***Big Data Analytics for Tracking COVID-19 Spread and Mortality Trends Across Regions***

**Abstract**

Each geographic region has seen significant changes in daily life, economics, and global health because of the COVID-19 pandemic. Significant regional differences in the virus’s transmission and fatality have surfaced despite massive crisis management measures. It is essential to comprehend these distinctions to enhance future pandemic preparedness and public health responses. Using Big Data analytics approaches, this paper attempts to investigate the global prevalence and death trends of COVID-19 across all continents.

A structured data pipeline that can clean, filter, and visualise time-series data on confirmed COVID-19 cases and deaths by continent is built using publicly accessible datasets. Data wrangling and analysis are carried out using Jupyter Notebook and Python programming, which makes it possible to identify peak infection periods, regional differences, and changing trends in the pandemic’s progression. By pointing out important continental patterns, pinpointing regions with high death rates, and providing comparative data, the study helps determine the scope and timing of COVID-19’s effects. These finding are presented in an understandable manner using data visualisation like line graphs and heat maps.

Through a better knowledge of the global dynamics of COVID-19, our research helps the global health community make data-driven decisions and improves readiness for future pandemics.

**Background of the study**

Novel coronavirus (COVID-19) is highly infectious and requires early detection, isolation, and treatment. We tried to find some useful information by analysing the COVID-19 screening data, so as to provide help for clinical practice​Big Data. The novel coronavirus Covid-19 originated in China in early December 2019 and has rapidly spread to many countries around the globe, with the number of confirmed cases increasing every day. Covid-19 is officially a pandemic. It is a novel infection with serious clinical manifestations, including death, and it has reached at least 124 countries and territories​.

Starting in December 2019, cases of pneumonia with unknown causes began to appear in Wuhan, Hubei province, China. Subsequently, the outbreak of this pneumonia quickly spread throughout the Hubei province, country, and world. This pneumonia was confirmed to result from a novel coronavirus infection according to whole-genome sequencing. On January 13, 2020, the World Health Organization tentatively named the virus as 2019 novel coronavirus (2019-nCoV). On February 7, 2020, China officially named this novel coronavirus pneumonia as NCP. Later, on February 11, 2020, the World Health Organization officially renamed the NCP as coronavirus disease 2019 (COVID-19)​Big Data.

The COVID-19 epidemic has caused a large number of human losses and havoc in the economic, social, societal, and health systems around the world. Controlling such epidemic requires understanding its characteristics and behaviour, which can be identified by collecting and analysing the related big data. Big data analytics tools play a vital role in building knowledge required in making decisions and precautionary measures.

Modelling the COVID-19 pandemic spread is challenging. But there are data that can be used to project resource demands. Estimates of the reproductive number (R) of SARS-CoV-2 show that at the beginning of the epidemic, each infected person spreads the virus to at least two others, on average. A conservatively low estimate is that 5% of the population could become infected within 3 months. Preliminary data from China and Italy regarding the distribution of case severity and fatality vary widely. About 15% of patients infected with COVID-19 have severe illness and 5% have critical

illness​.

Despite massive crisis management measures, significant regional differences in the virus’s transmission and fatality have surfaced. It is essential to comprehend these distinctions to enhance future pandemic preparedness and public health responses​Big Data.

Reliable and real-time analysis using big data techniques became critical for understanding virus hotspots, mortality risks, and intervention effects. Using big data analytics tools for combining available information of an infectious disease process enables researchers and policymakers to transform such information into practical knowledge, helping to detect and predict disease epidemics.

A regular monitoring and remote detection system for individuals assists in the fast-tracking of suspected COVID-19 cases. Such systems generate a huge amount of data, providing many opportunities for applying big data analytics tools. These tools are designed to operate in a cloud computing and distributed environment to assist in the development of scalable pandemic responses​.

Analysing health data in real-time using artificial intelligence (AI) techniques plays a vital role in predictive and preventive healthcare. For example, it helps predict the sites of infection and the flow of the virus. It also helps estimate the needs for hospital beds, medical specialists, and resources during such pandemic crises as well as in the diagnosis and characterization of the virus.

Although the ultimate course and impact of COVID-19 are uncertain, it is not merely possible but likely that the disease will produce enough severe illness to overwhelm the worldwide health care infrastructure. Emerging viral pandemics can place extraordinary and sustained demands on public health and health systems and on providers of essential community services. Modelling the COVID-19 pandemic spread is challenging. But there are data that can be used to project resource demands​.

The rapid spread of the pandemic, with its continuous evolving patterns and the difference in its symptoms, makes it more difficult to control. Moreover, the pandemic has affected health systems and the availability of medical resources in several countries around the world, contributing to the high death rate. A regular monitoring and remote detection system for individuals will assist in the fast-tracking of suspected COVID-19 cases. Moreover, using such systems will generate a huge amount of data, which will provide many opportunities for applying big data analytics tools​.

Because of its protracted asymptomatic period and virulence, COVID-19 can spread quickly unless strategic precautions are taken. This infection period would also be marked with an asymptomatic characteristic, meaning a host is infected but no symptoms are presenting. Public health efforts depend heavily on predicting how diseases such as those caused by COVID-19 spread across the globe. During the early days of a new outbreak, when reliable data are still scarce, researchers turn to mathematical models that can predict where people who could be infected are going and how likely they are to bring the disease with them.

Existing studies have used various data sources to quantify human movement and model the spread of infectious diseases. On a large scale, airline data are important sources in understanding global transmission of infectious diseases. For example, global spread of SARS simulation models has been generated with airline data (). Although airline data deepened our understanding of the transmission mechanism of infectious diseases at large geographical scales, the data have shown a limited usefulness for understanding transmission across short distances (). On a local or regional scale, mobile phone data have been used as a measurement of human mobility; such data improved our understanding of spatial transmission patterns of malaria (), cholera (), and influenza (). Due to privacy issues, mobile phone data are generally limited in terms of accessibility and are often limited to a local region or one country; therefore, this data cannot provide systematic global coverage (). Besides mobile phone data, commuting patterns derived from census data also play an important role in understanding virus spread patterns on a local scale ()().

One of the key challenges in managing the pandemic has been the dynamic and regionally varied nature of its spread and impact. Understanding when and where the virus surged, how mortality rates evolved, and how different regions responded provides critical insight into managing current and future pandemics. However, making sense of massive, multi-source datasets requires big data analytics​.

This report aims to provide a comparative analysis of COVID-19 spread and mortality across all continents. By leveraging time-series data from reliable public sources, the study seeks to highlight trends, disparities, and commonalities in how the pandemic unfolded globally.

