

# The application of biostatistics in COVID-19 epidemic

**Zixi Zhang**

School of AI and advanced computing, Xi'an Jiaotong Liverpool University, Suzhou, Jiangsu, China, 215127

Zixi.Zhang20@student.xjtlu.edu.cn

**Abstract.** Between the years 2020 and 2022, the COVID-19 pandemic is anticipated to emerge as the most severe global epidemic. The objective of this study is to examine the utilization of biostatistics in the domains of medication development, analysis of epidemic trends, and survival model analysis within the context of the COVID-19 pandemic. Through the utilization of a literature review method, this research delved into the examination of prospective therapeutic interventions employed in the realm of drug development studies. Specifically, the paper explored the efficacy and potential of camostat mesylate and remdesivir, alongside the exploration of immunotherapeutic strategies. Furthermore, the study examines the use of mathematical modeling in forecasting the trajectory of epidemic dissemination, and the significance of survival model analysis in comprehending patient longevity. The study revealed that medication development and immunotherapy play a crucial role in effectively combating novel coronavirus pneumonia. Furthermore, the utilization of mathematical modeling can provide valuable insights into forecasting the propagation of the epidemic. Additionally, survival model analysis can offer guidance in the allocation of medical resources and aid in decision-making processes. The findings of this research will contribute to a deeper comprehension and more effective mitigation of worldwide public health issues, such as the ongoing COVID-19 pandemic.

**Keywords:** COVID-19, Biostatistics, Models, Survival Analysis.

## 1. Introduction

In the year 2020, a novel strain of coronavirus emerged and exhibited a rapid global dissemination, leading to a consequential outbreak of pneumonia. This particular strain has since been recognized as the most severe pandemic of the 21st century [1]. The illness is characterized by symptoms resembling those of influenza, such as cough, fever, exhaustion, and dyspnea. In recent times, there has been a proliferation of the novel Omicron sub-variant EG.5 of the SARS-CoV-2 virus, resulting in the identification of infection cases across multiple geographical locations. In 2020, when the virus initially emerged on a significant magnitude, governments globally implemented extraordinary stringent measures, encompassing blockades, quarantines, widespread immunization campaigns, and obligatory adherence to mask-wearing protocols. Healthcare systems globally are currently facing significant strain, as healthcare professionals exert relentless efforts to assure the delivery of adequate treatment and care to patients. Biostatistics has significantly contributed to addressing the challenges posed by the COVID-19 pandemic. Its application has been crucial in aiding governmental bodies, health organizations, and researchers in several aspects, such as forecasting and modeling the outbreak,

conducting epidemiological inquiries, assessing vaccine efficacy, and formulating evidence-based public health strategies to enhance population well-being. Comprehend, mitigate, and address this pervasive international public health emergency.

Biostatistics is a fundamental discipline within the field of biology that use statistical models and methodologies to examine biological data, thereby elucidating patterns and interconnections in the realm of biology. Commonly employed biostatistical approaches encompass linear regression analysis, logistic regression analysis, survival analysis, and other related techniques. This study aims to critically examine the utilization of biostatistics in the context of the COVID-19 pandemic, focusing on four key areas: epidemiological testing, medication development experiments, analysis of epidemic trends, and survival model analysis. In a broader context, this research has the potential to make significant theoretical contributions to relevant academic disciplines, while also offering practical support in effectively managing pandemics. The analysis of the COVID-19 epidemic contributes to the enhancement of comprehension of this public health crisis. Furthermore, it facilitates the development of more effective strategies and instruments for response efforts. Additionally, the examination of this current situation yields valuable experiences and lessons that may be used to future scenarios involving comparable issues. The study holds value not only within the context of the present outbreak but also offers insights for broader global public health and epidemiological studies.

## **2. Epidemiological test**

The initial segment will center on the epidemiological investigation of COVID-19. Epidemiological research involves multiple facets, such as the investigation of transmission mechanisms of the novel coronavirus, the genetic variations of the virus, and the identification and evaluation of efficient control strategies. The primary mode of transmission for COVID-19 is by respiratory droplets. When an individual who is infected with a virus engages in activities such as coughing, sneezing, or talking, it results in the release of virus particles into the surrounding air. These particles have the potential to be inhaled by other individuals. Moreover, transmission of the virus can occur by direct contact with contaminated surfaces, subsequently leading to contact with the oral, nasal, or ocular regions. According to existing research, hospitals have been identified as a significant contributor to the secondary transmission of COVID-19, mostly due to the substantial influx of infected persons seeking admission. The transmission of the COVID-19 virus through contamination within patient rooms caring for individuals infected with COVID-19 has been documented in reports. In a recent study conducted by Santarpia et al., researchers collected surface samples from rooms accommodating patients who tested positive for SARS-CoV-2. The objective was to obtain evidence of viral RNA presence. The investigation revealed that commonly used goods, including washroom facilities, as well as air samples, exhibited positive results for SARS-CoV-2 [2]. The study collected multiple samples and found that individuals who received oxygen by nasal cannula had the highest virus concentration in terms of airborne transmission.

