study the climate data were collected from Google reports, using the monthly average temperature, UV-index, wind speed, relative humidity (%), and cloud cover in Riyadh. The findings show that the MERS-CoV virus is positively correlated with climate conditions, with an increasing incidence between April and August. High UV-index, high temperature, low wind speed, and low relative humidity increased MERS-CoV cases. The authors concluded that climatic factors in hospitals and community settings should be evaluated and integrated into guidelines to serve as a control measure to avoid the risk of transmission of the virus.

Albarrak et al. (2019) focused on evaluating knowledge, practices, and attitudes about MERS among physicians, nurses, pharmacists, and technicians. They conducted a cross-sectional study among the healthcare workers at King Khalid University Hospital in Riyadh, Saudi Arabia. The authors used a self-administered survey questionnaire to measure health workers' knowledge and practices toward MERS. The questionnaire consists of five sections, including demographic data, source of MERS information, attitude, knowledge, and practices of healthcare workers towards MERS patients. The findings showed that the health workers had adequate knowledge and a positive attitude toward MERS patients. But, when it came to the practices towards MERS patients in a clinical setting, the findings were average to low, with low compliance of facemasking. The results also showed that the physicians were more knowledgeable and had better attitudes than nurses, pharmacists, and technicians.

After reviewing the literature as discussed above, it can be concluded that clinical and non-clinical data covering infrastructure, environment, etc., are required for successful conduct of any research on this kind of pandemic and its different effects in our daily life. Therefore, even in our present case of COVID-19, every responsible body in all countries must share adequate data, including clinical and non-clinical data, to help researchers comprehend the effects of the virus further. This will help make more practical speculations about the future and significantly reduce the contamination level of the COVID-19 virus.

2.2 Growing Literature of COVID-19 and its Implications

The current literature on COVID-19 is relatively new (i.e., since 2020). The following reviews are classified into two categories. One group explores the clinical data, to study the clinical aspects of COVID-19. The other group analyzes the non-clinical data, to review economic and sociological policies and related consequences of COVID-19.

2.2.1 Selected Clinical Studies

Hussain, Baxi, Jamali, Nisar, & Hussain (2020) studied the commonality of diabetes in patients infected with COVID-19 and determined the frequency of intensive care unit (ICU) admission of dying COVID-19 patients with and without diabetes. The authors gathered data by searching the literature in PubMed, and the analysis was conducted from a total of 23,007 patients from 43 studies. The findings indicate a high frequency of diabetes comorbidity and higher mortality and ICU admission risk among COVID-19 patients with diabetes.

In another study, Giesen et al. (2020) aimed to describe the number of COVID-19 cases at Infanta Sofia University Hospital in Madrid. To conduct the research, the authors gathered descriptive case studies that had involved PCR testing from 27 February to 29 June 2020. The study's findings showed a biphasic epidemiological curve; and, in addition, the authors also mentioned that further studies are needed to extract more epidemiological characteristics of COVID-19.

2.2.2 Selected Socioeconomic or non-clinical studies

Desson, Weller, McMeekin, & Ammi (2020) conducted a study to compare the policy responses and epidemiological situation during the early COVID-19 pandemic in France, Belgium, and Canada. The authors gathered data from secondary sources, from official government reports, international databases, and local media reports. The findings

revealed that the policy responses and epidemiological situations varied among these three countries depending on the individual country's governance and healthcare structures.

Ashraf (2020) conducted a study to analyze the economic impact of government interventions on the stock market during the pandemic. The author extracted data from government announcements regarding COVID-19 from 77 countries from 22 January to 17 April 2020. G overnment responses such as awareness programs, quarantine, and testing policies, along with support packages, stimulated stock market returns.

In their study, Yoo & Managi (2020) examined the benefits of actions performed in response to global mortality. The authors used country-level data from January 2020 and projected the mortality cases up until September 2020. To examine the benefits, the authors calculated the age and individual country's value of statistical life (VSL). The result revealed that global actions such as social distancing, quarantine, temporary termination of schools, and isolation have remarkable economic benefits.

In view of the foregoing data reviews, it is imperative to consider clinical and non-clinical social, environmental, governance, and related data to conduct successful research on COVID-19 societal impact. The main body of research on COVID-19 started with clinical data at the early epidemic stages, of course. However, as time passed, the effect of the virus was not confined to just one sphere of concern. People from all walks of life and professions, including economists, social scientists, and business researchers, became invested in relevant research. This is why after a few months of its inception, when the pandemic started, researchers from all disciplines began coming forward with their new areas of research within their disciplines of expertise. The following section explains the existing data sources, the regions and countries each source focuses on, and each category's focus.

3. The Existing Data Sources

We started preparing this study when the COVID-19 pandemic was taking shape; and, at that time, only peripheral literature from newspapers and reports were available. Academicians did not start writing or publishing much on COVID-19 at that time, except for clinical studies. Socioeconomic studies began a later. That is why we have relied on Google for all the papers and reports scattered

haphazardly. SCOPUS index-based articles started appearing in the list mainly at a later stage.

The data sources can be broadly divided into four categories, such as: (1) international, (2) country-level, (3) sub-national, and (4) community-level. For further simplification, we have classified the data sources into four sub-categories: (a) broad areas focused, (b) specific focus of the data sources, (c) update or frequency, and (d) type of data. A summary of the dataset is shown in Table 1 below.

Table 1. Broad Classification of Business and Economic Data for COVID-19

Serial	**Types of Data**	**Main Sources**	**Comments**	**Updates/frequencies**
01	International data	Thirty-two active sources	Clinical data, hospital data, non-clinical data such as demographic data, mobility data	Twenty-seven of them are updated on a real-time daily basis
02	Country-level data	Seventy-six sources	Same as above	Sixty-six databases that update their data in real-time daily. Ten sources update their data on an irregular basis.
03	Sub-national data	Forty-seven primary sources and thirty-seven other sources	Same as above	Updated mostly during weekdays with some exceptions
04	Community level data	Sixty-three primary sources and fifty-three other sources	Same as above	Updated mostly during weekdays, with some exceptions

Serial	Types of Data	Main Sources	Comments	Updates/frequencies
05	Closely related data	Socioeconomic, political, environmental, and governance-related data	They are also used in other research areas except for COVID-19	They are mostly in annual frequency

Source: Own compilation

Each type of data is elaborated below.

3.1 International Data Sources

3.1.1 Broader Focus of the International Data Sources

The international database comprises 32 active sources. These include WHO, UNICEF, Worldometer, EU Open Data Portal, 1Point3 Acreshttps://coronavirus.1point3acres.com/, Corona data scraper, COVID-19 data portal, Apple Mobility report, and 24 others. These sources provide the most reliable data on COVID-19 and its different dimensions from a broad global perspective.

3.1.2 Specific Focus of the International Data Sources

These sources show the global status of COVID-19 by breaking down the cumulative global deaths, cases, and recoveries into country data. The data from these sources allow us to compare and understand the COVID-19 status of various countries. However, some sources focus on other countries or regions besides referencing the global data. Such sources include Euronews, Tableau, EU Open Data Portal, Ding Xiang Yuan, COTA, and John Hopkins University Coronavirus Resource Center. For instance, although showing international data of COVID-19 updates, EU Open also shows detailed country data mainly of European countries as published by the European Centre for Disease Prevention and Control.

The majority of these sources tend to focus on clinical data such as COVID-19 confirmed cases, number of deaths, mortality rate, the total number of tests, and the rate of incidence. Such sites include WHO, Worldometer, Our World in Data, Euronews, Tableau, HDX, EU Open Data Portal, 1point3Acres, The Kaggle Novel Coronavirus Dataset,

Ding Xiang Yuan, European Centre for Disease Prevention and Control (ECDC), Corona Data Scraper, COVID-19 public datasets, National Institutes of Health, CSH, Bing, Facebook, Johns Hopkins University Coronavirus Resource Center, COVID-19 Data Hub, Github (Microsoft/Bing COVID-19 data), Yahoo, Github (Microsoft/Bing COVID-19 data), nCoV2019.live. In contrast, Facebook shows more detailed data, such as the number of people with COVID-19 symptoms and flu symptoms of the total confirmed COVID-19 cases collected using surveys. However, many of these sites lack details about hospital data across countries. The hospital data**,** such as ICU availability, hospital beds availability, and ventilation equipment data of individual countries were primarily unavailable in the international level category, except for the COVID-19 Data Hub. Among the available sources, there are non-clinical data available as well.