The contributions of this report include:

* Data acquisition from global repositories (e.g., Johns Hopkins, WHO)
* Data cleaning and preparation for time-series analysis
* Visualization of COVID-19 cases and death trends by continent
* Identification of peak periods and regional impact

**Definition, Characteristics, and Prevention Difficulties of Covid-19**

Public health incidents refer to sudden outbreaks of major infectious diseases, mass unexplained diseases, major food and occupational poisoning, and other events that have caused or may cause serious damage to the public health. The World Health Organisation (WHO) defines “unexplained public health incidents of international concern” as unusual public health risks to other countries through the international spread of disease that may require coordinated international event response. The WHO warning is divided into six major levels, from level one to level six, with level six being the highest level, indicating that the epidemic is spreading globally (a pandemic). Following this, each country also has their own policies; for example, China divides natural disasters, accidental disasters, and public health incidents into particularly important (level I), significant (level II), large (level III), and general (level IV) according to factors such as the degree of social harm and the scope of the influence.

The outbreak of major public health incidents usually has the characteristics of suddenness, uncertainty, unpredictability, high hazard, high social concern, chain reaction, timely disposition, and evasive prevention (). The explanations of these of COVID-19 are shown in Table 1.

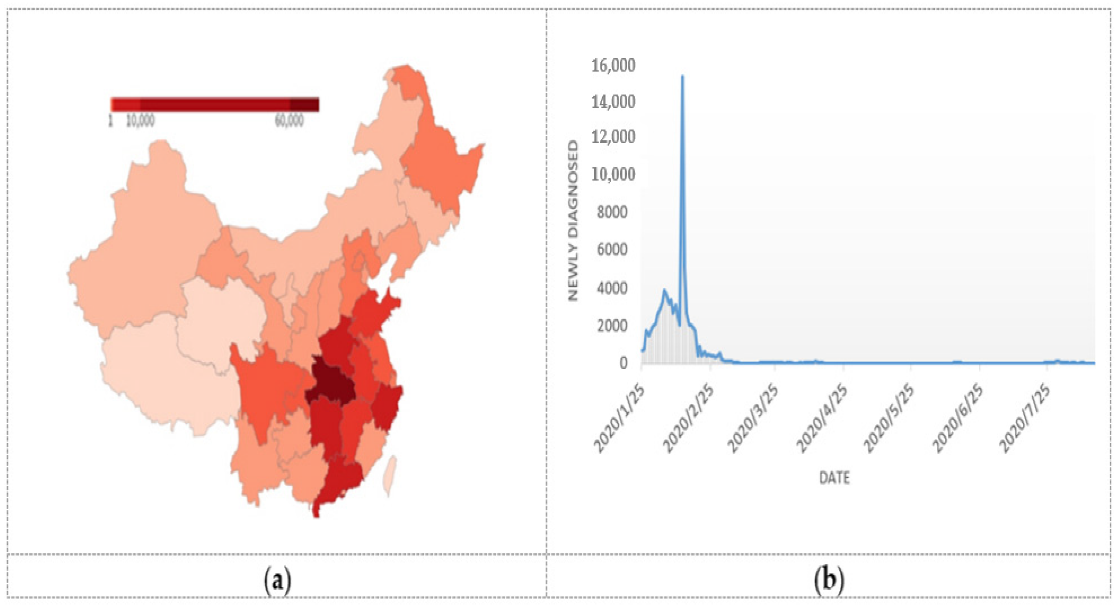
Table 1. Characteristics of COVID-19 incidents. WHO: World Health Organization

A white and black text on a white background

AI-generated content may be incorrect.

Outbreaks of viruses such as Covid-19 are arduous to prevent. A key feature is in the fact that is an unknown, new type of coronavirus. People have very limited previous knowledge that is relevant. It also has high uncertainty and unpredictability. Since the virus is extremely contagious, with a high degree of suddenness and a strong chain reaction, it has caused a high degree of social attention due to its rapid spread around the world. Although its rate of mortality is about 5.6% lower than SARS, its spreading ability (diagnosis rate of close contacts) is very high, resulting in a high number of infections. Therefore, it is incumbent upon governments to deal with it in a timely fashion to avoid spread on a larger scale (). In addition, although transportation brings convenience to citizens, in the case of limited cognition and a lack of effective control measures, it can also accelerate the spread of virus and infect large numbers of people with high death incidence in a short period of time. For example, from 20 January 2020 to 08 March 2020, 80,735 accumulated confirmed cases were reported in China, 58,600 cases were cured, 3119 persons had died from COVID-19, and 674,760 close contacts were tracked. The number of close contacts observed was 20,146 ().

Faced with the emergency situation, the Chinese government responded quickly and enforced strict controls and timely resource deployment to minimize the loss and gradually control the epidemic. As shown **Figure 1**, as of July 2020, under the policy intervention of the Chinese government with big data applications, the number of newly confirmed cases continues to decrease, while in other parts of the world, the confirmed cases do not yet appear to be controlled effectively, as shown by the number of confirmed cases in **Figure 2**. Overall, the global pandemic situation is not yet optimistic. By 18 August, 21,826,342 accumulated confirmed cases were reported in the world, 13,888,301 cases were cured, and 773,152 persons had died from COVID-19 (). Fighting against COVID-19 has become a long-term topic for the whole world, which is also transforming the situations of economics, politics, and societies.



**Figure 1.** (**a**) Cumulative confirmed cases of COVID-19 pandemic in China [[**13**](https://www.mdpi.com/1660-4601/17/17/6161#B13-ijerph-17-06161)]; (**b**) The development trend of COVID-19 pandemic in China: New diagnosed cases [[**13**](https://www.mdpi.com/1660-4601/17/17/6161#B13-ijerph-17-06161)].

A map of the world

AI-generated content may be incorrect.

**Figure 2.** Global confirmed cases of the COVID-19 pandemic [[**14**](https://www.mdpi.com/1660-4601/17/17/6161#B14-ijerph-17-06161)]. (**a**) COVID-19 pandemic across the globe; (**b**) Global daily new cases of the COVID-19 [[**14**](https://www.mdpi.com/1660-4601/17/17/6161#B14-ijerph-17-06161)].