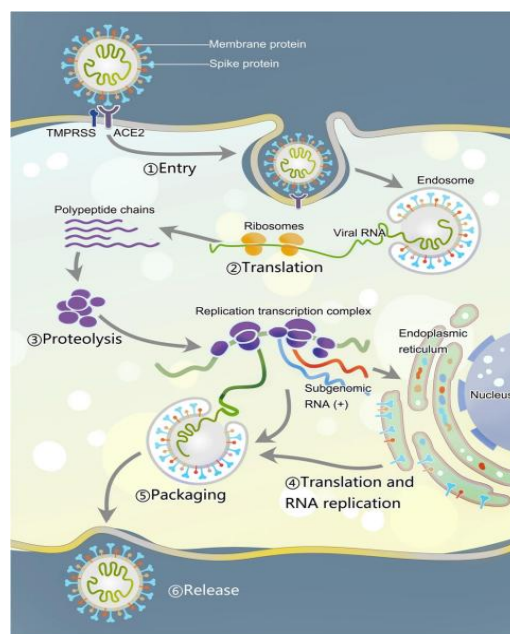
The ongoing mutation of SARS-CoV-2 has resulted in the appearance of several variations. Mutations play a fundamental role in the process of evolution, enabling natural selection to promote advantageous characteristics in the virus, such as heightened virulence, adaptability, and evolutionary progression [3]. Certain variations demonstrate increased transmissibility and have the potential to influence the efficacy of vaccinations. The variants are being actively monitored by researchers, who are continually upgrading vaccinations to effectively respond to emerging virus strains. According to a recent study conducted by Tang et al., there is evidence to imply that SARS-CoV-2 is undergoing continuing evolutionary changes [4]. The present work conducted a comparative analysis of the genomes of SARS-CoV-2 obtained from a cohort of 100 patients, consisting of 73 individuals from Wuhan and 27 individuals from regions outside of Wuhan. A comprehensive analysis of 100 genomes revealed the presence of 149 mutation sites, among which two single nucleotide polymorphisms (SNPs) exhibited a robust correlation. This finding suggests the existence of two different subtypes of the SARS-CoV-2 virus. The initial single nucleotide polymorphism (SNP) is situated at position 8782 (orf1ab: T8517C, synonymous), whereas the second SNP is found at position 28,144 (ORF8: C251T,

non-synonymous - S84L). Following that, the genomes were classified according to the second single nucleotide polymorphism (SNP), which determines whether the amino acid at position 84 in the open reading frame 8 (ORF8) is serine (S) or leucine (L). Through comparative genome study with closely related viruses, it has been determined that the S type represents the ancestral form. However, it has been observed that the L type is more prevalent, particularly in the region of Wuhan. The variants are being actively monitored by researchers, who are continually modifying vaccinations to accommodate emerging viral strains. This highlights the significance of closely monitoring and conducting research on viral variants.

### 3. Drug development experiments

The urgent and important challenge of developing medications for COVID-19 has been of paramount importance among the worldwide medical community since the onset of the epidemic. Since the initiation of the global pandemic, researchers have been diligently and continuously engaged in the pursuit of identifying efficacious therapeutic approaches. This paper aims to provide an overview of essential elements pertaining to experimental medication development for COVID-19.

To facilitate the development of appropriate pharmaceutical interventions, it is imperative to conduct a comprehensive investigation into the mechanism of viral transmission exhibited by the novel coronavirus. The COVID-19 virus, which has its origins in bats, induces pathogenesis by invading immune cells. The infection process is visually depicted in Figure 1. Hence, it is plausible that protease inhibitors and RNA polymerase inhibitors could exhibit favorable therapeutic efficacy against the virus. An example of a prospective treatment for COVID-19 is Camostat mesylate (CM), a protease inhibitor that was discovered in Japan during the 1980s. It has gained attention due to its ability to prevent the entry of SARS-CoV and HCoV-NL63 into HeLa cells [5]. In murine models of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), the administration of CM had a significant protective effect, resulting in a survival rate of 60% among the animals. The compound CM exhibits a plasma half-life of approximately 100 minutes, and its concentration in the plasma diminishes significantly within a time frame of 4 to 5 hours. Therefore, it is plausible that the regular administration of a 600mg dose of CM could serve as an efficacious method for managing SARSCOV-2 infections. One of the primary advantages of CM is its widespread utilization in Japan, owing to its perceived safety and cost-effectiveness. The proposed approach exhibits a high likelihood of efficacy in the treatment of COVID-19.