Most sources show cumulative data on global COVID-19 status, which is divided into sub-categories in terms of cumulative data per country. Thus, detailed data for each country is not shown. Demographic data such as the age and sex of COVID-19 patients is shown only in Our World in Data, Euronews, and Worldometer. These sites used variables like cases and deaths by age and sex, age structure, and death analysis by age group.

Similarly, transmission and mobility data were scarce in number in the given international sources. Only sites like Google mobility and Apple mobility are dedicated sites that show mobility and transmission data.

Other non-clinical data can be found in the COVID-19 Data Portal, such as sequence data related to COVID-19, SARS-CoV data for drug target identification and prioritization, latest literature on SARS CoV-2, and a variety of resources for studying the SARS CoV-2 coronavirus and COVID-19 disease. Another source, Kaggle, shows a list of countries and the dates each country went into lockdown. The dataset comprises country/region, province, lockdown date, type of lockdown, and information reference. Detailed analysis is available in a structured format (see Table A1).

3.1.3 Database Update and Frequency

Only five of the 32 international data sources tend to upload their data regularly. However, in some cases, a specific range of data is updated periodically, while some are updated irregularly. The rest of the 27 sources are updated on a real-time daily basis.

3.1.4 Type of Data

The datasets in the given sources are mostly panel data, while in some cases cumulative world data are presented in time-series. Data from the majority of sources are shown using descriptive statistical tools.

3.2 Country-level Data

3.2.1 Broader Focus of the Country-level Data Sources

The country-level database contains 76 sources. The sources include the USA CDC, Johns Hopkins University Coronavirus Resource Center, China CDC, Taiwan CDC, the Government of Pakistan, The U.K. government, the UK Office of National Statistics, the Scottish Government, Bengal Institute, Ministry of Health Singapore, Italy Ministry of Health, Ministry of Health of the Republic of Serbia, and 64 other sources.

3.2.2 Specific Focus of the Country-level Data Sources

The country-level sources mainly focus on the COVID-19 status of specific countries. The US CDC only focuses on the United States' COVID-19 status data. Bangladesh's DGHS focuses only on Bangladesh's COVID-19 data. Similarly, there are other sources, each focusing on a single country's data. The sources show detailed information on COVID-19 in each country by breaking down each division/district/state/region status. For instance, the Indian Ministry of Health and Family Welfare shows the country's statewide COVID-19 level, showing each state's severity of the virus. Likewise, Japan's COVID-19 Coronavirus Tracker shows the national and regional COVID-19 status of Japan. These data breakdowns help the users understand the exact COVID-19 situation of the entire country and comparative status across-states/divisions/districts/regions.

Some sources also focus specifically on clinical and hospital data such as the number of COVID-19 cases/deaths, tests, and hospital data such numbers of patients in ICU, patients on ventilators, etc. The Government of Ireland, for example, provides detailed datasets at their website about COVID-19 cases, deaths, and statistics about hospitalization and ICUs. The dataset includes the geographic distribution of COVID-19 cases in Ireland. The series of dashboards also provide the latest information on confirmed COVID-19 cases in ICU and acute hospitalizations and the newest data on laboratory testing.

Likewise, the Ministry of Health of Singapore focused on hospital and clinical data. Their site shows the total number of imported cases in Singapore, active and discharged case summary in Singapore, the number of swab tests, and the latest updates on COVID-19. The case summary includes hospitalization cases and community facilities cases. All the database sources used standard variables, e.g., cumulative /new cases, deaths, number of tests, cases per 1000 people, recovered cases , number of COVID-19 associated cases, antibody surveillance, asymptomatic infected persons, hospitalization by region, recovery and total PCR, mortality ratio, number of people in isolation, number of people in ICU, the age distribution of COVID-19 ICU patients to date, number of PCR tests, number of patients hospitalized to show the COVID-19 status within the country. Among all these sources, only US CDC and Taiwan CDC contain data on pre-existing health conditions such as pneumonia in COVID-19 patients. However, the non-clinical data sources did not have these variables in common.

Apart from the above information, some sites include demographic data such as age, sex, and ethnicity. For instance, the US CDC, Scottish Government, Government of Luxembourg, National Records of Scotland, Hong Kong Department of Health, National Records of Scotland, Government of Ireland, Government of Canada, Austrian Federal Ministry for Health, Terviseamet (Estonia), Finland Department of Health, Norwegian National Institute of Public health, Government of Peru, Covid-19 Datelazi (Romania), Federal Office of Public Health (Switzerland) shows datasets of age and sex. Among these sources, New Zealand's Ministry of Health, the US CDC, also offers data on cases by ethnicity.

Besides the clinical information, some sites include data on COVID-19 transmission and mobility**.** New Zealand's Ministry of Health, the Government of Ireland, Github's portal of South Africa, and China CDC provide information about the transmission of the COVID-19 virus in their respective countries. Among these sites, US CDC includes data on mobility and transmission data simultaneously.

Although the lion's share of these sources tends to focus on clinical data, some sources focus on non-clinical data**.** For instance, "Five Thirty-Eight" focuses on the Americans' concerns and their thoughts on the Trump Administration responses to the outbreak. "Notion" focuses on how schools are affected by COVID-19. "Kaggle" shows the updated dataset containing stay-home order dates for different U.S. states and

counties. The Stanford Libraries shows the statistical code to replicate analysis presented in the pre-print named "Using rapid online surveys to assess perceptions during infectious disease outbreaks: a cross-sectional survey on Covid-19 among the public in the United States and the United Kingdom."[1]

3.2.3 Update or Frequency

The country-level database breaks down country-level information into the district/province level. For instance, the Bangladesh DGHS collects its data from secondary sources, like local hospitals and health institutions. Soon after collecting the data, it is published on their official site. A total of 63 database sources update their data in real-time daily. The National Records of Scotland and Esri Australia update some of their data on a real-time basis and some irregularly. However, there are a total of 10 sources that update their data on an irregular basis. These include Bangladesh Care, UK Office of National Institute, Bengal Institute, US CDC, Github (South Africa), Kaggle (U.S.), Notion (U.S.), Stanford Libraries (U.S. and U.K.), Government of Azerbaijan, COVID-19 Mauritius.

3.2.4 Type of Data

It is important to note that most of these sources show panel data, allowing for comparison between states/divisions/districts/regions over time. However, some sources have time-series and cross-sectional data as well. For example, The Office of National Statistics contains time series variables and Stanford Libraries is a database that contains cross-sectional data from the U.K. and the U.S.

3.3 Sub-national Data

3.3.1 Broader Focus of the Sub-national Data Sources

The sub-national level database of the USA contains 47 sources. The sources include Michigan.gov, Washington State Department of Health, Maryland Department of Health, New York State Department of Health, Colorado Department of Public Health & Environment, Virginia Department of Health, Missouri Department of Health, Illinois Department of Public Health, Indiana State Department of Health, Oregon Health Authority, and 37 other sources.

[1] Detailed Appendix Table A2 is available upon request.

3.3.2 Specific Focus of the Sub-National Data Sources

The sub-national category of the COVID-19 database emphasizes state-level cities. The database mainly contains data for the states of the United States. For instance, the Michigan.gov data source focuses on the state of Michigan only; the Washington State Department of Health data source focuses on the state of Washington only. Likewise, other sources concentrate on the COVID-19 status of the conditions in the United States. However, aside from focusing only on the United States, the database contains data on a few other cities, including Kerala, London, New South Wales, Calcutta, and Delhi. Each of these sources focuses on these cities specifically. Most of these sources demonstrate their updates on COVID-19 using a data dashboard, and the information shown in the dashboard is very detailed and categorized at the community level. For instance, the Washington State Department of Health demonstrates Washington's COVID-19 status of severity by county. Similarly, the data source of Maryland.gov shows the rate of COVID-19 variables like cases, deaths, tests, and hospitalized by the county of Maryland. These classifications of COVID-19 status by county help the users comprehend a more precise and detailed image of the entire state and cross-compare between counties of specific conditions.