**Visual Analysis of COVID-19 Spread and Mortality Trends Across Continents Using Big Data Charts.**

The outbreak of COVID-19 in late 2019 and its rapid global spread through the marked one of the most significant public health crises in the recent history. Unlike seasonal influenza and pneumonia, COVID-19 demonstrated a far steeper and more deadly trajectory in many regions. According to the Office for National Statistics (ONS), weekly deaths in England and Wales due to COVID-19 quickly surpassed those from flu and pneumonia combined by early April 2020, highlighting the virus’s acute impact on the public health systems as shown in the **Figure 3** (BBC News ,2020).

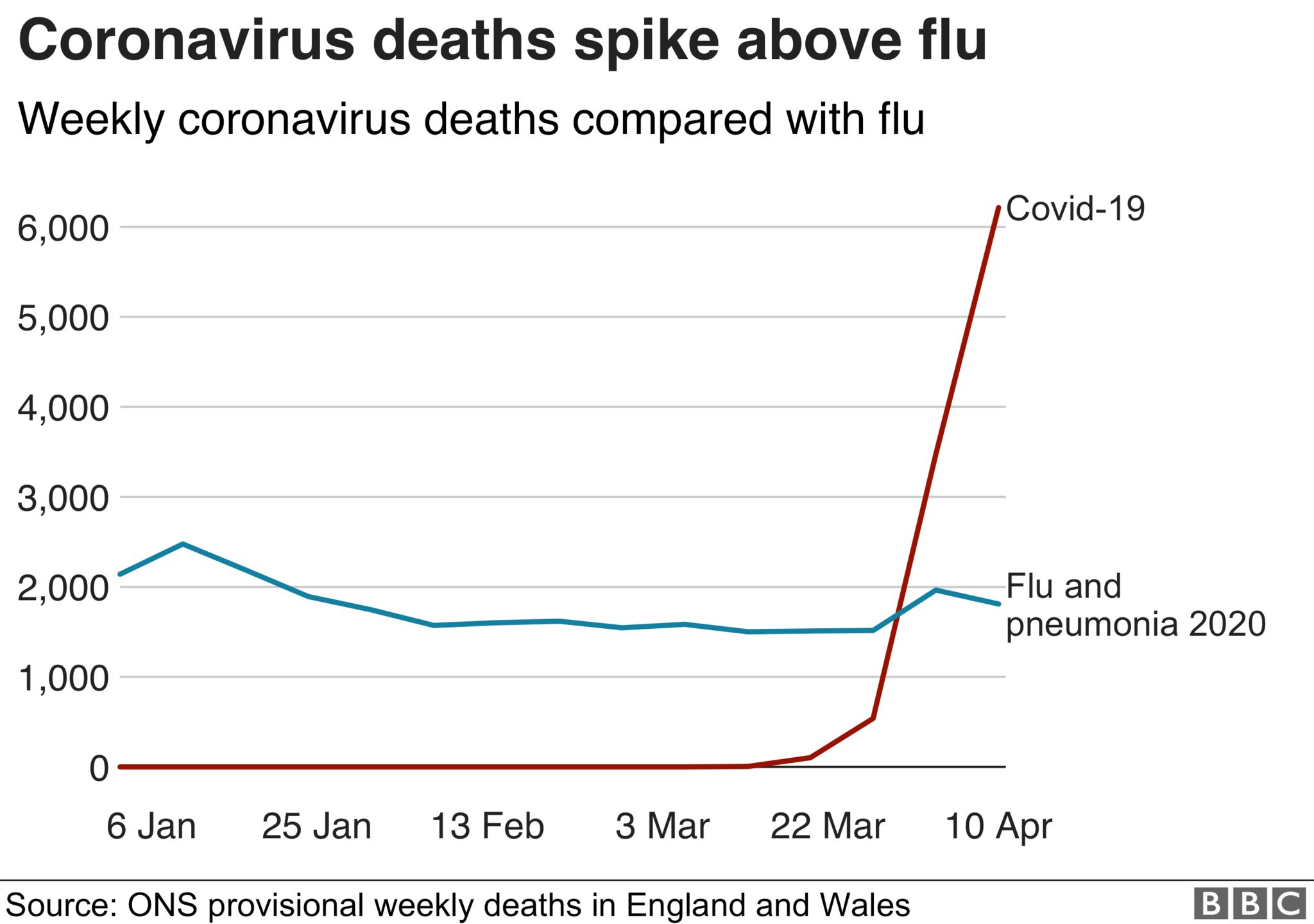
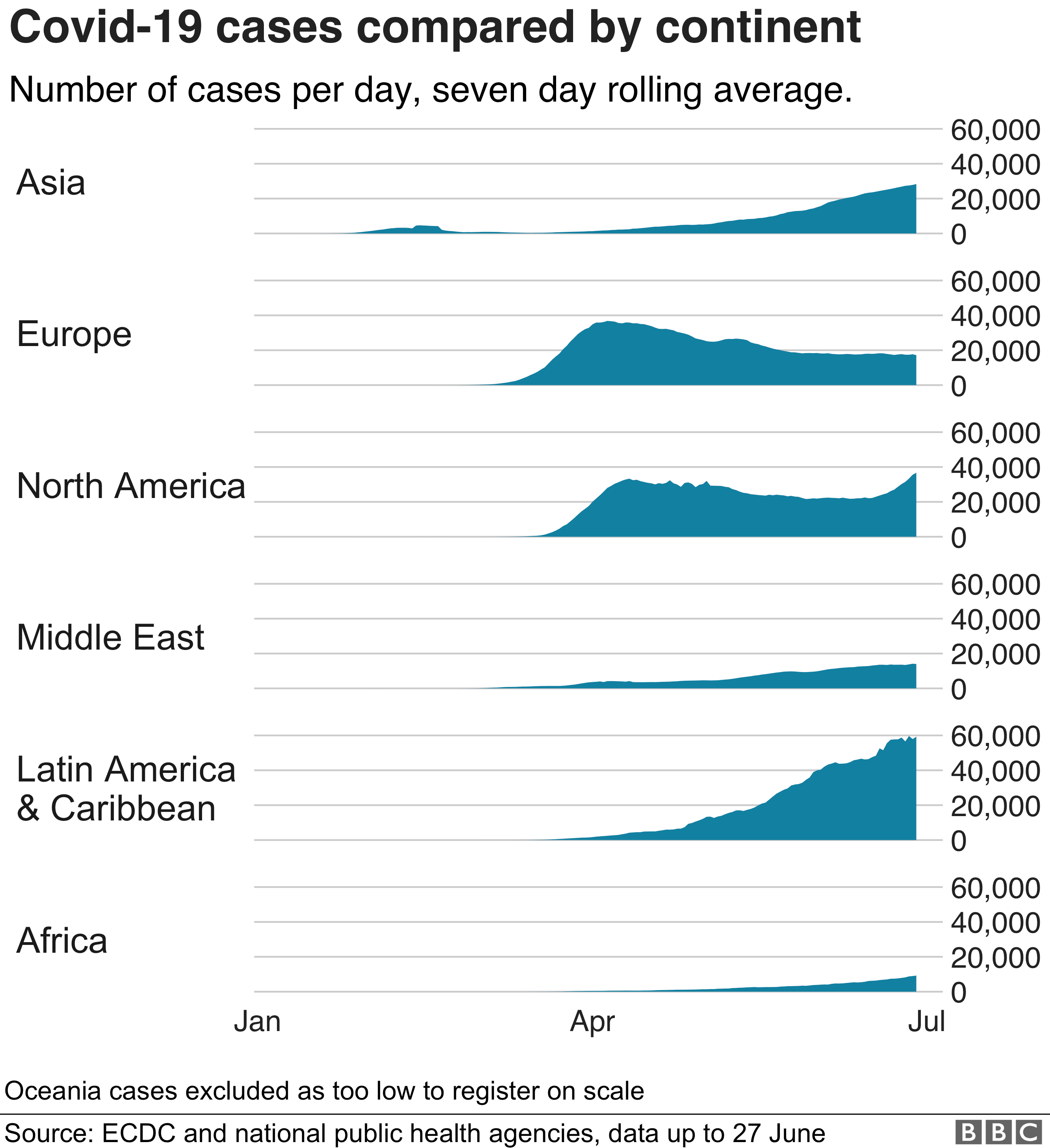