**Figure 1.** The membrane fusion, replication, packaging and release of SARS-Cov-2 [5].

Remdesivir (RDV) is an adenosine C-nucleotide analog with a 1'-cyano substitution, exhibiting the ability to disrupt viral transcription. This characteristic suggests that Remdesivir has a wide range of possible medical uses. In both tissue culture and animal investigations, therapeutic effects against SARS have been shown. These effects include a reduction in lung viral loads, improvement in clinical symptoms, and enhancement of respiratory function [6]. RDV has been extensively utilized in the United States and Europe as a therapeutic intervention in the battle against COVID-19. Notably, the initial case of COVID-19 in the United States was effectively managed with the administration of RDV on the seventh day of hospitalization.

Immunotherapy is a widely respected field of study within the realm of COVID-19 treatment. Corticosteroids have demonstrated efficacy in modulating autoimmune disorders and inflammatory processes, and their effectiveness has been confirmed in a range of indications, including severe instances of Acute Respiratory Distress Syndrome (ARDS) [7].

An alternative approach to immunotherapy is employing anti-cytokine therapies to mitigate the excessive activation of the host immune response triggered by potent proinflammatory cytokines. This intervention serves to avert detrimental consequences, including acute respiratory distress syndrome (ARDS) and multi-organ dysfunction, in individuals afflicted with COVID-19 [8]. Tocilizumab, a recombinant humanized monoclonal antibody, has been developed to specifically target both soluble and membrane-bound IL-6, hence potentially offering a viable therapeutic approach.

#### 4. Epidemic trends

To develop mathematical models for the estimation and prediction of the COVID-19 pandemic's spread, a study was conducted wherein three models were tested: the transfer vector model employing transfer functions as components, the AutoRegressive with eXogenous (ARX) model, and the AutoRegressive Moving Average with eXogenous (ARMAX) model [9], as shown in table 1. In order to evaluate the precision of these models, it is important to undertake certain phases within the data analysis procedure. The procedures involved in this process encompass several key tasks, such as confirming the temporal gaps between days and the accuracy of the information provided. Additionally, any missing data must be addressed and appropriately handled. Subsequently, the data is processed on a daily basis, ensuring that each day's information is properly analyzed. Consistent and reliable information is then selected, while any inconsistent data is filtered out and eliminated from consideration. Finally, the organized data is represented in the form of vectors or matrices for further analysis. Relevant input data may encompass many factors such as the quantity of tests administered, both indoor and outdoor temperatures, outdoor humidity levels, and quantifiable metrics pertaining to the COVID-19 containment measures implemented by governing bodies in order to limit the transmission of the pandemic.

**Table 1.** Examples of input data [9].

Date	Inside Temperature (°C)	Outside Temperature (°C)	Humidity (%)	No. of Tests	Quantified Value of COVID-19 Measures (%)
28 February 2020	21°C	10.714°C	72.286%	8	0%
28 March 2020	21.5°C	16.571°C	68.857%	86	86.85%
28 April 2020	19°C	22.857°C	71.571%	759	87%
28 May 2020	21.8°C	18°C	71.286%	1271	73%
28 June 2020	20.3°C	32°C	65.714%	1689	47.5%
28 July 2020	18.8°C	33.571°C	61.714%	2374	49.5%
28 August 2020	20.7°C	34°C	60%	3498	45.5%
28 September 2020	19.9°C	25°C	64%	4632	31.5%

Numerous studies have demonstrated that artificial intelligence (AI) technologies, particularly those employing neural networks, constitute a substantial component of the forthcoming advancements in various domains, including mathematical modeling. However, when dealing with dynamic systems that have multiple input signals, neural models frequently demonstrate a substantial level of complexity and necessitate considerable computer resources for the training process. Moreover, it is widely acknowledged that neural nonlinear models tend to exhibit superior performance compared to linear models, hence presenting additional complexities in their numerical execution. In contrast, it should be noted that Multiple-Input Single-Output (MISO) models based on Transfer Functions are characterized by linearity, hence facilitating their numerical implementation in a straightforward manner. Moreover, neural models possess inherent limits when confronted with training signals that are subject to interference and noise, as well as situations where there is an inadequate amount of available data or an overwhelming abundance of data. Classical modeling techniques are frequently employed to effectively tackle these circumstances.

Through the analysis of data pertaining to the COVID-19 pandemic and the development of suitable models, governmental bodies and healthcare institutions can enhance their comprehension of the dissemination patterns and trends associated with the virus. By employing comprehensive data gathering and rigorous analysis, it becomes possible to identify regions that are particularly susceptible to the pandemic. This knowledge facilitates the timely implementation of focused interventions, such as quarantine protocols and efficient vaccine distribution. Forecasting probable surges in the pandemic in the near future can be facilitated through the utilization of models that incorporate historical data and real-time information. This proactive approach assists in enhancing preparedness efforts by guaranteeing the availability of sufficient medical equipment, drugs, and staff. Additionally, this allows decision-makers to more precisely adjust preventive efforts, hence reducing unwanted economic and societal consequences.