Some sources in the sub-national category database contain clinical and hospital data, e.g., cumulative cases, deaths, tests, hospitalization, recovery, ICU beds, ventilators, laboratory testing, and many others. Corona Dashboard contains the hospital data for Delhi. The dashboard comprises three variables COVID-19 beds, ICU beds with ventilators, and ICU beds without ventilators. The variables are further classified into total, occupied, and vacant. The dashboard is very detailed, and the information on these three variables is broken down into Delhi's hospitals.

Similarly, the Wyoming Department of Health in the USA also focuses on hospital and clinical data. Wyoming's site demonstrates laboratory cases, probable cases, COVID-19-related deaths, cases by county, reported hospitalized COVID-19 patients, statewide ICU and ventilator capacity, hospital COVID-19 testing, and total reported resources (over time by hospital and county). Other data dashboard series demonstrate Wyoming COVID-19 reporting by state and county, map and statistics, and testing. This site provides very detailed and categorized information for COVID-19 in Wyoming. However, aside from containing only hospital and clinical data, sources have three kinds

of data: clinical, hospital, and non-clinical. For instance, the GoK Dashboard is a COVID-19 dashboard for Kerala in India, providing clinical, hospital, and non-clinical data. The dashboard demonstrates the total confirmed cases, active cases, recovered, deaths, samples tested, number quarantined by district, total under observation, total hospitalized, and numbers under home isolation. All the data sources used in this COVID-19 database under the sub-national category have used clinical data (cumulative cases, deaths, number recovered, and number of tests).

Aside from clinical and hospital data, a few sites in the USA under the sub-national category include demographic data (age, sex, ethnicity, and county). Such sources include Michigan.gov, Maryland Department of Health, New York State Department of Health, New York State Department of Health, Colorado Department of Public Health & Environment, Missouri Department of Health, and others. However, among these sources, there are few other sources like the Washington State Department of Health and New York State Department of Health, where variables such as cases or deaths are categorized under variables such as weekly illness onset and test date.

Besides the clinical, hospital, and demographic information, one specific source deals with mobility data. The site of the University of Virginia contains a global surveillance dashboard for COVID-19, ~~a~~ population contact rates by age and county dataset, and Baidu mobility data for January through March 2020. There is no source other than the University of Virginia, which provides mobility data.

Besides clinical, demographic, hospital, and mobility data, some sources in the USA focus on non-clinical data. For instance, Mass.gov focuses on residents subjected to COVID-19 quarantine. Similarly, besides providing clinical and hospital data, the Government of the District of Columbia also offers non-clinical data, such as data on public safety and human service agencies, and Department of Motor Vehicles. All these non-clinical data focus on information such individuals quarantined or returned to work. These individuals are the employees from the public safety and service departments, such as Fire and Emergency Medical Services (FEMS), Metropolitan Police Department (MPD), Saint Elizabeth's Hospital, and Department of Disability Services.[2]

[2] Detailed Appendix Table A3 is available upon request.

3.3.3 Update or Frequency

The update or frequency of the sub-national category database in the USA varies according to the source. Three specific sources update their data in an irregular pattern. For instance, the Washington State Department of Health updates its data dashboard weekly on Sundays, the Kansas Department of Health and Environment updates its status of COVID-19 on Monday, Wednesday, and Friday by 12:30 pm., and the Rhode Island Department of Health update their data dashboard on Monday to Friday by 12 pm. However, the list of sources with regular updates covers a significant portion of the sub-national category database—for instance, Michigan.gov (Michigan), Maryland Department of Health, New York State Department of Health, Colorado Department of Public Health & Environment, and 44 other sources.

3.3.4 Type of Data

All the dataset under the sub-national category contains panel data, allowing users to compare counties over time. However, a few sources may collect cross-sectional and time-series data.

3.4 Community-level Data

3.4.1 Broader Focus of the Community-level Data Sources

The database under the category of Community-level for the USA contains 63 sources. The sources include Google (global), Access Kent (Kent County), Alameda County Public Health Department (Alameda), Alpine County California (Alpine), Amador County California (Amador), Butte County California (Butte), Calaveras County (Calaveras), Colusa County California (Colusa), Contra Costa Health Service (Contra Costa), County of Del Norte California (Del Norte) and 53 other sources.

3.4.2 The Specific Focus of the Community-level Data Sources

The primary focus of the community-level data sources is the COVID-19 situation in counties in the USA. For instance, Access Kent focuses on the COVID-19 status in Kent County only; Alameda County Public Health Department focuses on the COVID-19 status in Alameda County only; Alpine County California focuses on Alpine County only. Similarly, each of the data sources of this category focuses on each county's COVID-19 status. The community-level database mainly contains data sources of counties under the United States, except for

Google since Google is the only well-recognized source in the community-level database focusing on the global community. There are sources in the community-level database broken down into categories like demographics and city/jurisdiction. For example, Access Kent focuses on the status of COVID-19 in Kent county. The source categorizes the variables of COVID-19, e.g., cases, deaths, and recoveries, under demographic variables. Alameda County Public Health Department also focuses on COVID-19 variables under demographics and city/jurisdiction. Similarly, there are sources such as Google focusing on mobility data; Alpine County, California, which concentrates on clinical, and hospital data; and Marin Health and Human Service focuses on non-clinical data. These classifications of COVID-19 variables help users understand a detailed picture of the COVID-19 situation in different counties.

Sources dealing with clinical and hospital data mainly involve variables like cumulative cases, deaths, testing, diagnostic testing, hospital supplies, laboratory testing, recovered, and hospitalized factors. For instance, the site of Calaveras County contains Calaveras community-level status of the COVID-19 variables by the numbers. Such variables include active cases, currently hospitalized, recovered, deaths, totally tested, place of residence under investigation, and total cases factors. The category of demographics also includes all these variables. The status of COVID-19 in Calaveras county is shown in a PDF file.

Similarly, the County of Del Norte California site also contains COVID-19 clinical and hospital data. However, the only difference between the location of Calaveras and Del Norte is that the data is demonstrated in a data dashboard. The data hub of Del Norte focuses on variables such as confirmed cases, total tests, currently hospitalized, total deaths, COVID-19 test results for Del Norte County, and transmission cases by age and gender. The data site of Fresno County contains cumulative variables of COVID-19. Along with that, the site also holds a testing site, and data dashboard. The site is detailed with information, and there is an additional table where the risk level of COVID-19 is broken down into four groups: widespread, substantial, moderate, and minimal. A different dashboard is incorporated into the site, containing COVID-19 surveillance for Fresno County. In addition to the dashboard, there is an additional section on demographics and mortality shown by graphs and pie charts. The data site of Fresno County focused on variables including California case statistics, case demographics, California testing results, positive cases by county, total

cases, active cases, recovered, deceased, source of exposure, cases over time, and cases by demographics. All the sources in the community-level database hold clinical and hospital data, except for Google. Aside from keeping clinical and hospital data, these sources' datasets are also classified into demographics.

Nevertheless, aside from focusing on clinical and hospital variables, there are sources too where the variables of COVID-19 impact are classified into demographics, city/jurisdiction, zip code, and date. For instance, the Imperial County site categorized the variables of COVID-19 into zip code, location, and community. Similarly, the Alameda County Public Health Department's site sorts out their city/jurisdiction and demographics variables. Butte County sorts its variables by region, day, week, month, and demographics. There are other sources in the database where variables are sorted out in demographics, city/jurisdiction, zip code, and date (for instance, Google COVID-19 Community Mobility Reports, Access Kent, Calaveras County, County of Del Norte California).

Aside from the clinical and hospital information, the community-level database includes data on COVID-19 mobility. The Google COVID-19 community Mobility Report is the only data source in the community-level database focusing on COVID-19 mobility. The dataset shows mobility trends by region and different categories of place. The dataset is not aimed at clinical purposes; it only shows how mobility in grocery stores and parks changes in each geographic region. Datasets of specific countries or areas show mobility trends in retail and recreation, grocery and pharmacy, park, transit stations, workplaces, and residential. The mobility report can be searched either by country or region. However, most countries are not subcategorized at the regional level except for a few countries, including Argentina, Australia, Austria, Belgium, Bolivia, Brazil, Bulgaria, Cameroon, Canada, Chile, Colombia, and 28 others. The dataset is retrievable in pdf form and, for a few countries the dataset can be retrieved in two languages (English and Persian). Moreover, the data site has a few options for users to understand the dataset. Such an opportunity includes an "overview" to help the user understand how public health officials can use this dataset, understand the data, and demonstrates examples to interpret common patterns in the dataset.

