Analyzing the Accuracy of Reporting of SARS-CoV-2 Infections Across Six Countries

Hyunjin Lim

Abstract

The outbreak of the COVID-19 pandemic prompted a global response that involved hundreds of government bodies and international organizations. As part of this response, large amounts of data on infection rates, hospitalizations, and fatalities was reported. Verifying the accuracy, reliability, and consistency of reported data is important for establishing public trust and scientific consensus. Here, I search for inconsistencies in COVID-19 case reporting and analyze their various causes, including variations in testing capabilities, healthcare infrastructure, governmental policies, and data collection methodologies. I analyzed reported COVID-19 case data from South Korea, the United States, the United Kingdom, France, China, and Japan. For each nation, I compared time-course data on the number of positive cases, the number of deaths, and the per-case mortality rates to identify discrepancies. I found that almost all nations showed consistent trends in case rates, indicating global waves of infections. I also found a consistent decrease in the per-case mortality rate, possibly caused by access to better prophylactics and medicines along with SARs-CoV-2 evolving to be less virulent. However, we found massive anomalies in China's state reported data when compared with other nations. I think this is likely due to government censorship of infection data. I hope that the work I present here will help to increase confidence in reported data and will provide tools and strategies for international organizations to build trust in data collected from collaborating nations during a global pandemic.

Introduction

The COVID-19 epidemic led to a global response from governments and organizations.Sharing data on infection rates, hospitalizations, and deaths was crucial for coordinating an effective response to the pandemic. However, there have been concerns about the consistency, accuracy, and dependability of the data released.

It is impossible to overstate the importance of reliable data during a global health emergency. Valid and reliable data form the basis for global collaboration and response actions, affecting public health plans and high-level decision-making. However, without methods to check consistent data collection and reporting, it is difficult to ensure the truthfulness of such data.

In this paper, I search for consistencies in reported data among six countries: China, Japan, South Korea, France, the United States, and the United Kingdom. I compare time course data using Matplotlib to plot the number of positive cases, the number of deaths, and the per-case mortality rates in an attempt to find any possible patterns or anomalies. I compare trends in the data to various world events, such as the release of the first vaccines and the emergence of new SARS-CoV-2 variants. I also analyze data collection methods of each country in order to explain any observed anomalies.

The results of this study will help clarify the precision and correctness of COVID-19 data globally and offer international organizations strategies and tools for fostering confidence among cooperating countries in times of global health emergencies. For public health policies and decision-making to be effective, trust in the data provided is necessary. Our research should

increase confidence and strengthen the capacity of the international community to handle similar crises in the future.

Data and Methods

To evaluate the accuracy and precision of COVID-19 data supplied by different countries, I use a set of statistical techniques that compare epidemiological data from South Korea, the United States, the United Kingdom, France, China, and Japan. This analysis aims to assess the accuracy of the underlying data generating procedures as well as to find any conflicts between distinct pieces of data that may exist.

Data Sources

I have gathered COVID-19 data from multiple sources to ensure the robustness of our study. Our primary sources include official government health agencies and international organizations, such as the U.S. Centers for Disease Control and the World Health Organization (WHO). I use the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, which collects and standardizes data from these primary governmental and international organizations. The data collected includes the number of reported new cases and deaths from 2020 through 2023.

Reporting Methods

To fully understand the context and nuances of the reported COVID-19 data, it is crucial to examine the reporting methods employed by each nation. Different countries have adopted different approaches to data collection and reporting. These disparities and complexities in the data must be accounted for and standardized across nations and to limit statistical bias in this analysis. Reporting considerations for each nation are as follows:

South Korea: To increase its ability for extensive testing and contact tracking, South Korea passed laws comparatively early to other countries. The nation took a transparent and proactive stance when it came to reporting, providing comprehensive details on test findings, contact tracking, and quarantine procedures. South Korea, a tiny but affluent country with a strong healthcare system, was able to supply pretty accurate and standardized data regarding COVID-19 cases.

United States: The United States used a decentralized approach to data reporting, giving each state some degree of independence in terms of how data is gathered and reported. Differences in reporting requirements and data quality were brought about by this decentralized method. While the Centers for Disease Control and Prevention (CDC) acted as a central organizing agency, data gathering was hampered by state-level variations.

