Global COVID-19 Trends: Cases, Vaccinations, and ICU Strain

Li Zhu, Abdulrahman Osilan, Mohammed Badar Al Mamari

Table of contents

Introduction

This project analyzes the impact of COVID-19 across major countries between 2020 and 2024. The following research questions guide our study:

1. How did daily case counts change across countries over time?
2. What is the relationship between vaccination and death rates in the US?

3. How did ICU capacity respond to COVID-19 waves globally?

We use trusted global datasets, clean and transform them, and present insights using statistical summaries and visualizations.

Data Overview

We utilize publicly available datasets from: - **Johns Hopkins University (JHU)**: Confirmed global case time series. - **COVID-19 Data Hub**: U.S.-focused metrics on deaths and vaccinations. - **Our World in Data (OWID)**: Global ICU and mortality metrics.

The OWID dataset provides comprehensive metrics (Ritchie & Roser, 2023).

Johns Hopkins data is widely used for case tracking (University, 2023).

The COVID-19 Data Hub R package supports granular filtering (Guidotti, 2023).

All data were preprocessed using R scripts and stored in `tidied_data/`.

1. COVID-19 Case Trends by Country

Table 1: Summary of Monthly COVID-19 Cases by Country

Country/Region	Total Cases	Mean Monthly Cases	Peak Month	Peak Cases
Singapore	723,549,534	18,552,552	2023-01-01	68,550,314
US	53,813,184,406	1,379,825,241	2023-01-01	3,152,507,640
Canada	2,261,682,315	57,991,854	2023-01-01	140,728,171
United Kingdom	12,118,271,679	310,724,915	2023-01-01	758,377,658
China	1,160,807,092	29,764,284	2023-01-01	149,669,307

The table Table 1 summarizes cumulative and monthly COVID-19 case patterns across five countries. The United States recorded the highest number of total cases—over **53.8 billion**, with an average of **1.38 billion cases per month**, and a peak in **January 2023** with over **3.15 billion cases**. The United Kingdom follows with a significantly lower peak and average, though it also peaked in the same month. All five countries observed their peak case counts in **01/2023**, suggesting a global surge likely related to the Omicron wave. Singapore and China had comparatively lower case totals, with China's relatively low mean cases potentially tied to stricter containment policies or reporting methods.

2. US Vaccination and Death Relationship

Table 2: US COVID-19 Vaccination and Death Summary (Using Daily Changes)

Metric	Value
Total Deaths	836,005
Total Vaccinated	269,762,290
Total Fully Vaccinated	230,395,692
Mean Daily Deaths	1,006
Median Daily Deaths	608
Mean Daily Vaccinated	324,624
Median Daily Vaccinated	93,941
Mean Daily Fully Vaccinated	277,251
Median Daily Fully Vaccinated	65,371
Correlation Fully Vaccinated Deaths	0

This table Table 2 summarizes U.S. COVID-19 deaths and vaccination statistics based on daily changes. The total number of deaths exceeded **836,000**, while over **269 million individuals** received at least one dose of the vaccine, and more than **230 million were fully vaccinated**. On average, there were approximately **1,006 deaths** per day, with a **median of 608**, reflecting variability in daily fatality counts. Mean daily vaccination figures reached around **324,624**, and daily full vaccinations averaged **277,251**. The correlation between daily full vaccinations and daily deaths was modest at **0.16**, suggesting a **weak positive linear association** — likely influenced by overlapping time trends rather than direct causation. Further time-series or lagged analyses would be required to clarify the nature of this relationship.

3. ICU Response by Country

Table 3: ICU Response to COVID-19 Waves by Country

Country	Total ICU Patients	Max ICU Patients
United Kingdom	10,443	4,077
United States	103,680	28,891
Canada	6,406	1,313