Along with the clinical, hospital, demographic, and mobility data, the community-level database also contains a non-clinical level dataset.

For instance, the site of Alpine County updates non-clinical information for employers and displaced workers due to COVID-19 along with clinical and hospital information. Similarly, Contra Costa health services contain homeless persons data, homeless and at-risk data, hotel placement data, and clinical and hospital data. The data site of Sonoma County contains COVID-19 impact planning report data along with clinical data such as cases, death, and others for Sonoma County, USA, and California. A few other sources focus on non-clinical data other than those mentioned above, for instance, Lander County, Nevada, and Yolo County.[3]

3.4.3 Database Update or Frequency

The update frequency under the community-level database category is of two different types, real-time daily, and irregular patterns. The sources with a daily update frequency of real-time tend to update their dataset within a day. Some with irregular patterns update datasets irregularly, for instance, twice a week or five days a week. Most of the sources in the community-level database update their dataset daily in real-time. Such sources include Google COVID-19 Community Mobility Reports, Access Kent (Kent County), Alameda County Public Health Department, Amador County, California, Contra Costa Health Service, and 39 other sources.

One specific source provides updates of some of its data regularly and some irregularly. The site of Pershing County Nevada, e.g., updates the data of Pershing County in real-time daily and updates the data of Nevada in an irregular pattern. Among all these sources in the database, the number of sources with varying update frequencies is lower than sources with a regular update frequency. Sources with irregular update frequency include Butte County, where the dataset is updated Mondays through Fridays by 4 pm. For the site of Calaveras County, the latest is updated each Tuesday and Friday. The information provided by Colusa County's location is updated from Monday to Friday at 10 am. There are overall 17 sources that update their data with such irregular frequency.

3.4.4 Type of Data

The data from community sources are mostly panel data. However, some sources might contain time-series and cross-sectional data, as well. Most of these data are readily downloadable in various formats for free.

[3] Detailed Appendix Table A4 is available upon request.

3.5. Closely Related Data

Apart from using clinical data to assess the impact of COVID-19, a few other variables passively related to COVID-19 have been used to determine the virus's effect. Epidemiological and clinical variables have been used frequently to assess the COVID-19 situation. However, environmental, socioeconomic, sociodemographic, governance and many others have been used less in determining the COVID-19 impact.

Few studies have been conducted assessing the impact of coronavirus by examining environmental indicators. For instance, Aman and Salman (2020) evaluated the effects of COVID-19 by analyzing environmental indicators in Ahmedabad, India. To conduct such a study, the authors used the water and air quality index. The authors used remote sensing data to demonstrate the air and water quality over time to assess the effects of COVID-19.

Similarly, Saha and Chouhan (2020) conducted a study in India to determine the risk factors related to COVID-19 among children below five years of age by assessing indicators such as morbidity and environmental factors including indoor air pollution and fatality, and coronavirus recovery rate in India. The authors collected data from the fourth round of the National Family Health Survey (NFHS) from 2005 to 2016 and the Ministry of Health and Family Welfare on May 18, 2020.

In another study, Mandal and Pal (2020) assessed the impact of COVID-19 lockdown on environmental quality by analyzing components of air quality such particulate matter (PM), land surface temperature, river water quality, and noise using images. The field-derived data were collected from before and during the lockdown.

Variables like socioeconomic, sociodemographic, and economic structures have been used in assessing the impact of COVID-19. For instance, Shahbazi and Khazaei (2020) incorporated the Human Development Index (HDI) as an indicator and aimed to determine the role of such an indicator in assessing the incidence and mortality rate of COVID-19 worldwide. The authors collected data on incidence and mortality rates from the Worldometer and data of the HDI 2019, compiled from the World Bank database.

Similarly, Shen, Cai and Li (2020) assessed the role of restrictions in socioeconomic activities during the outbreak of COVID-19 globally. The authors collected data from the satellite readings of nitrogen dioxide, a pollutant emitted from socioeconomic activities.

Sociodemographic factors have been essential, too, in assessing the COVID-19 pandemic. Witteveen (2020) evaluated the impact of COVID-19 on earning and sociodemographic groups exposed to economic poverty. The author used data from the UK household longitudinal study COVID-19 supplement to conduct this study.

Governance plays an essential role in examining the COVID-19 outbreak. Sharma, Borah & Moses (2020) focused on the role of governance, healthcare infrastructure, and lessons from past epidemics in assessing the responses to COVID-19. To examine how these three variables function, the authors reviewed the literature on governance structure, past pandemics, and investment in healthcare infrastructure, and collected data from various sources.

Thus, sociodemographic, socioeconomic, environmental, and governance variables are important to assessing other dimensions of COVID-19's impact. Apart from the clinical variables of the COVID-19 pandemic, such variables are also needed for the purpose of efficacious policy-making.

The sources used in these kinds of data are diverse. Researchers mostly use data from the World Bank, IMF, UNDP, WTO, WHO, and UNICEF, depending on the nature of the project. They combine different indicators into COVID-19 related variables and fit different models based on their specific research requirements. Our World in Data is a noble initiative to connect clinical and other data types for all the countries and areas of the world into a daily panel format. The data are updated at regular intervals. This dataset is extremely useful for conducting advanced panel data-based modeling of COVID-19. But the major limitation is that time-invariant variables have a repetitive pattern.

4. Uses of COVID-19 Data

In the 21st century, mitigation of any crisis requires data and information. Nevertheless, acquiring a dataset with accuracy has become a challenge. As COVID-19 has become a global concern, policymakers need considerable data input to shape policies that would fit best for the world. Hence, this section aims to discuss the usefulness of the COVID-19 database and how it can be utilized appropriately.

The analysis of the COVID-19 database has revealed a few indicators that are directly related to COVID-19, e.g., the transmission

of the virus, cumulative cases, hospital beds, cumulative deaths, and many others. Nevertheless, a few indicators of COVID-19 impact are indirectly related, e.g., the stock market, exchange rate, environment, governance, etc. All indicators show the trend, implication, and impact of COVID-19, which is helpful in conducting a situation analysis for an individual country over time. As mentioned earlier, the whole COVID-19 database has mostly panel data type, which is useful to examine changes in variables over time.

Ozyigit (2020) conducted a study to examine the effect of temperature and health behavior on the transmission rate of COVID-19. The author used panel data estimation for EU-15 countries to surmise the impact of temperature and health behavior on the transmission rate.

Yu, Lei, Li, Wang, Liu, Fan and Li (2020) conducted a study investigating the clinical features and risk factors of COVID-19 mortality and predicting the mortality risk in COVID-19 patients. The authors collected data on COVID-19 patient mortality, recovery, demographics, clinical and laboratory data on admission. They used multivariate logistic regression for this analysis. Similarly, Wang et al. (2020) also conducted a study involving real-time estimation and prediction of mortality caused by COVID-19. These authors have collected data from publicly available data sources for this study.

Bhattacharjee, Kumar and Patel (2020) aim to predict the declination of COVID-19 using the recovery rate and case rate. The authors collected data on the case rate and recovery rate from a publicly available data hub to conduct this study.

Yilmazkuday (2020a, b, c) examines the changes in consumption patterns due to COVID-19, demographic discrimination in social distancing, and other issues of economic interest such as lower inter-country mobility due to COVID-19 transmission and death.

Assessment of the COVID-19 pandemic impact requires indicators like case rate, transmission rate, death rate, and others. Therefore, Khan, Haleem and Javaid (2020) analyzed the effect of the COVID-19 pandemic across the 13 most infected countries through COVID-19 cases, deaths, and recoveries.