Europe: Without standardized reporting procedures, the European Union (EU), which is made up of several member states, faced the difficulty of harmonizing data from disparate healthcare systems. Although the EU established rules for data gathering, each member state was required to modify these recommendations to fit their unique situation. The inconsistent reporting criteria created difficulties for uniformity.

China: Especially during the early phases of the pandemic, China implemented strict policies for data collecting and reporting. During the early stages of the outbreak, the Chinese authorities came under fire for possible underreporting and a lack of transparency. However, reporting techniques changed throughout time, which can have an impact on data consistency.

Japan: Japan adopted a novel strategy for COVID-19 data reporting, giving privacy issues a priority while preserving efficient public health protocols. To create digital tools, the nation brought together extensive testing, effective contact tracing, and partnership with the commercial sector. Local health authorities were crucial in promoting community-based monitoring; nevertheless, there were difficulties in standardizing data reporting throughout various regions. Japan managed the delicate balancing act between the demands of public health and individual rights by adapting international norms to its cultural setting. Japan was better prepared to handle future health crises thanks to the ongoing improvement of reporting techniques, which reflected a process of learning from both successes and setbacks.

Data Visualization and Statistical Analysis
Data was analyzed with Python, including packages such as NumPy and Pandas. Matplotlib was used to create clear images that enabled data visualization and analysis.

Results
For all countries, the total number of cases grew at differing rates; however, one could see a common rise in cases just at the start of 2022, in which the rate of change in cases was the largest for most of the countries, excluding China and Japan (Figure 1).
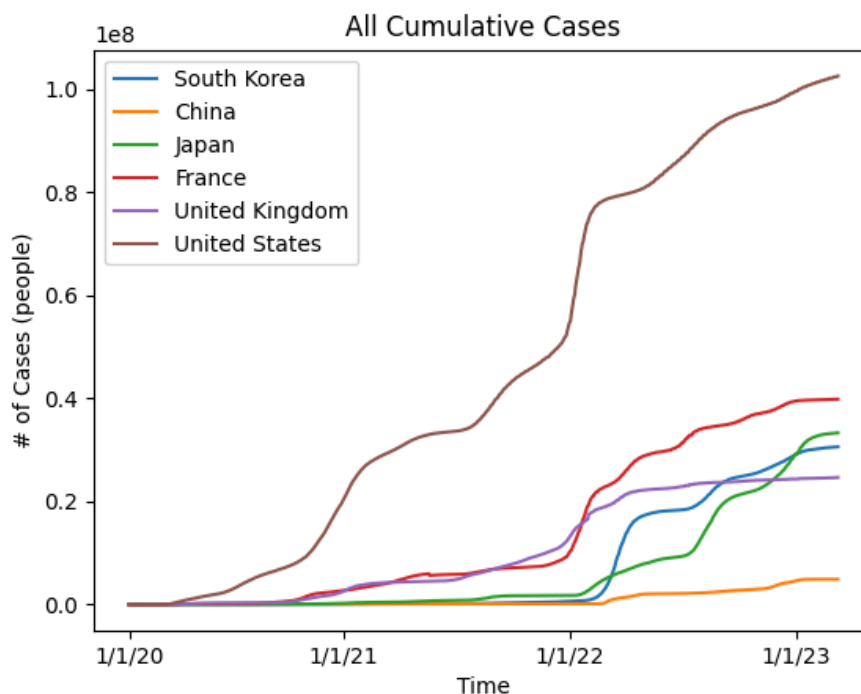
<Figure 1> A Timeline for Cumulative COVID-19 Cases for All Countries Studied

This could be explained by the introduction of the omicron (the BA.1 variant) and delta variants, which "was causing daily case numbers in the U.S. [as well as other countries] to skyrocket to over a million." The spread of the omicron (BA.1) variant led to the spawning of sub variants such as the EG.5 (nicknamed "Eris") and BA.2.86 (nicknamed "Pirola") (Katella). For Japan, the rate of change for the number of cases was the highest in the mid-2022s, as BA.5, a new omicron subvariant that had mutated from the original BA.1 variant, was introduced in Japan in the mid 2022s (Otake).