Figure 3: Weekly COVID-19 deaths vs Flu and Pneumonia deaths. Source: BBC News, 2020

Globally, the pandemic’s progression varied by continent, with Europe and North America experiencing sharp rises in daily cases early on, followed by significant increases in Latin America and Asia by mid-2020 as shown in the **Figure 4 & 5** (BBC News,2020).



**Figure 4:** COVID-19 cases by continent.

A map of the world with different colored countries/regions

AI-generated content may be incorrect.

**Figure 5**

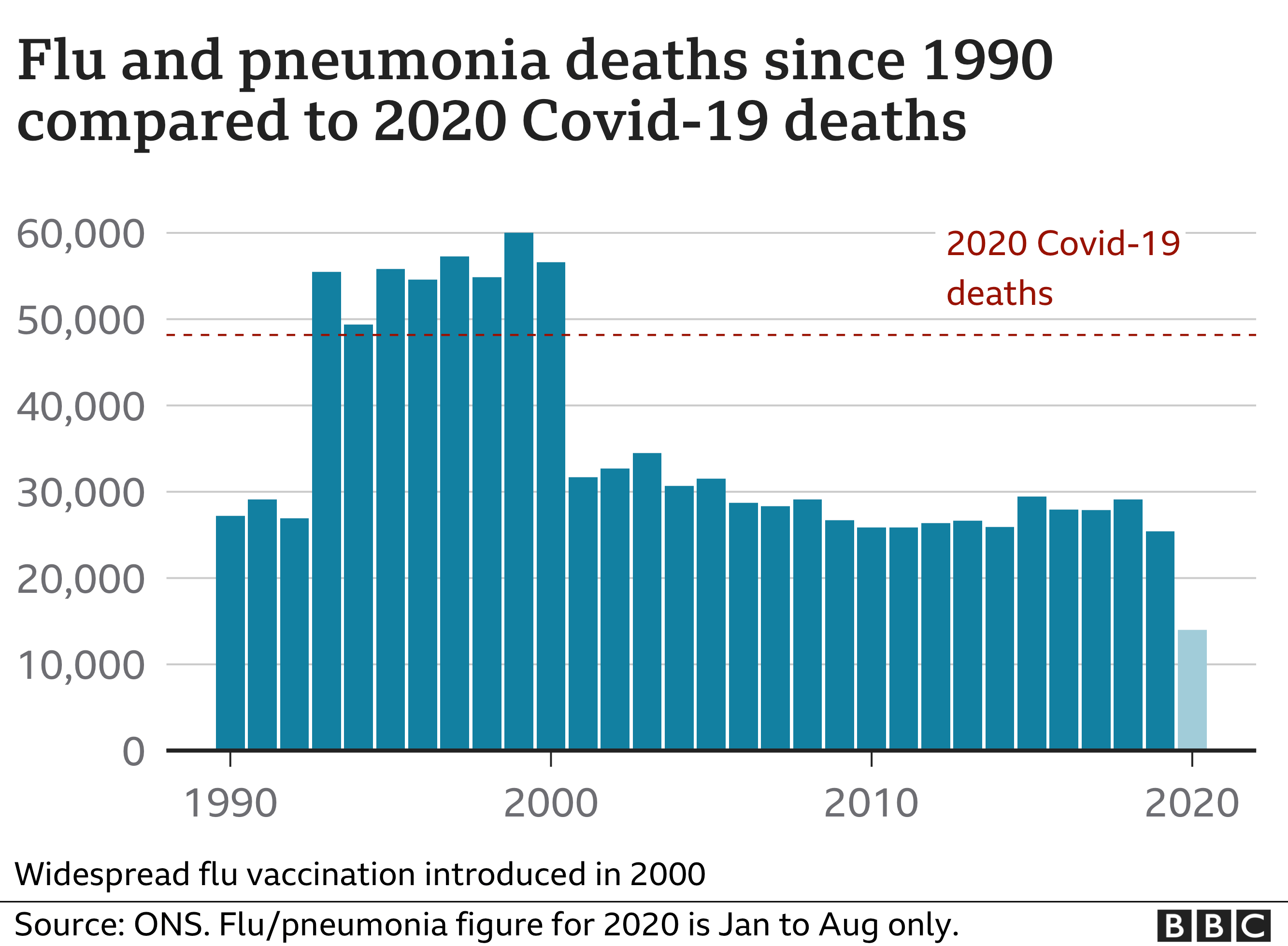
The United Sates, Brazil and India recorded the highest absolute numbers of the COVID-19 deaths, reflecting the vast scale of the pandemic’s toll as shown in the **Figure 6** (BBC News, 2020).

A graph of a number of people with red squares

AI-generated content may be incorrect.

**Figure 6:** Top 5 countries with the most COVID-19 deaths.

When comparing the total mortality, the number of deaths from COVID-19 in 2020 alone exceeded the annual flu and pneumonia deaths in the UK for many prior years, even before the introduction of mass vaccinations as shown in the Figure 7(BBC News, 2020). Excess mortality statistics further underscored the impact, with countries like Peru and Russia showing exceptionally high rates of excess deaths per 100,000 people, significantly above the global average as shown in the **Figure 8** (BBC News,2021).