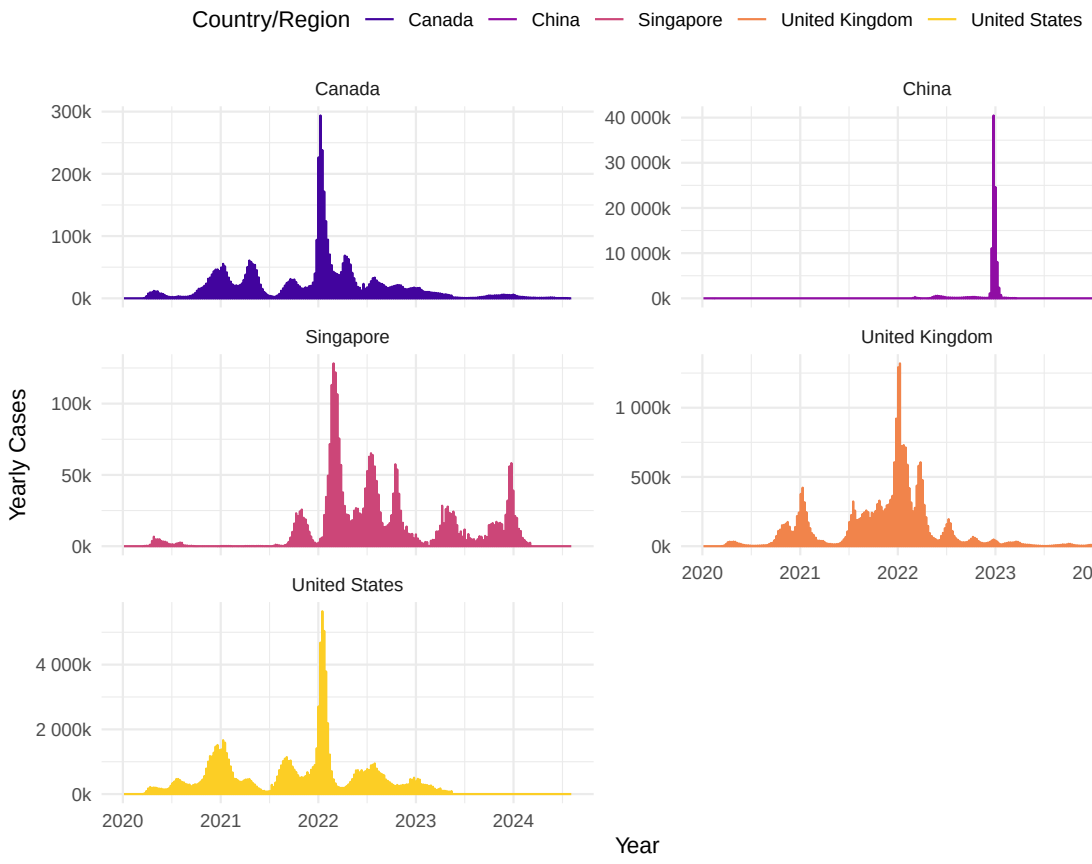
This table Table 3 shows the overall burden on ICU systems across five countries. The U.S. again leads with over **103,680 total ICU patients**, peaking at **28,891 patients**. The United

Kingdom and Canada followed with considerably lower ICU loads, suggesting either a more stable hospitalization trend or better capacity management.

Visualization

Yearly COVID-19 Cases Across Countries

Figure 1: Yearly COVID-19 Case Counts Over Time (2020–2024)