Furthermore, we saw a massive under-reporting of the true number of cases. For example, the population in Japan is about 123.3 million, and the cumulative number of cases at the end of 2023 is slightly less than 40 million people, indicating around 30% of the population was infected (Figure 1). However, although China has a population of over 1.4 billion people (O'Neill), China's reporting shows that the cumulative number of cases is less than 10 million people, which is less than 0.7% of the total population.
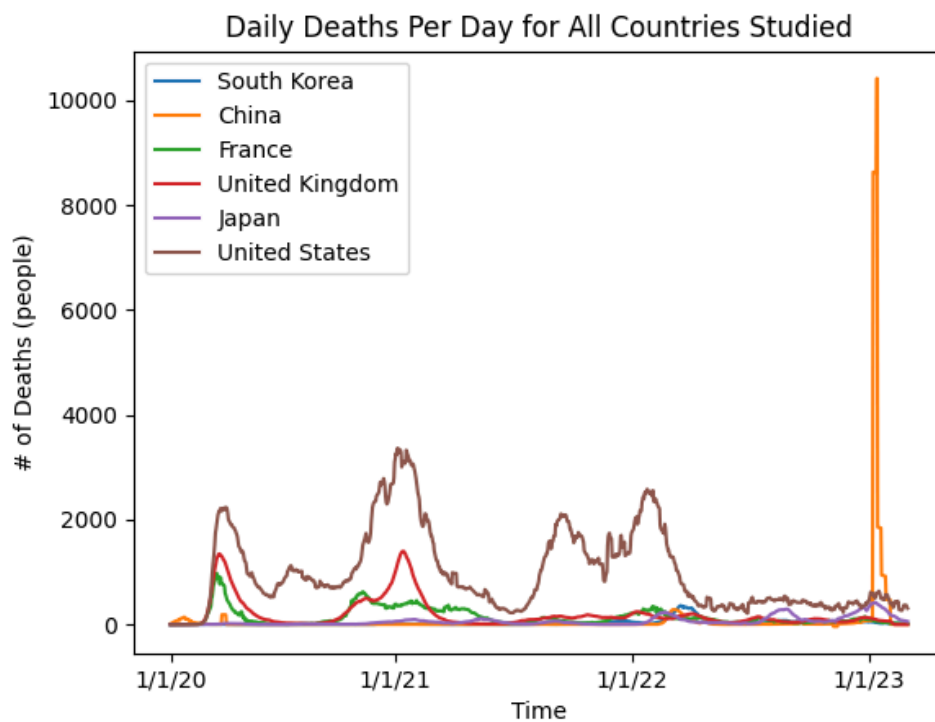

<Figure 2> Daily Deaths for All Countries Studied

Next, I studied the number of deaths reported daily for each country, smoothed over a seven day moving average (Figure 2). For all countries, with the exception of China, there were two primary peaks in the number of reported deaths; the first peak occurred approximately February to March of 2020, and the second occurred at the beginning of 2021. The first peak can be explained by the introduction of the virus to new populations without immunization or acquired immunity and without reliable antivirals, causing a surge in deaths. The peak in 2022 was likely caused by increased mortality from infection during the winter, as has been observed for other

respiratory illnesses such as the Spanish Flu. However, while all other countries' daily deaths did not exceed 4000 per day, China had over 10,000 deaths per day in 2023 and skyrocketed down after the reporting, suggesting a clear anomaly in their data.

Finally, I calculated and plotted the cumulative mortality rate, which is the total number of deaths over the total number of cases (Figure 3).
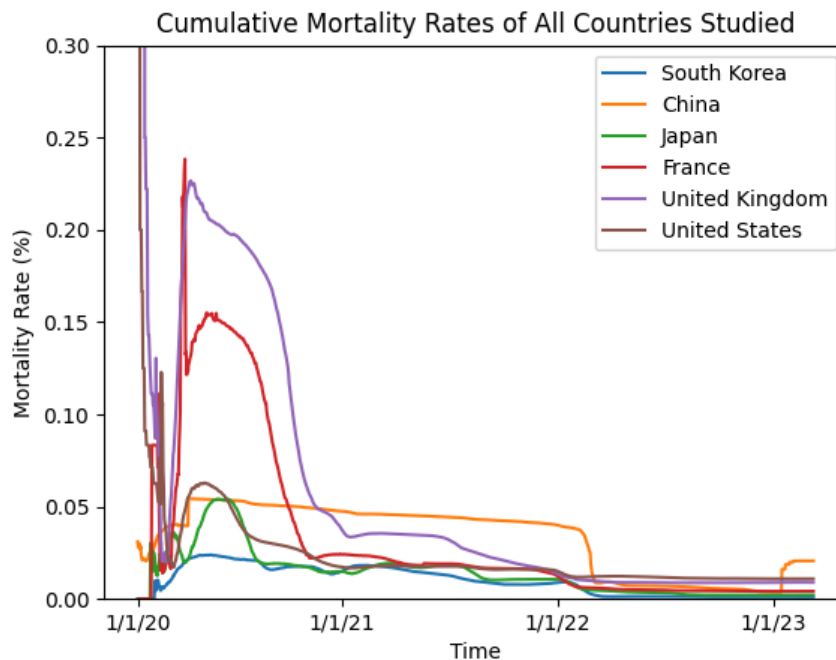
<Figure 3> Cumulative Mortality Rates of All Countries Studied