## **5. Survival model analysis**

Survival analysis plays a crucial role in comprehending patient survival and illness progression within the framework of a pandemic, hence informing the distribution of medical resources and facilitating decision-making processes. The Cox proportional hazards model, Kaplan-Meier curves, and survival tree models are frequently employed techniques in the analysis of COVID-19 survival.

The Cox proportional hazards model is a frequently employed survival model that examines the duration of patient survival and the associated odds of survival. Within the framework of the COVID-19 pandemic, this particular model can be utilized to assess the influence of various elements on patient survival, encompassing age, gender, underlying health issues, and additional variables. This aids in the identification of high-risk populations for the implementation of early preventive actions. The Kaplan-Meier curves serve as a non-parametric technique utilized for the estimation of survival curves, commonly employed to assess and compare survival periods across distinct groups. This methodology can be employed to assess and contrast the survival rates among individuals who have had various treatment regimens or are located in disparate geographical areas.

Survival tree models, conversely, refer to decision tree models employed for the purpose of determining patient survival probabilities by considering various combinations of features. This approach can be employed to customize individualized treatment strategies, thereby enhancing patient outcomes in terms of survival rates. Survival analysis techniques are valuable tools for healthcare professionals as they seek to enhance their comprehension of patient survival during the COVID-19 pandemic. These techniques aid in optimizing the allocation of resources, refining treatment strategies, and furnishing scientific evidence to inform government decision-making in effectively addressing this global challenge.

## **6. Conclusion**

This work emphasizes the pressing need for drug development experiments in the context of the COVID-19 pandemic, encompassing a variety of potential therapeutic interventions, including

camostat mesylate and remdesivir, alongside immunotherapeutic strategies. This offers guidance in identifying efficacious therapy alternatives. While the study raises significant points, it exhibits major deficiencies in several aspects. The insufficiency of extensive clinical trial data necessitates further validation of the efficacy and safety of these prospective treatments. Furthermore, there is a lack of clarity in providing comprehensive and specific details regarding data sources, including their completeness and availability. Subsequent study should aim to provide a more thorough description of the quality and trustworthiness of the data. Ultimately, while survival models were acknowledged, there was a lack of comprehensive analysis on their practical use and constraints within the context of the COVID-19 pandemic. Future study should aim to conduct a more comprehensive investigation into the specific uses and enhancement techniques of these models. Future research should prioritize enhancing the research and development endeavors pertaining to novel pharmaceutical interventions for the treatment of coronary pneumonia. Additionally, it is imperative to conduct extensive clinical trials on a broad scale to ascertain the efficacy and safety of potential therapeutic agents. In the realm of mathematical modeling, it is imperative to do more research on more intricate models in order to enhance the precision of predicting the propagation and dynamic patterns of the epidemic. Simultaneously, the utilization of survival model analysis holds promise in informing medical resource allocation and treatment decisions. Future investigations can delve further into additional parameters to enhance the predictive efficacy of the model. These endeavors will contribute to a more effective approach in tackling worldwide public health issues, such as the ongoing COVID-19 pandemic.

## References

- [1] Mehta, A. & Hong, Q.-J. QJHong Model for Novel Coronavirus Disease 2019 (COVID-19) in the United States. *MedRxiv* 1 (2023).
- [2] Santarpia, J. L. et al. Transmission Potential of SARS-CoV-2 in Viral Shedding Observed at the University of Nebraska Medical Center. (2020) doi:<https://doi.org/10.1101/2020.03.23.20039446>.
- [3] Duffy, S. Why are RNA virus mutation rates so damn high? *PLOS Biology* 16, (2018).
- [4] Tang, X. et al. On the origin and continuing evolution of SARS-CoV-2. *National Science Review* 7, (2020).
- [5] Xiao, Y. et al. Update on treatment and preventive interventions against COVID-19: an overview of potential pharmacological agents and vaccines. *Molecular Biomedicine* 1, 5 (2020).
- [6] Sheahan, T. P. et al. Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses. *Science Translational Medicine* 9, (2017).
- [7] Hu, Y. et al. Prevalence and severity of corona virus disease 2019 (COVID-19): A systematic review and meta-analysis. *Journal of Clinical Virology* 127, (2020).
- [8] Xu, Z. et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *The Lancet Respiratory Medicine* 0, (2020).
- [9] Cristina-Maria Stancioi et al. Solution for the Mathematical Modeling and Future Prediction of the COVID-19 Pandemic Dynamics. *Applied sciences* 13, 6–8 (2023).