However, the indirectly related indicators of COVID-19 are relevant to assessing the impact of COVID-19. Topcu and Gulal (2020) examined the effect of COVID-19 on stock markets from 10 March to

30 April 2020. The authors obtained data on exchange rates, oil price shocks, and infection rates from 10March to 30 April 2020 for 26 stock markets by MSCI.

In another study, Iqbal, Fareed, Shahzad, He, Shahzad and Lina (2020) assessed the connection between weather, the COVID-19 outbreak in Wuhan, and the Chinese economy. To conduct this study, the authors used daily average temperature, daily new confirmed cases of COVID-19 in Wuhan, and the exchange rate of the Chinese currency to represent the weather, the COVID-19 outbreak, and the Chinese economy.

Hence, assessing the impact of the COVID-19 pandemic requires indicators like epidemiological and public health data such as mortality, transmission rate, recovery cases, case rate, number of hospital beds, but also social impact data such as stock market, exchange rate, environment, and others.

5. Limitations of COVID-19 Data

In the given situation, the COVID-19 related data are vital in for the performance of the healthcare systems and government policymaking worldwide. However, besides the usability of the data sources mentioned above, it is also essential to consider the limitations.

It should be kept in mind that not all data sources are collected or created using the same methods. There is an insufficiency of alignment among the players battling the spread of COVID-19 disease and thus about what is being compared and measured. In many countries it has been seen that if a COVID-19 patient dies from other diseases, his/her death is reported under COVID-19 Death (Henriques, 2020). This often fails to show the real impact of COVID-19, creating uncertainty and inconsistency among the users of these data. These above-mentioned factors pose a significant limitation for many researchers in answering the death rate discrepancies across countries.

Similarly, aside from other constraints of COVID-19 data, the data collection methods are also to be critiqued. For instance, the data collection method in public health is mainly based on the manual data collection and coding, which is slower to collect and difficult to comprehend. Also, the collection of such datasets requires a large amount of time. Hence, considering the current situation, the collection of datasets in a traditional way can be quite time-consuming (Callaghan,

2020). So far, depending on the strain of the virus causing COVID-19 infection, in many cases the disease is considered mild and, therefore, self-treated. The number of COVID-19 patients who are seeking healthcare facilities is reported as COVID-19 as both cases and deaths. Using hospital records to estimate the mortality or morbidity rate of COVID-19 can be misleading sometimes.

It has been noticed that most countries (e.g., low-income countries) do not have proper sites for COVID-19 data. There is only a limited amount of data in existing sources in those countries, mainly showing the cumulative number of deaths and cases, along with newly added rates/numbers of deaths and cases. Also, the data at the community level is unavailable primarily in these sources. Although government testing in low-income countries for COVID-19 is free, it reaches a limited number of citizens.

Furthermore, testing for COVID-19 in hospitals in low-income countries is not readily affordable for many citizens. Hence, only the people who are tested are reported. This can result in under-reporting of the actual number of COVID-19 cases.

In contrast, high-income countries of Europe, countries like the USA and Australia, etc., have detailed and elaborated data on their population regarding COVID-19. Compared to the tests conducted in low-income countries, these high-income countries have a better testing system. In most high-income countries, the tests are done for free by the government and are available for everyone. This enables them to report many cases compared to low-income countries. The methods used by sources in these high-income countries to collect these reports are often well defined. When writing this paper, we saw that the USA was the only country that provided the highest number of community-level data. These sources include total cases, deaths, cases by sex, age, race, ethnicity, and other detailed variables that could not be found in data sources from lo-income countries.

The irregularly updated data creates a data lag when compared to the sources that update data on a real-time daily basis. In many countries, data are not updated on national/weekly holidays. This creates irregularly updated data sources, thus to under-reporting the COVID-19 cases compared to the sources that regularly update their data.

Therefore, the user is advised to consider the factors mentioned above while using the data sources.

6. Recommendations for future research

We used the archival method to compile all the literature reviewed here because we conducted this project when COVID-19 was appearing as a pandemic. The main body of academic literature was developed in 2021 and 2022 when some countries experienced the COVID-19 pandemic. That's why now we have thousands of papers available in this area and data are also readily available. Researchers are also conducting a bibliometric literature review of disease transmission, mortality, vaccination rates, and socioeconomic impact studies (Goswami and Labib, 2022). The same data review may now be extended using this modern and sophisticated method. A systematic review would be another area where there is an avenue for fulfilling the research gap.

7. Concluding Remarks and Policy Suggestions

Clinical data are essential for COVID-19 research for all disciplines because it prepares the foundation for primary information about diseases, viruses, patients, and other related matters. Thereafter, these clinical data must be complemented by other non-clinical or closely related data to examine the nature of transmission, mortality, the role of government, politics, institutions, environment, ecology, economic condition, social and cultural issues, the overall governance structure of a nation, as well as international organizations, diplomatic matters, the role of political leaders and their farsightedness in fighting the pandemic. The role of scientists in implementing successful trials before launching a vaccine is also important. To deal with all these issues, researchers may conduct primary surveys or repeat surveys and collect first-hand data to support the world community in combating this pandemic. However, these depend on an institution's preparedness to conduct such surveys during a crisis such as COVID-19. Another option is to rely exclusively on secondary data from hospitals or testing centers compiled by each government at the community, state, or county levels. These data are subsequently processed by international organizations or think tanks and made available and free for conducting research and prescribing policies and forecasts for comprehending the future dimensions of this disease. Data are essential in the sense that policy decisions based on merely ad hoc judgments are subject to severe mistakes and may guide us in the wrong direction.

The purpose of this paper is to contribute a little in this area to compile the sources of secondary data for COVID-19 available up until

the middle of 2020 and make this information available to research scholars so that they get a quick review of the available sources for the conduct of meaningful research in business and economics.

We have also explored the categories and limitations of data while researchers use them. This is especially important because we observed some inherent underreporting of COVID-19 data, especially in the case of less developed countries. We have come across several puzzles in COVID-19 related information. For example, many countries fail to conduct sufficient testing. As a result, the transmission rates are underreported in those countries. In death rates, the counts are sometimes overestimated because it is challenging to distinguish between a COVID-19 death and deaths from other causes. There is no mechanism to deal with asymptomatic cases and deaths. These matters have not yet been resolved. COVID-19 is primarily random in its transmission. All governments are trying hard to understand the virus's underlying nature, with limited success because of the virus's frequent changes in its genome and thus to produce variants and sub-variants. This results in several waves of infection in many countries, even after the successful containment of earlier waves by way of vaccination programs.

We hope this kind of data review article will help policy makers to have access to hands-on materials that can be used for research purposes and to guide their respective governments to deal with future pandemics properly, so that transmission and mortality may be minimized and the general public are encouraged to take vaccination properly.

Conflict of Interest: *The authors declare that there is no conflict of interest in preparing this manuscript.*

Funding: *There was no financial involvement in preparing this manuscript.*

References

Albarrak, A., Mohammed, R., Al Elayan, A., Al Fawaz, F., Al Masry, M., Al Shammari, M., & Miaygil, S. (2019). Middle East Respiratory Syndrome (MERS): Comparing the knowledge, attitude and practices of different health care workers. Journal of Infection And Public Health. doi: 10.1016/j.jiph.2019.06.029

Alburikan, K., & Abuelizz, H. (2020). Identifying factors and target preventive therapies for Middle East Respiratory Syndrome sucsibtable patients. *Saudi Pharmaceutical Journal*, *28*(2), 161-164. doi: 10.1016/j.jsps.2019.11.016

Al-Raddadi, R., Shabouni, O., Alraddadi, Z., Alzalabani, A., Al-Asmari, A., & Ibrahim, A. et al. (2020). Burden of Middle East respiratory syndrome coronavirus infection in Saudi Arabia. *Journal of Infection and Public Health*, *13*(5), 692-696. doi: 10.1016/j.jiph.2019.11.016\

Altamimi, A., & Ahmed, A. (2020). Climate factors and incidence of Middle East respiratory syndrome coronavirus. *Journal of Infection and Public Health, 13*(5), 704-708. doi: 10.1016/j.jiph.2019.11.011

Aman, M. A., & Salman, M. S. (2020). COVID-19 and its impact on environment: Improved pollution levels during the lockdown period–A case from Ahmedabad, India. *Remote Sensing Applications: Society and Environment*, 100382.