The cumulative mortality rates fluctuated significantly at first but, as time passed, the mortality rates eventually stabilized. The initially high mortality rates were likely caused by a combination of higher early virulence in SARS-CoV-2 and smaller testing capacity, biasing positive cases to people who were severely ill and hospitalized. In late 2020, vaccines became available for the elderly and other vulnerable populations, which significantly decreased the mortality rates. Also, although there were new variants that became more contagious, particularly the Omicron (BA.1) variant, mortality rates continued to decline, suggesting less virulence than previous variants (Katella).

Discussion

This study has identified that COVID-19 data reported by the six countries under investigation (China, Japan, South Korea, France, United States, and United Kingdom) have shown both consistencies as well as inconsistencies. To define, consistencies are common or similar patterns witnessed in several nations that show how they reported about the pandemic and how they were affected by it. However, inconsistencies may depict variations in statistics amongst these countries. China had notable abnormalities (i.e: unexpectedly high or low case counts for

some days, discrepancies in death numbers, or irregularity of general movement of the pandemic), suggesting that its reports on COVID-19 might be flawed and censored while other countries' numbers portrayed uniform changes across different categories of patients with downward trend of deaths per case. Conversely, the suggestion "flaws and censorship" in China's reporting points to the fact that such peculiarities within COVID-19 data might not be due to accidental mistakes or collection errors only. Rather, this means that there could be deliberate actions of withholding or changing information leading to observed inconsistencies.

China's reporting irregularities cast doubt on the accuracy of the data, especially when it comes to the overall number of cases and daily deaths. The rapid decline in mortality rates in 2022, with the lack of instances reported in a populous nation, raises the possibility of discrepancies and censorship.

China's data discrepancy aligns with past research and analyses that have questioned the legitimacy, openness, and methodology of China's COVID-19 reporting (The Diplomat, Nature, The Guardian). A number of research investigations and reports have cast doubt on the dependability and accuracy of COVID-19 reporting, especially in China. For example, in an article on China's missing COVID-19 data, The Diplomat (2023) emphasized the challenges in acquiring reliable information because of official censorship. The authenticity of China's COVID-19 deaths was also questioned by The Guardian (2022), who pointed out discrepancies and potential underreporting.

Research published in PLOS (2023) revealed widespread underreporting of COVID-19 cases and deaths, with different degrees of accuracy across countries. The CFR (Council on Foreign Relations) revealed the extent of COVID-19 case and death underreporting worldwide, emphasizing the importance of open reporting (CFR). A comparison of key COVID-19 data sources (OpenNews) revealed variances in reporting procedures, emphasizing the need of recognizing these changes when interpreting global COVID-19 data.

Implication on Public Health
Inaccurate COVID-19 reporting has consequences for resource allocation, policymaking, and health outcomes. It is crucial to have data during a pandemic to make decisions and develop effective public health strategies. For instance, if case counts are manipulated, leading to a false sense of security and insufficient resource allocation, it could hinder the timely deployment of critical medical supplies, personnel, and facilities.

During the stages of the pandemic there was a real world example where inaccurate data had effects. Some areas initially underestimated the severity of the virus resulting in delayed reactions and a failure to implement precautions. This led to virus spread and increased strain on healthcare systems. Conversely, other areas that reported promptly and accurately were able

to take measures such as lockdowns and targeted testing which mitigated the impact of the virus. Furthermore, accurate information is crucial for public health initiatives. Vaccination programs rely on data regarding infection rates, high risk populations and intervention effectiveness. Inaccurate reporting can hinder the success of vaccination efforts by offering instructions on how to distribute vaccines and not adequately protecting individuals at a risk.