**Figure 7:** Flu and pneumonia deaths since 1990 vs 2020 COVID deaths.

A graph of different countries/regions

AI-generated content may be incorrect.

**Figure 8:** Excess deaths per 100,00 in 2020 and 2021.

`1`

**Related Work**

The unprecedented scale and impact of the COVID-19 pandemic prompted a rapid and expansive body of research in the area of big data analytics. Researchers and practitioners across the globe have turned to big data to develop methods for real-time monitoring, understanding patterns of spread, predicting mortality, and supporting healthcare decision-making. The following section provides an in-depth review of existing literature relevant to our study, focusing on ten significant works that represent a comprehensive cross-section of the current landscape. Each of these studies contributes uniquely to the collective understanding of how big data has been used to track the spread of COVID-19 and analyse its mortality trends across regions.

1. **Applications of Big Data Analytics to Control COVID-19 Pandemic**

This foundational work emphasizes the role of big data in pandemic response, especially in the early stages of outbreak detection, population movement analysis, and healthcare system preparedness. It provides an overview of how government agencies and health institutions utilized real-time data collection from various sources such as GPS tracking, public health databases, and online activity. The article highlights practical implementations in South Korea and China where surveillance, contact tracing, and policy decision-making were driven by analytics tools built upon massive datasets.

In relation to our study, this work sets the stage for understanding how large datasets were not only gathered but actively used to mitigate risk and optimize resources. It also discusses ethical considerations in using sensitive data for the public good, which forms a crucial part of our analysis when evaluating cross-regional implementation. Our approach builds on this work by focusing on how these applications varied across regions and how they influenced mortality trends.

A diagram of a health care procedure

AI-generated content may be incorrect.

**Figure 9:** Potential application areas of big data analytics for COVID-19.

As illustrated by a diagram in the original article, the diagnostic applications include COVID-19 pre-symptomatic detection, detecting COVID-19 cases, tracking symptoms, analysing medical characteristics, and monitoring discharged patients. Risk prediction involves determining care levels, patient priority, psychological impacts, defining infectious places, estimating epidemic spread, and contact tracing. Healthcare decision-making includes determining care and medical resource needs, assessing health staff needs, and deciding on quarantine measures. Pharmaceutical applications focus on accelerating drug discovery and development, improving clinical trials, more effectively targeting patient samples, gaining better patient behaviour insights, improving risk management, and enhancing marketing insights.

1. **Big Data Driven COVID-19 Pandemic Crisis Management**

This paper takes a crisis management perspective, exploring how big data supported different phases of pandemic response—from detection and decision-making to response coordination and recovery planning. It categorizes big data applications into strategic, tactical, and operational levels, aligning them with pandemic stages such as outbreak escalation, peak, and de-escalation.

The strategic use of predictive analytics and scenario modelling is particularly relevant to our study. For instance, this paper analyses how dashboards and early-warning systems, developed using data from hospitals, transportation networks, and social media, were employed by national governments. Our work extends this research by focusing on the consistency and disparity of these tools across multiple regions, especially in relation to mortality rates and outcomes. Moreover, the emphasis on integrating unstructured and structured data offers a methodological framework that we incorporate into our own comparative analysis.

A diagram of a data processing process

AI-generated content may be incorrect.

**Figure 10:** Big data technology for COVID-19 pandemic

1. **Big Data Analytics Show COVID-19 Spread, Outcomes by Region**

This source provides one of the most direct contributions to the topic of our report. It explores regional disparities in both the spread of COVID-19 and corresponding health outcomes using big data analysis. The authors deploy spatial analysis and machine learning models to examine how socioeconomic factors, healthcare access, and policy interventions correlate with infection and mortality rates across various regions. The findings reveal significant inequalities, highlighting those areas with lower socioeconomic status faced disproportionately higher infection and mortality rates. This underscores the urgent need for targeted health policies that address these disparities and improve healthcare access for vulnerable populations.

What sets this work apart is its emphasis on geospatial data and the use of regional heatmaps to visualise disparities. These visual representations not only highlight areas of concern but also facilitate targeted interventions by policymakers. By identifying high-risk regions, the study aims to inform strategies that could mitigate the impact of future health crises. It identifies those areas with higher population densities and lower healthcare resources had significantly higher mortality rates, even when controlling for age and comorbidities. Our study builds upon these insights by conducting a deeper multi-dimensional analysis, integrating variables such as vaccination rates, testing capacity, and public compliance, which were not extensively covered in this work.

1. **Meaningful Big Data Integration for a Global COVID-19 Strategy**

This article explores the challenge of integrating diverse datasets for a coordinated global response to the pandemic. It advocates for the harmonization of clinical, epidemiological, mobility, and economic data to enable more holistic decision-making. The authors discuss interoperability challenges and propose a framework for international data collaboration, particularly through open data portals and federated learning systems.

In the context of our research, this paper is crucial in highlighting the fragmentation of data across regions. While our study does not seek to implement a unified global model, we address similar challenges when comparing datasets from different countries and regions, especially when standardization practices vary. This necessitates a careful approach to data integration and analysis, ensuring that findings are both accurate and relevant within the specific cultural and operational contexts. Ultimately, our research aims to contribute to the ongoing discourse on how to effectively navigate these complexities while still providing meaningful insights. The concept of meaningful integration also aligns with our methodology, which involves aligning regional data sources for comparative mortality and spread analysis. By leveraging advanced analytical techniques and fostering collaboration among local stakeholders, we can enhance the robustness of our findings. This collaborative effort not only enriches our understanding of the data but also promotes a sense of ownership among the regions involved, thereby facilitating more effective public health interventions.

1. **Demystifying COVID-19 Mortality Causes with Interpretable Data Mining**

This study applies interpretable data mining methods to uncover the underlying risk factors and feature combinations associated with COVID-19 mortality, particularly among elderly patients. Using a retrospective cohort of 1,917 COVID-19 patients admitted to Xiangya Hospital between December 2022 and March 2023, association rule mining was employed to systematically identify leading causes of death. The researchers first used Affinity Propagation clustering to extract key features from the dataset, followed by the Apriorist Algorithm to discover six groups of abnormal feature combinations that significantly increased mortality risk. Rigorous data preprocessing ensured high data integrity: patients with less than 75% data completeness were excluded, missing values were handled via multiple imputations, and only patients diagnosed with COVID-19 at admission were included. Deceased patients were further categorized based on whether death occurred during hospitalization or after treatment withdrawal and discharge, with careful cross-verification of outcomes using medical records to ensure accurate death labelling. The resulting association rules linked specific clinical features and patient profiles with increased mortality, providing actionable insights for clinicians and public health officials.