Ashraf BN. Economic impact of government interventions during the COVID-19 pandemic: International evidence from financial markets. J Behav Exp Finance. 2020 Sep; 27:100371. doi: 10.1016/j.jbef.2020.100371. Epub 2020 Jun 29. PMID: 32835011; PMCID: PMC7323651.

Bhattacharjee, A., Kumar, M., & Patel, K. K. (2020). When COVID-19 will Decline in India? Prediction by Combination of Recovery and Case Load Rate.

Callaghan S. COVID-19 Is a Data Science Issue. Patterns (N Y). 2020 May 8;1(2):100022. doi: 10.1016/j.patter.2020.100022. Epub 2020 Apr 7. PMID: 32562484; PMCID: PMC7144860.

Desson Z, Weller E, McMeekin P, Ammi M. An analysis of the policy responses to the COVID-19 pandemic in France, Belgium, and Canada. Health Policy Technol. 2020 Dec;9(4):430-446. doi: 10.1016/j.hlpt. 2020.09.002. Epub 2020 Sep 5. PMID: 33520640; PMCID: PMC7832234.

Giesen C, Diez-Izquierdo L, Saa-Requejo CM, Lopez-Carrillo I, Lopez-Vilela CA, Seco-Martinez A, Prieto MTR, Malmierca E, Garcia-Fernandez C; COVID Epidemiological Surveillance and Control Study Group. Epidemiological characteristics of the COVID-19 outbreak in a secondary hospital in Spain. Am J Infect Control. 2021 Feb;49(2):143-150. doi: 10.1016/j.ajic.2020.07.014. Epub 2020 Jul 11. PMID: 32663494; PMCID: PMC7834034.

Goswami, G.G.; Labib, T. Modeling COVID-19 Transmission Dynamics: A Bibliometric Review. *Int. J. Environ. Res. Public Health* 2022, *19*, 14143. https://doi.org/10.3390/ ijerph192114143

Henriques, M. (2020, April 2). Coronavirus: Why death and mortality rates differ. The BBC. Retrieved from https://www.bbc.com/future/ article/20200401-coronavirus-why-death-and-mortality-rates-differ

Hess RG Jr, Weaver SH, Speroni KG. Shared Governance During a Pandemic. Nurse Lead. 2020 Oct;18(5):497-499. doi: 10.1016/j.mnl.2020.05.008. Epub 2020 Jul 11. PMID: 32837348; PMCID: PMC7351658.

Hussain S, Baxi H, Chand Jamali M, Nisar N, Hussain MS. Burden of diabetes mellitus and its impact on COVID-19 patients: A meta-analysis of real-world evidence. Diabetes Metab Syndr. 2020 Nov-Dec;14(6):1595-1602. doi: 10.1016/j.dsx.2020.08.014. Epub 2020 Aug 20. PMID: 32862098; PMCID: PMC7439970.

Najaf Iqbal, Zeeshan Fareed and Farrukh Shahzad et al. Nexus between COVID-19, temperature and exchange rate in Wuhan City: New findings from Partial and Multiple Wavelet Coherence. *Sci. Total Environ.* DOI: 10.1016/j.scitotenv.2020.138916

Khan, I., Haleem, A., & Javaid, M. (2020). Analyzing COVID-19 pandemic through cases, deaths, and recoveries. *Journal of Oral Biology and Craniofacial Research*, *10*(4), 450-469.

Mandal, I., & Pal, S. (2020). COVID-19 pandemic persuaded lockdown effects on environment over stone quarrying and crushing areas. *Science of The Total Environment*, *732*, 139281.

Ozyigit A. Understanding Covid-19 transmission: The effect of temperature and health behavior on transmission rates. Infect Dis Health. 2020 Nov;25(4):233-238. doi: 10.1016/j.idh.2020.07.001. Epub 2020 Jul 8. PMID: 32674969; PMCID: PMC7345381.

Saha J, Chouhan P. Indoor air pollution (IAP) and pre-existing morbidities among under-5 children in India: are risk factors of coronavirus disease (COVID-19)? Environ Pollut. 2020 Nov;266(Pt 2):115250. doi: 10.1016/j.envpol.2020.115250. Epub 2020 Jul 15. PMID: 32693324; PMCID: PMC7362832.

Shahbazi F, Khazaei S. Socio-economic inequality in global incidence and mortality rates from coronavirus disease 2019: an ecological study. New Microbes New Infect. 2020 Nov;38:100762. doi: 10.1016/j.nmni.2020.100762. Epub 2020 Sep 16. PMID: 32953124; PMCID: PMC7492853.

Sharma A, Borah SB, Moses AC. Responses to COVID-19: The role of governance, healthcare infrastructure, and learning from past pandemics. J Bus Res. 2021 Jan; 122:597-607. doi: 10.1016/j.jbusres.2020.09.011. Epub 2020 Sep 10. PMID: 33518844; PMCID: PMC7834581.

Shen X, Cai C, Li H. Socioeconomic restrictions slowdown COVID-19 far more effectively than favorable weather-evidence from the satellite. Sci Total Environ. 2020 Dec 15; 748:141401. doi: 10.1016/j.scitotenv.2020.141401. Epub 2020 Jul 31. PMID: 32798876; PMCID: PMC7836816.

Topcu M, Gulal OS. The impact of COVID-19 on emerging stock markets. Financ Res Lett. 2020 Oct; 36:101691. doi: 10.1016/j.frl.2020.101691. Epub 2020 Jul 10. PMID: 32837378; PMCID: PMC7348595.

Wang, L., Li, J., Guo, S., Xie, N., Yao, L., Cao, Y., Day, S.W., Howard, S.C., Graff, J.C., Gu, T., Ji, J., Gu, W., & Sun, D. (2020). Real-time estimation and prediction of mortality caused by COVID-19 with patient information-based algorithm. *The Science of the Total Environment, 727*, 138394 - 138394.

Witteveen D. Sociodemographic inequality in exposure to COVID-19-induced economic hardship in the United Kingdom. Res Soc Stratif Mobil. 2020 Oct; 69:100551. doi: 10.1016/j.rssm.2020.100551. Epub 2020 Sep 6. PMID: 32921869; PMCID: PMC7474909.

Yilmazkuday, Hakan, Changes in Consumption in the Early COVID-19 Era: Zip-Code Level Evidence from the U.S. (October 6, 2021). Journal of Risk and Financial Management, Forthcoming, Available at SSRN: https://ssrn.com/abstract=3658518 or http://dx.doi.org/10.2139/ssrn.3658518

Yilmazkuday H. COVID-19 and unequal social distancing across demographic groups. Regional Science Policy & Practice. 2020 Dec;12(6):1235–48. doi: 10.1111/rsp3.12329. Epub 2020 Sep 18. PMCID: PMC7461490.

Yilmazkuday H. COVID-19 spread and inter-county travel: Daily evidence from the U.S. Transp Res Interdiscip Perspect. 2020 Nov; 8:100244. doi: 10.1016/j.trip.2020.100244. Epub 2020 Oct 22. PMID: 34173479; PMCID: PMC7580684.

Yoo S, Managi S. Global mortality benefits of COVID-19 action. Technol Forecast Soc Change. 2020 Nov; 160:120231. doi: 10.1016/j.techfore.2020.120231. Epub 2020 Jul 27. PMID: 32834139; PMCID: PMC7383174.

Yu C, Lei Q, Li W, Wang X, Liu W, Fan X, Li W. Clinical Characteristics, Associated Factors, and Predicting COVID-19 Mortality Risk: A Retrospective Study in Wuhan, China. Am J Prev Med. 2020 Aug;59(2):168-175. doi: 10.1016/j.amepre.2020.05.002. Epub 2020 May 27. PMID: 32564974; PMCID: PMC7250782.