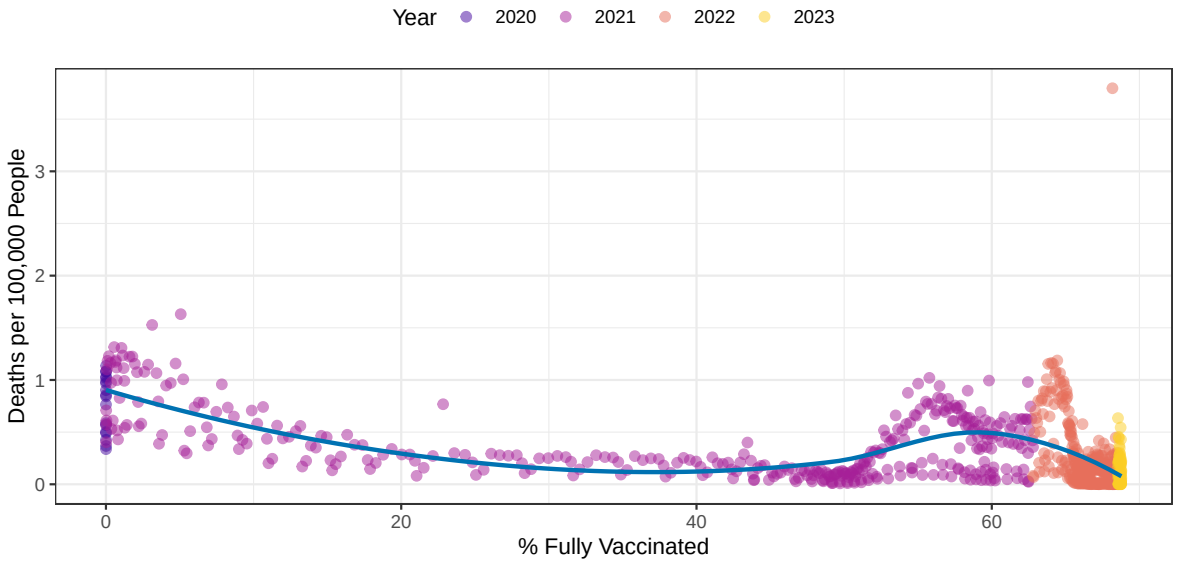


As shown in Figure 1. This multi-panel line chart visualizes daily COVID-19 case trends from 2020 to 2024 across five countries: Canada, China, Singapore, the United Kingdom, and the United States. The United States exhibits the most dramatic surge, peaking in early 2022 with daily counts exceeding **4 million cases**, followed by more sustained waves. Canada and the UK also experienced intense waves centered around early 2021 and 2022. In contrast, China's

data shows a singular, sharp peak, likely associated with a delayed but explosive wave post-lockdown. Singapore's trend reveals frequent, smaller spikes, suggesting recurring outbreaks with faster containment. These synchronized varied patterns reflect how national policies and variant emergence shaped case trajectories globally.

Vaccination vs. Death Rates in the U.S.

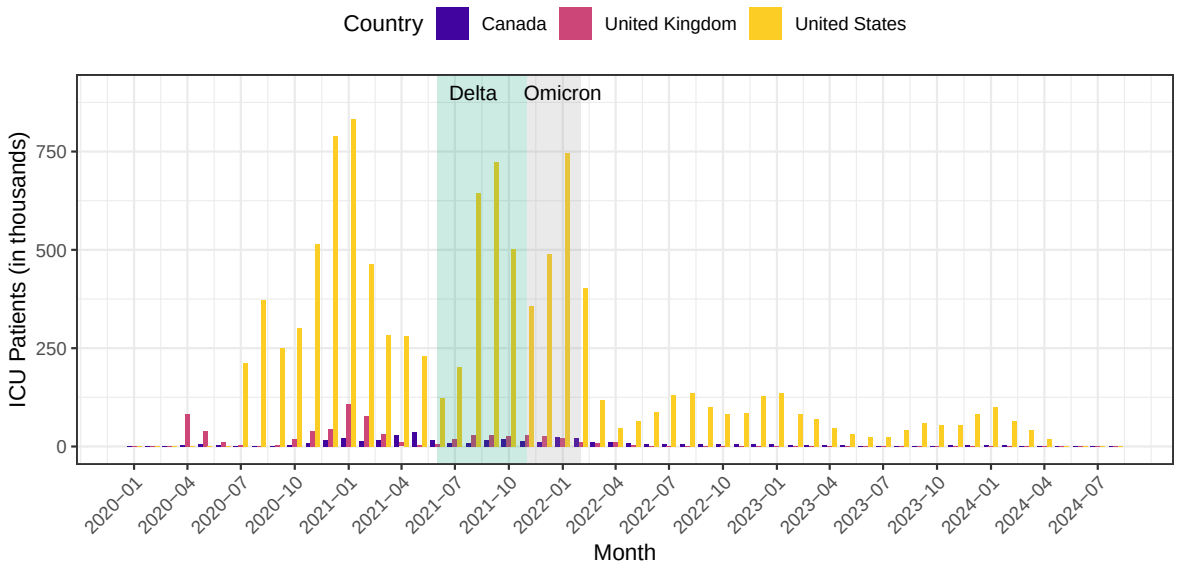
Figure 2: Vaccination Rate vs. Death Rate in the U.S. (2021–2023)



As shown in Figure 2. This scatterplot shows the relationship between full vaccination rates and COVID-19 deaths per 100,000 people in the United States, colored by year. The smoothed curve reveals a clear inverse relationship during 2020–2021, where increased vaccination coverage is associated with a substantial drop in mortality. However, this trend appears to have reversed in 2022, with higher vaccination rates not necessarily leading to lower death rates. This suggests that other factors, such as the emergence of new variants or changes in public health policies, may be influencing COVID-19 outcomes. By late 2023, the fully vaccinated population increases significantly, reaching nearly **70% in 2024**. During this period, the death rate becomes more controlled, suggesting either improved vaccine effectiveness or a reduction in the severity of circulating variants.