This work stands out for its comprehensive and interpretable approach to analysing COVID-19 mortality. Unlike many studies that focus solely on individual risk factors, this research reveals not only the independent effects but also the interactive effects of multiple variables on patient outcomes. By leveraging interpretable data mining techniques, the study offers transparent and clinically meaningful rules that can guide targeted interventions, resource allocation, and patient management strategies for high-risk groups, especially the elderly.

1. **Effects of COVID-19 Pandemic on Life Expectancy and Premature Mortality in 2020: Time Series Analysis in 37 Countries**

This longitudinal study analyses changes in life expectancy and premature mortality across 37 countries during the year 2020. Using national statistics and health records, it finds significant declines in life expectancy in countries with high COVID-19 death rates. The authors attribute these trends to both direct and indirect pandemic effects, including healthcare disruption and mental health deterioration. These findings underscore the profound impact of the pandemic on global health, highlighting the need for robust public health strategies to mitigate long-term consequences. Furthermore, the study calls for increased investment in healthcare systems to enhance resilience against future health crises.

This work enhances our understanding of mortality trends at a macro level. While we focus on regional differences, we draw on this paper to contextualize our findings within broader national and international mortality patterns. Furthermore, the time-series methods used in this study are mirrored in our own trend analysis, allowing us to identify not just levels of mortality but also their progression over time in different areas. This progression is crucial for tailoring interventions that address the specific needs of each region. By recognising these disparities, policymakers can allocate resources more effectively and implement targeted health initiatives that aim to mitigate the adverse effects of mortality trends.

**Changes in life expectancy in 2020**

In all the countries between 2005 and 2019, an increasing trend was observed in life expectancy at birth, both in men and women (supplementary figure S3). However, most countries showed a reduction in life expectancy in 2020, with the largest overall reduction in life expectancy at birth (in years) in Russia (−2.32, 95% confidence interval −2.55 to −2.11), the US (1.98, −2.16 to 1.82), Bulgaria (−1.75, −2.09 to −1.41), Lithuania (−1.61, −1.92 to −1.29), and Poland (−1.36, −1.55 to −1.17). Reductions in life expectancy in Italy, Spain, and England and Wales were −1.35 (−1.72 to −0.99), −1.27 (−1.57 to −0.99), and −1.02 (−1.27 to −0.78), respectively. In contrast, a gain in life expectancy was observed in New Zealand (0.66, 0.41 to 0.89) and Taiwan (0.35, 0.14 to 0.54); no evidence was found of a change in life expectancy in South Korea (0.11, −0.09 to 0.30), Norway (0.07, −0.03 to 0.17), or Denmark (−0.09, −0.24 to 0.06).

A graph of different numbers

AI-generated content may be incorrect.

**Figure 11:** changes in life expectancy in 2020 across countries by Gender

**Changes in years of life lost in 2020**

Years of life lost declined in most countries in both men and women between 2005 and 2019, except Canada, Greece, Scotland, Taiwan, and the US. The observed YLL in 2020 was higher than expected in all countries except Taiwan and New Zealand, where there was a reduction in YLL, and Iceland, South Korea, Denmark, and Norway, where there was no evidence of a change in YLL in 2020. In the remaining 31 countries, more than 222 million (130 million in men and 92.6 million in women) years of life were lost in 2020, which is 28.1 million (95% confidence interval 26.8m to 29.5m) YLL higher than expected. The excess YLL in men and women were 17.3 million (16.8m to 17.8m) and 10.8 million (10.4m to 11.3m), respectively.

A chart of different countries/regions

AI-generated content may be incorrect.

**Figure 12: Trends** in Years of Life Lost (YLL) per 100,000 by Gender Across Countries (2005–2020)

1. **Applications of Big Data analytics to control COVID-19 pandemic**

The study titled *"Applications of Big Data Analytics to Control COVID-19 Pandemic"* explores how big data technologies were pivotal in managing the COVID-19 crisis, especially in the early stages of the outbreak. It outlines various data sources—including electronic health records, mobile location data, social media activity, and public health databases—that were harnessed to monitor virus transmission, identify hotspots, and support public health interventions. The paper highlights case studies from countries like China and South Korea, where real-time contact tracing and population surveillance were enabled through integrated big data platforms. These systems allowed health authorities to track the movement of infected individuals, issue timely alerts, and allocate medical resources more efficiently. The study also discusses how machine learning algorithms were employed to predict case surges, assess patient risk levels, and model the effects of intervention strategies. Importantly, the paper raises ethical concerns related to privacy and data security, emphasizing the need for responsible data governance. In the context of our research, this work serves as a foundational reference for understanding the core applications of big data in pandemic response. It provides critical insights into how large-scale data collection and analytics can influence both infection control and mortality outcomes. Our study builds upon this by focusing on cross-regional comparisons and examining how the deployment and effectiveness of such big data applications varied depending on regional healthcare infrastructure and policy implementation.

1. **COVID-19 Outbreak Data Analysis and Prediction Modeling Using Data Mining Technique**

progression of the COVID-19 pandemic and to build predictive models for outbreak forecasting. The study utilizes real-world datasets, including daily reported cases, death rates, and recovery counts, to apply techniques such as classification, clustering, and regression. Through these methods, the authors identify key trends in the virus’s spread and develop models capable of short-term forecasting of infection rates. The work highlights how data mining enables the discovery of hidden patterns in vast datasets, such as correlations between demographic factors and infection severity or the impact of lockdown measures on case trends. Furthermore, the study evaluates the performance of different algorithms (e.g., decision trees, support vector machines, and random forests) in terms of accuracy, precision, and computational efficiency. This research is particularly relevant to our study as it provides a methodological foundation for modelling both spread, and mortality trends based on historical data. While the paper concentrates mainly on case prediction, it reinforces the role of data mining in early warning systems and outbreak management. Our research extends this approach by applying similar techniques across multiple regions to investigate how regional dynamics and healthcare capacity influence the relationship between spread and mortality. Additionally, we explore the integration of these techniques with other big data sources, such as mobility and clinical data, to enhance prediction accuracy and provide region-specific insights.

A graph showing a number of different colored lines

AI-generated content may be incorrect.