Appendix A1

Appendix Table A1. International Data

Sl. No.	Data Source	Data Link	The Focus of Data Source	Variables	Up-dates/ Frequency
1	WHO	https://covid19.who.int/?gclid=EAIaIQobChMItdfnyKbR6QIVWKSWCh1sXg7REAAYASAAEgK2sPD_BwE	The World Health Organization focuses on cumulative deaths and cases across regions and countries. The dataset is retrievable and demonstrated by using a dashboard in daily or weekly timelines. The data is classified in Country and WHO region-level data by new cases, cumulative cases, new deaths, and cumulative deaths.	Global situation of confirmed cases and deaths, Situation analysis by WHO region, and situation by country territory or area	Real-time daily
2	World-ometer	https://www.worldometers.info/coronavirus/	Worldometers focus on cumulative cases, deaths, and recovered. Their breakdown of cases, deaths, and recovery are classified in details. For instance, Global active cases of severe and mild condition, close cases of recovery and deaths, daily new cases, and daily deaths. Their dashboard of COVID-19 updates is categorized by region and country. Worldometer	Covid-19 cases, deaths, recovered, active cases, closed cases, daily cases, daily deaths, reported cases by country and territory,	Real-time daily

			visualizes their updates through the graph, and they are of various types of cumulative cases, deaths, and recovery. Worldometer analysis is very detailed as they provide useful information on the COVID-19 death rate, incubation period, and death by age, sex, and medical condition. The data type is of cumulative cases, deaths, and recovery of COVID-19 by country and region.		
3	Our World in Data	https://ourworldindata.org/coronavirus-data	Our world in data updates the status of COVID-19 on confirmed cases and deaths. This site has visualized updates of COVID-19 through various descriptive statistical tools. Trends of Cumulative confirmed cases and deaths by country and region, breakdown of trajectories of deaths by country and region, and many more categories are used in analyzing the trends of COVID-19 across the world. The dataset is of confirmed cases, and deaths are also shown by country and region.	Tests, cases and deaths by country, cases, and deaths globally, confirmed deaths by country and region, testing over time-relative to the time of the outbreak, confirmed cases by country and region, case fatality rate, comparison of data source, healthcare capacity, age structure, risk	Real-time daily

				factors, and comorbidities, comparison of COVID-19 cases and GDP per capita	
4	Euro-news	https://www.euronews.com/2020/05/25/covid-19-coronavirus-breakdown-of-deaths-and-infections-worldwide	Euronews focus on the latest numbers of cumulative cases and deaths by country. The data is shown using graphs, charts, and tables. The data can be sorted according to the severity of Covid-19 affected regions/countries. The source also allows a global breakdown of COVID-19 by deaths, active cases, and recovery rates into country-levels. Analysis of the COVID-19 trajectory was also shown in a precise order. The site mainly covers COVID-19 updates from the parts of Europe. The data is of Cumulative cases and deaths by country and Europe.	Number of cases, deaths, new deaths by country, breakdown of global COVID-19 cases, infections and death overtime, daily new deaths, and cases worldwide, deaths per one million of population, analysis of death by age group in Italy, analysis of deaths by country	Real-time daily
5	WHO	https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports	WHO's situation reports on COVID-19 focuses on assessing cumulative and regular cases and deaths across all regions. The report breaks down the situation in graphs and charts and provides a	Number of confirmed COVID-19 cases by country, territory or area, number of confirmed COVID-19	Real-time daily

			breakdown of surveillance across all the countries and regions. The report covers the cumulative cases and deaths by countries, territories, or areas across WHO regions.	cases, by date of the report and WHO region, countries, territories or areas with confirmed COVID-19 cases and deaths by WHO region.	
6	UNICEF	https://data.unicef.org/resources/data-to-inform-the-covid-19-response/	UNICEF focuses on the situation of children and their vulnerabilities to COVID-19. The information is shown under different categories. For instance, the situation can be shown by country, sex, residence, or wealth. The site also shows other various country-level data regarding children.	Sex, Residence, Wealth quintile	Global data tracker updated on a weekly basis; the case starter is updated on a regular basis
7	tableau	https://www.tableau.com/covid-19-coronavirus-data-resources	Tableau provides a global coronavirus tracker. The site focuses on confirmed cases and deaths, the spread of coronavirus, and the places impacted the most. The cumulative data is broken down in terms of countries, and the United States' data is broken down in terms of its states. The dataset is country-level and district-level data by race/ethnicity.	New positive cases, New deaths, new cases, reported deaths by country, confirmed cases and deaths by race, total cases	Real-time daily

8	HDX	https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases	This site's dataset includes province/state, country/region, last Update, confirmed, suspected, recovered, and deaths. The dataset is a time-series data. It is displayed by using a map. The dashboard contains total confirmed, global Death, U.S. state-level death, and recovered confirmed cases by country/region.	Province/State, Country/Region, Last Update, Confirmed, Suspected, Recovered, Deaths.	Real-time daily
9	Eu Open Data Portal	https://data.europa.eu/euodp/en/data/dataset/covid-19-coronavirus-data	The dataset holds the latest available public data on COVID-19 (EU/EEA and the U.K., worldwide). Every day a team of epidemiologists collects the latest figures. The dataset can be visualized by graph, table, or map. The documentation about the data collection method.	Daily situation updates of E.U., U.K., and Worldwide.	Real-time daily
10	1 point 3 Acres	https://coronavirus.1point3acres.com/en	1Point3Acres is a site that focuses on the international Covid-19 database. This site shows the information detailed Covid-19 information on any country as desired by the user. The dashboard displays the summary of total/new (cases, deaths, recovered, fatality rates) of the desired	Case Summary figures of country and international level: Cases, Deaths, Recovered, Fatality Case summary graphs of the desired	Real-time daily

			country and the world. Different time series graphs accompany the information to assist the user in better analysis. The site has a section on trends of cases and mobility that are mostly shown by graphs and numbers. The figures are well-defined, and the countries' graphs can be compared with others and world figures. The data is a congregate of Cumulative and time-series data of all countries are available at the international level/individual country level.	country level: Confirmed cases, death and recovered cases, Active cases Trends: Community mobility level of U.S./Canada only based on destination type, Coronavirus confirmed cases Trajectories by country/ Region/ territory, Active cases by country, Confirmed cases by State in the U.S.	
11	The Kaggle Novel Corona-virus Dataset	https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset	This dataset has daily level information on the number of affected cases, deaths, and recovery from 2019 novel coronavirus. This site contains a column description of the dataset, and it contains other country-level datasets too.	Observation date, Province/ state, country/region, the cumulative number of confirmed cases, deaths, and recovered	Irregular pattern
12	Ding Xiang Yuan	https://ncov.dxy.cn/ncovh5/view/en_pneumonia?from=dxy&source=	The dataset of DXY focuses on providing updates of active cases, total confirmed deaths, and recovery cases. The dataset is shown in	Active cases, total confirmed, deaths, recovered by country/regio	Real-time daily

		&link= &share=	a trajectory graph and table categorized by country/region, active cases, totally confirmed, total deaths, and total recovery cases. A trajectory graph of daily new cases and death are shown for China and the USA, Russia, France, Italy, Germany, Brazil, Spain, and the U.K. The data is a congregate of country/region-level time-series data.	n	
13	European Centre for Disease Prevention and Control (ECDC)	https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases	ECDC focuses on the geographic distribution of COVID-19 cases and death worldwide. The dataset can be retrieved in CSV, JSON, or XML format. The data file is the latest available public data on Covid-19. The dataset is broken down into the number of new cases and deaths reported per day and per country. The data is a congregate of Country/region level data by cases and deaths.	Reported cases, deaths, country, population	Real-time daily
14	COTA	https://covid19.rnbo.gov.ua/	This page is a dashboard of Covid-19 for Worldwide and in Ukraine. It focuses on confirmed deaths and recovered, active, and	Confirmed, deaths, recovered, existing, suspicion by region	Real-time daily

			suspected cases worldwide and in Ukraine. The statistics of worldwide and Ukraine are displayed on a map. The variables are categorized by region.		
15	United Nations	https://covid-19-data.unstatshub.org/search?groupIds=0439998c8542406ea5edefb5ad7ae23a	This page provides Global datasets about Covid-19. This page has all kinds of data from worldwide cases, deaths, and tests. This page has a filter option, and content type, source, and categories can filter it. The datasets are updated in an irregular pattern.	Content type, Source, Categories	Irregular pattern
16	Apple Mobility report	https://www.apple.com/covid19/mobility	The charts and CSV file on this site present the relative volume of directions requests per country/region, sub-region, or city compared to a baseline volume on January 13, 2020. This data is developed by counting the number of claims made to Apple Maps for directions in select countries/regions, sub-regions, and cities	Mobility trends	Real-time daily
17	Corona data scraper	https://coronadatascraper.com/#home	Corona Data Scraper extracts information from a variety of openly available world government data sources and curated datasets. This page provides a worldwide	Cases, Recovery, Tests, Deaths	Real-time daily