Erroneous statistics can have direct repercussions as well as hinder efforts to determine the exact cost of illness. It is more difficult to evaluate the long-term health effects, allot funds for post-acute treatment, and make plans for possible future outbreaks if the provided data understates the true number of patients or distorts the severity of the illness. In order to identify the areas of greatest need, efficiently manage public health crises, and customize interventions to individual needs, honest and trustworthy reporting is essential.

Conclusion

This study provides a thorough examination of the correctness and consistency of COVID-19 data supplied from six different countries. While the majority of other countries' data showed consistent trends, China's statistics showed some noticeable abnormalities that sparked concerns about potential censorship and inaccuracy. The study emphasizes the necessity of transparent reporting as well as the difficulties in locating reliable information globally.

To address these issues, more international cooperation, transparent communication, standardized reporting practices, and independent verification are required. The importance of precise statistics for effective policy making, resource allocation, and, ultimately, lessening the effects of global health crises is highlighted by obvious public health implications. I believe that this study provides valuable information that will guide future efforts to improve the standard of COVID-19 data reporting.

References

[1] Costigan, J. M., & Ye, J. (2023). China's Missing COVID-19 Data. Retrieved from https://thediplomat.com/2023/01/chinas-missing-covid-19-data/.

[2] CSSEGISandData. (n.d.). *CSSEGISANDDATA/COVID-19: Novel coronavirus (COVID-19) cases, provided by JHU CSSE*. GitHub. https://github.com/CSSEGISandData/COVID-19.

[3] Fischer-Hwang, I., & Mayo, J. (2020, May 5). *A comparison of four major COVID-19 data sources*. Source. https://source.opennews.org/articles/comparison-four-major-covid-19-data-sources/.

[4] Guardian News and Media. (2022, December 22). *How accurate are China's Covid death numbers?* The Guardian. https://www.theguardian.com/world/2022/dec/23/how-accurate-are-chinas-covid-death-numbers.

[5] Katella, K. (2023, September 1). *Omicron, Delta, Alpha, and more: What to know about the coronavirus variants*. Yale Medicine. https://www.yalemedicine.org/news/covid-19-variants-of-concern-omicron.

[6] Klobucista, C. (2021, May 10). *By how much are countries underreporting COVID-19 cases and deaths?*. Council on Foreign Relations. https://www.cfr.org/in-brief/how-much-are-countries-underreporting-covid-19-cases-and-deaths.

[7] Lee, C. (2022, October 6). *Deaths from covid-19 spiked over the summer, especially among people 65 and older, before dipping again in September*. KFF. https://www.kff.org/coronavirus-covid-19/press-release/deaths-from-covid-19-spiked-over-the-summer-especially-among-people-65-and-older-before-dipping-again-in-september/.

[8] Otake, T. (2022, July 7). *New surge in COVID-19 cases across Japan prompts alarm*. The Japan Times. https://www.japantimes.co.jp/news/2022/07/07/national/surge-covid-japan-prompts-alarm/.

[9] O'Neill, A. (2023, November 7). *Countries with the largest population 2023*. Statista. https://www.statista.com/statistics/262879/countries-with-the-largest-population/.

[10] Tsou, HH., Kuo, SC., Lin, YH. *et al.* A comprehensive evaluation of COVID-19 policies and outcomes in 50 countries and territories. *Sci Rep* 12, 8802 (2022). https://doi.org/10.1038/s41598-022-12853-7.

[11] Rotulo, A., Kondilis, E., Thwe, T., Gautam, S., Torcu, Ö., Vera-Montoya, M., Marjan, S., Gazi, Md. I., Putri, A. S., Hasan, R. B., Mone, F. H., Rodríguez-Castillo, K., Tabassum, A., Parcharidi, Z., Sharma, B., Islam, F., Amoo, B., Lemke, L., & Gallo, V. (2023, April 21). *Mind the gap: Data availability, accessibility, transparency, and credibility during the COVID-19 pandemic, an international comparative appraisal*. PLOS Global Public Health. https://journals.plos.org/globalpublichealth/article?id=10.1371%2Fjournal.pgph.0001148.