**Figure 13:** Visualizing of Worldwide COVID-19 cases

1. **Integration of Clinical and Open Data for COVID-19 Mortality and Spread Analysis**

The article provides an in-depth exploration of how mortality data related to COVID-19 can be accurately interpreted and utilized for epidemiological insights. The study emphasizes that mortality statistics are among the most reliable indicators for assessing the true impact of the pandemic, as they are less affected by testing coverage or asymptomatic underreporting. It outlines key challenges in mortality data interpretation, such as the differentiation between deaths caused directly by COVID-19 and those where the virus was a contributing factor, along with the need for standardizing death certification processes across regions.

A central contribution of the paper is its discussion on statistical and machine learning models used to analyse mortality patterns. The authors evaluate methods such as time series forecasting, excess mortality estimation, and survival analysis. These techniques are employed to assess not only temporal trends but also the effects of demographic variables, healthcare capacity, and pre-existing conditions on COVID-19 fatality outcomes. The paper also highlights the importance of incorporating non-medical data—such as socioeconomic indicators and regional healthcare infrastructure—to improve model accuracy and policy relevance.

From a comparative standpoint, the article sheds light on regional disparities in mortality rates and attempts to model the underlying factors contributing to these differences. This aligns closely with the goals of our study, which investigates how big data analytics can track COVID-19 spread and mortality trends across various regions. The paper’s focus on methodological rigor and interpretability reinforces our approach, particularly in applying explainable models that enable policymakers to understand not just what the trends are, but why they exist. It also supports our multi-layered analysis strategy, combining epidemiological, demographic, and infrastructural data to derive actionable insights on regional mortality dynamics.

A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

AI-generated content may be incorrect.

**Figure 14:** Global Comparison of COVID-19 Mortality Rates per Million Population by Country and Region (as of 14th July 2020)

Deaths per million population for 55 countries selected to represent five main socioeconomic and geo-political groups from around the world. Sweden, Brazil and Belarus are highlighted as countries that have pursued a policy of no lockdown. The two graphs differ only in the logarithmic versus linear scale of their vertical axis, allowing us to put in perspective the relative death toll across different countries