			dataset from various sources. Most of the datasets are time-series data. This page also cross-checks reports of cases, deaths, tests, and recovery cases for different countries. The dataset in JSON and CSV format. A total of 3,195 cross-check reports were generated. The datasets are in order of quality, from good to bad.		
18	Covid-19 public datasets	https://console.cloud.google.com/marketplace/browse?filter=solution-type:dataset&filter=category:covid19	This page contains datasets for Covid-19 research. It has datasets from various resources, and they are explained with additional details, overview, and sample of a particular dataset.	Overview of the dataset, the sample of the dataset, and additional details	Irregular pattern
19	National Institute of Health	https://datascience.nih.gov/covid-19-open-access-resources	This repository contains resources that have been aggregated by The Office of Data Science Strategy and posted for scientific and public health interests. The datasets are categorized by support, resource description, and data type.	Resource, Resource description, Data type, and NIH funded	Real-time daily
20	Covid-19 Data Portal	https://www.covid19dataportal.org/	The COVID-19 Data Portal enables researchers to upload, access, and analyze COVID-19 related reference data and	Sequences, Expression data, Proteins, Structures, Compounds, Targets,	Irregular pattern

			specialist datasets as part of the more extensive European COVID-19 Data Platform. It provides raw and assembled sequences related to the COVID-19 outbreak, SARS-CoV-2 and SARS-CoV data for drug target identification and prioritization, latest literature about SARS-CoV-2, and a variety of related resources for studying the SARS-CoV-2 coronavirus and the COVID-19 disease. The data is in an irregular pattern. The datasets are categorized by collection date, country, particular keywords.	Literature, related resources	
21	CSH	http:// covid19-interventions.com/	This page contains datasets from various public sources. The dataset describes the implemented NPIs for 54 countries; measures implemented at the subnational level (state, region, city) are also included. The dataset also represents the measure as found in the text data source. This page has a dashboard that displays graphic visualization of the dataset.	Id, Country, Three-letter country code, State, Region, Date, Themes, Categories, Sub-categories, Codes, Status, Comment, and Source	Real-time daily

22	Bing	http://www.bing.com/covid/local/bangladesh	Bing COVID-19 data holds confirmed, fatal, and recovered cases from all regions. The dataset is updated daily in a .csv file and is updated according to correction or Update. To ensure the stability of the data is released with a 24-hour delay. Their datasets are filtered by country and displayed by using maps, charts, and graphs. They also present a table that displays COVID-19 statistics by region.	Spread over time, new cases by region, compare cases by region, distribution of confirmed cases, top 10 regions by active cases, confirmed cases, active cases, statistics by region	Real-time daily
23	Kaggle	https://www.kaggle.com/jcyzag/covid19-lockdown-dates-by-country#countryLockdowndates.csv	This page holds a list of countries and the dates that each country went into lockdown. The dataset comprises country/region, province, lockdown date, type of lockdown, and reference of the information.	Country/region, province, date of lockdown, type of lockdown, and reference of the information.	Irregular pattern
25	Face book	https://covid-survey.dataforgood.fb.com/?date=2020-06-26&dates=2020-04-17_2020-06-26&region=WORLD	This site includes a COVID-19 interactive map and dashboard. The map covers the worldwide Situation of COVID-19. COVID-19 symptoms can categorize it, Flu symptoms of total confirmed COVID-19 cases, New confirmed COVID-19 cases, the percentage of people staying in places, and	Symptoms and cases, physical distancing and risk factors	Real-time daily

			changes in movement. The data was collected by inviting people to participate in surveys about COVID-19 symptoms and risk factors and then use the survey data on a map and show other publicly available aggregate data to provide a more detailed view of the COVID-19 pandemic.		
26	Johns Hopkins univer-sity & Medi-cine (Corona virus Resour-ce Center)	https:// coronavirus .jhu.edu/ map.html	This site is a dashboard for COVID-19, which contains a world and U.S. map. The world map's dashboard includes total confirmed, global deaths, U.S. state-level deaths, and recovered, confirmed cases by country/region/ sovereignty. The U.S. map consists of the top 50 confirmed cases by county, top 20 countries by the number of deaths, confirmed by population, confirmed cases, fatalities, and fatality rate dataset categorized by states/territories/ county.	Word Map-total confirmed, global deaths, U.S. state-level deaths and recovered, confirmed cases by country/regio n/ sovereignty U.S. map- top 50 confirmed cases by county, top 20 countries by number of deaths, confirmed by population, confirmed cases, deaths, and fatality rate dataset categorized by states/ territories/ county	Real-time daily

27	Covid-19 Data Hub	https://covid19datahub.io/index.html	The goal of this COVID-19 Data Hub is to provide the research community with a unified data hub by collecting worldwide fine-grained case data merged with exogenous variables to better understand COVID-19. The COVID-19 data hub covers around 180 countries. The dataset highlights on tests, confirmed cases, deaths, recovery, hospitalization date, ventilation date, and ICU date.	cumulative number of tests, the cumulative number of confirmed cases, the cumulative number of deaths, the cumulative number of recovered, hospitalized on a date, requiring ventilation on a date, ICU (intensive therapy on date)	Real-time daily
28	Github (microsoft/Bing covid-19 data)	https://github.com/microsoft/Bing-COVID-19-Data	This site is the repository for Bing COVID-19 data. Bing COVID-19 data includes confirmed, fatal, and recovered cases from all regions, updated daily in a CSV file.	confirmed, fatal, and recovered cases from all regions	Real-time daily
29	Yahoo	https://yahoo.github.io/covid-19-dashboard/#/Earth	This page presents a dashboard of Yahoo for Covid-19. It displays global cases and deaths over time. This site has a map that shows the global situation by providing fatality cases. This dashboard focuses on global confirmed cases, fatalities, and fatality cases broken down by countries and states.	Global Confirmed cases, confirmed cases by country, cases over time, daily change, and 7-day moving average	Real-time daily

30	Github (Yahoo COVID-19 data)	https://github.com/yahoo/covid-19-data	This site presents Yahoo Knowledge Graph COVID-19 Datasets that feed into Yahoo properties like Yahoo News, Yahoo Finance, and Yahoo Weather. The COVID-19 datasets include country, state, and county-level information updated on a rolling basis, with updates occurring approximately hourly. The COVID-19 datasets are developed entirely from primary (government and public agency) sources. The data is logically organized by region and time. It provides general information about the areas covered in the dataset, such as geographic location, detailed case counts of COVID-19 in each region on any given date in local time, and the latest figures for each area.	Geographic location, case counts of COVID-19 by region, and latest COVID-19 figures for each region	Real-time daily
31	nCoV2019.live	https://ncov2019.live/data	This site is a data dashboard for World Covid-19 statistics. The dashboard illustrates cumulative variables of Covid-19 for the world, all regions, USA, Canada, Australia, Italy, Ireland, Russia,	Cumulative Confirmed, Critical, Deceased, Active, and Recovered, Vaccines in Development	Real-time daily

			Europe, Asia, Africa, Oceania, South America, and North America. Every country/region is categorized according to its state. The dashboard is updated daily.		
32	Google Data studio	https://datastudio.google.com/u/0/reporting/a529e043-e2b9-4e6f-86c6-ec99a5d7b9a4/page/yY2MB?s=ho2bve3abdM	This site is a dashboard for COVID-19 mobility. The dashboard shows COVID-19 trends over time. The dataset is categorized by country, region, and sub-region. The dataset is updated daily.	Residential percent change from baseline, parks percent change from baseline, grocery and pharmacy percent change from baseline, retail and recreation percent change from baseline, transit stations percent change from baseline, workplaces percent change from baseline	Real-time daily

Source: Own compilation