**References**

1. R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding Lancet, 395 (2020) (2019), pp. 565-574
2. World Health Organization, WHO Director-General’s remarks at the media briefing on 2019-nCoV on 11 February 2020 [EB/ OL], <https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020> (accessed 2 February 2020).
3. Y. Liu, Y. Yang, C. Zhang, C. Zhang, F. Huang, F. Wang, *et al.* Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury Sci. China Life Sci., 63 (2020), pp. 364-374, [10.1007/s11427-020-1643-8](https://doi.org/10.1007/s11427-020-1643-8)
4. X.W. Xu, X.X. Wu, X.G. Jiang, K.J. Xu, L.J. Ying, C.L. Ma, *et al.* Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series BMJ, 27 (2020), p. 368
5. C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China Lancet, 15 (2020), p. 395
6. C. Rothe, M. Schunk, P. Sothmann, G. Bretzel, G. Froeschl, C. Wallrauch, *et al.* Transmission of 2019-nCoV infection from an asymptomatic contact in Germany N. Engl. J. Med., 382 (2020), pp. 970
7. J.F. Chan, S. Yuan, K.H. Kok, K.K. To, H. Chu, J. Yang, *et al.* A familial cluster of pneumonia associated with the novel coronavirus indicating person-to-person transmission: a study of a family cluster Lancet, 395 (2020) (2019), pp. 514-523, [10.1016/s0140-6736(20)30154-9](https://doi.org/10.1016/s0140-6736(20)30154-9)
8. Y. Huang, M. Tu, S. Wang, S. Wang, S. Chen, W. Zhou, *et al.* Clinical characteristics of laboratory confirmed positive cases of SARS-CoV-2 infection in Wuhan, China: a retrospective single center analysis
9. Travel Med Infect. Dis. (2020), p. 101606, [10.1016/j.tmaid.2020.101606](https://doi.org/10.1016/j.tmaid.2020.101606)
10. Hufnagel L, Brockmann D, Geisel T. Forecast and control of epidemics in a globalized world. Proc Natl Acad Sci U S A 2004 Oct 19;101(42):15124-15129 [[FREE Full text](http://www.pnas.org/cgi/pmidlookup?view=long&pmid=15477600)] [[CrossRef](https://dx.doi.org/10.1073/pnas.0308344101" \t "_blank)] [[Medline](https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=15477600&dopt=Abstract)]
11. Brockmann D, David V, Gallardo A. Human mobility and spatial disease dynamics. In: Schuster HG, editor. Reviews of Nonlinear Dynamics and Complexity, Volume 2. Weinheim, Germany: Wiley; 2009:24.
12. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. Proc Natl Acad Sci U S A 2009 Dec 22;106(51):21484-21489 [[FREE Full text](http://www.pnas.org/cgi/pmidlookup?view=long&pmid=20018697)] [[CrossRef](https://dx.doi.org/10.1073/pnas.0906910106" \t "_blank)] [[Medline](https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=20018697&dopt=Abstract)]
13. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, et al. Quantifying the impact of human mobility on malaria. Science 2012 Oct 12;338(6104):267-270 [[FREE Full text](http://www.sciencemag.org/cgi/pmidlookup?view=long&pmid=23066082)] [[CrossRef](https://dx.doi.org/10.1126/science.1223467" \t "_blank)] [[Medline](https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=23066082&dopt=Abstract)]
14. Finger F, Genolet T, Mari L, de Magny GC, Manga NM, Rinaldo A, et al. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks. Proc Natl Acad Sci U S A 2016 Jun 07;113(23):6421-6426 [[FREE Full text](http://www.pnas.org/cgi/pmidlookup?view=long&pmid=27217564)] [[CrossRef](https://dx.doi.org/10.1073/pnas.1522305113" \t "_blank)] [[Medline](https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=27217564&dopt=Abstract)]
15. Tizzoni M, Bajardi P, Decuyper A, Kon Kam King G, Schneider CM, Blondel V, et al. On the use of human mobility proxies for modeling epidemics. PLoS Comput Biol 2014 Jul 10;10(7):e1003716 [[FREE Full text](https://dx.plos.org/10.1371/journal.pcbi.1003716)] [[CrossRef](https://dx.doi.org/10.1371/journal.pcbi.1003716" \t "_blank)] [[Medline](https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=25010676&dopt=Abstract)]
16. Wesolowski A, Buckee C, Engø-Monsen K, Metcalf C. Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data. J Infect Dis 2016 Dec 01;214(suppl\_4):S414-S420 [[FREE Full text](http://europepmc.org/abstract/MED/28830104)] [[CrossRef](https://dx.doi.org/10.1093/infdis/jiw273" \t "_blank)] [[Medline](https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=28830104&dopt=Abstract)
17. Dalziel BD, Pourbohloul B, Ellner SP. Human mobility patterns predict divergent epidemic dynamics among cities. Proc Biol Sci 2013 Sep 07;280(1766):20130763 [[FREE Full text](http://europepmc.org/abstract/MED/23864593)] [[CrossRef](https://dx.doi.org/10.1098/rspb.2013.0763" \t "_blank)] [[Medline](https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=23864593&dopt=Abstract)]
18. Wang, J. Research on the Model and Effectiveness Evaluation of Emergency Plan System. Ph.D Thesis, Dalian University of Technology, Dalian, China, 2014. [[**Google Scholar**](http://scholar.google.com/scholar_lookup?title=Research+on+the+Model+and+Effectiveness+Evaluation+of+Emergency+Plan+System&author=Wang,+J.&publication_year=2014)]
19. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, *395*, 497–506. [[**Google Scholar**](http://scholar.google.com/scholar_lookup?title=Clinical+features+of+patients+infected+with+2019+novel+coronavirus+in+Wuhan,+China&author=Huang,+C.&author=Wang,+Y.&author=Li,+X.&author=Ren,+L.&author=Zhao,+J.&author=Hu,+Y.&author=Zhang,+L.&author=Fan,+G.&author=Xu,+J.&author=Gu,+X.&publication_year=2020&journal=Lancet&volume=395&pages=497%E2%80%93506&doi=10.1016/S0140-6736(20)30183-5)] [**[CrossRef](http://doi.org/10.1016/S0140-6736(20)30183-5" \t "_blank)**] [[**Green Version**](http://www.thelancet.com/article/S0140673620301835/pdf)]
20. Chan, J.F.W.; Yuan, S.; Kok, K.H.; To, K.K.W.; Chu, H.; Yang, J.; Xing, F.; Liu, J.; Yip, C.C.Y.; Poon, R.W.S.; et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *Lancet* **2020**, *395*, 514–523. [[**Google Scholar**](http://scholar.google.com/scholar_lookup?title=A+familial+cluster+of+pneumonia+associated+with+the+2019+novel+coronavirus+indicating+person-to-person+transmission:+A+study+of+a+family+cluster&author=Chan,+J.F.W.&author=Yuan,+S.&author=Kok,+K.H.&author=To,+K.K.W.&author=Chu,+H.&author=Yang,+J.&author=Xing,+F.&author=Liu,+J.&author=Yip,+C.C.Y.&author=Poon,+R.W.S.&publication_year=2020&journal=Lancet&volume=395&pages=514%E2%80%93523&doi=10.1016/S0140-6736(20)30154-9)] [**[CrossRef](http://doi.org/10.1016/S0140-6736(20)30154-9" \t "_blank)**] [[**Green Version**](http://www.thelancet.com/article/S0140673620301549/pdf)]
21. Xu, X.W.; Wu, X.X.; Jiang, X.G.; Xu, K.J.; Ying, L.J.; Ma, C.L.; Li, S.B.; Wang, H.Y.; Zhang, S.; Gao, H.N.; et al. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: Retrospective case series. *BMJ* **2020**, *368*, 606. [[**Google Scholar**](http://scholar.google.com/scholar_lookup?title=Clinical+findings+in+a+group+of+patients+infected+with+the+2019+novel+coronavirus+(SARS-Cov-2)+outside+of+Wuhan,+China:+Retrospective+case+series&author=Xu,+X.W.&author=Wu,+X.X.&author=Jiang,+X.G.&author=Xu,+K.J.&author=Ying,+L.J.&author=Ma,+C.L.&author=Li,+S.B.&author=Wang,+H.Y.&author=Zhang,+S.&author=Gao,+H.N.&publication_year=2020&journal=BMJ&volume=368&pages=606&doi=10.1136/bmj.m606)] [**[CrossRef](http://doi.org/10.1136/bmj.m606" \t "_blank)**] [[**Green Version**](https://www.bmj.com/content/bmj/368/bmj.m606.full.pdf)]
22. National Health Commission of the People’s Republic of China. The Latest Situation of New Coronavirus Pneumonia. Available online: [**http://www.nhc.gov.cn/xcs/yqtb/list\_gzbd.shtml**](http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml) (accessed on 16 August 2020)
23. Sina News. Real-Time Dynamic Tracking of the Novel Coronavirus Pneumonia Epidemic. Available online: [**https://news.sina.cn/zt\_d/yiqing0121**](https://news.sina.cn/zt_d/yiqing0121) (accessed on 16 August 2020).

Related work

1. Alsunaidi, S.J., Almuhaideb, A.M., Ibrahim, N.M., Shaikh, F.S., Alqudaihi, K.S., Alhaidari, F.A., Khan, I.U., Aslam, N. and Alshahrani, M.S., 2021. Applications of big data analytics to control COVID-19 pandemic. *Sensors*, *21*(7), p.2282. (Figure 1)
2. Lv, Y., Ma, C., Li, X. and Wu, M., 2021. Big data driven COVID-19 pandemic crisis management: potential approach for global health. *Archives of Medical Science: AMS*, *17*(3), p.829.
3. Kent, J., 2020. Big data analytics show covid-19 spread, outcomes by region. *HealthITAnalytics. Retrieved February*, *15*, p.2021.
4. Costa, J.P., Grobelnik, M., Fuart, F., Stopar, L., Epelde, G., Fischaber, S., Poliwoda, P., Rankin, D., Wallace, J., Black, M. and Bond, R., 2020. Meaningful big data integration for a global COVID-19 strategy. *IEEE Computational Intelligence Magazine*, *15*(4), pp.51-61.
5. Qian, X., Zuo, Z., Xu, D., He, S., Zhou, C., Wang, Z., Xie, S., Zhang, Y., Wu, F., Lyu, F. and Zhang, L., 2024. Demystifying COVID-19 mortality causes with interpretable data mining. *Scientific Reports*, *14*(1), p.10076.
6. Islam, N., Jdanov, D.A., Shkolnikov, V.M., Khunti, K., Kawachi, I., White, M., Lewington, S. and Lacey, B., 2021. Effects of covid-19 pandemic on life expectancy and premature mortality in 2020: time series analysis in 37 countries. *bmj*, *375*.