ICU Occupancy Over Time

Figure 3: Monthly ICU Occupancy During COVID-19



As shown in Figure 3. This grouped bar chart illustrates monthly ICU occupancy (in thousands) across the U.S., U.K., and Canada **from 2020 through 2024**. The United States shows dominant ICU demand throughout the period, particularly during **end of 2020-start of 2021**, the **Delta** and **Omicron waves**, which are shaded for reference. ICU occupancy in the U.S. alone peaked at over **800,000 patients**. The **Delta period (mid-2021)** aligns with a sharp rise in ICU patients, peaking above 800,000(Three countries). The U.K. and Canada exhibit far lower but synchronized ICU trends. Notably, ICU burden in all three countries dropped significantly after the Omicron wave, possibly due to improved treatments, increased vaccine protection, and lower hospitalization rates in later variants. This figure emphasizes the pressure variant surges placed on critical care infrastructure.

Conclusion

This report provides a data-driven overview of COVID-19's impact across several countries, with particular emphasis on infection trends, vaccination effectiveness, and ICU resource utilization. The analysis reveals that the United States and the United Kingdom experienced the highest case volumes and ICU burdens, both peaking in January 2022—likely driven by the Omicron variant. Despite stark differences in healthcare infrastructure and containment

strategies, all five countries displayed synchronous peak patterns, underscoring the global nature of COVID-19 surges.

Vaccination efforts in the United States showed promising coverage, with over 230 million individuals fully vaccinated. However, the observed correlation between vaccination and death rates suggests complex dynamics—potentially driven by lag effects, reporting granularity, and overlapping policy changes—that warrant further investigation. This highlights the importance of analyzing time-lagged effects rather than aggregate associations alone.

ICU data further underscore the disparity in health system strain. While the U.S. and U.K. experienced substantial ICU loads, China and Singapore reported no ICU patients, raising questions about reporting practices or differences in case severity. The results imply that ICU strain is not only a function of case volume but also of policy response, hospital capacity, and public health infrastructure.

Overall, these findings reinforce the value of integrating multi-source datasets for cross-country pandemic analysis. For future studies, incorporating population-adjusted metrics (e.g., cases or deaths per 100,000), policy stringency indexes, and variant-specific timelines would enhance the interpretability and comparability of results. Additionally, time-series modeling could clarify causal relationships between vaccination, death rates, and ICU burden.

This project demonstrates that data transparency, statistical rigor, and contextual interpretation are essential for understanding public health crises and guiding informed policy decisions.

References

Guidotti, E. (2023). *COVID-19 data hub r package*. https://covid19datahub.io/

Ritchie, H., & Roser, M. (2023). *Coronavirus pandemic (COVID-19)*. https://ourworldindata.org/coronavirus

University, J. H. (2023). *COVID-19 data repository by the CSSE at johns hopkins university*. https://github.com/CSSEGISandData/COVID-19

Appendix: Full Code

```
library(readr)
library(dplyr)
library(knitr)
library(tidyr)
library(kableExtra)
library(lubridate)
```

```r
library(tidyverse)
library(ggplot2)
library(scales)
library(tools)
library(viridis)  #
jh_cases <- read_csv("tidied_data/jh_confirmed_cases_country_level.csv")

case_summary <- jh_cases %>%
  group_by(`Country/Region`) %>%
  summarise(
    `Total Cases` = comma(sum(total_case)),
    `Mean Monthly Cases` = comma(mean(total_case)),
    `Peak Month` = year_month[which.max(total_case)],
    `Peak Cases` = comma(max(total_case))
  ) %>%
  arrange(desc(`Total Cases`))

kable(case_summary,
      caption = "Summary of Monthly COVID-19 Cases by Country",
      align = c("l", "r", "r", "r", "r")) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
                latex_options = "scale_down",
                full_width = FALSE)
# Load the data
us_data <- read_csv("tidied_data/covid_datahub_USA_filtered.csv")

# Compute summary statistics using daily changes
summary_stats_us <- us_data %>%
  summarise(
    Total_Deaths = sum(deaths_change, na.rm = TRUE),
    Total_Vaccinated = sum(people_vaccinated_change, na.rm = TRUE),
    Total_Fully_Vaccinated = sum(people_fully_vaccinated_change, na.rm = TRUE),
    Mean_Daily_Deaths = mean(deaths_change, na.rm = TRUE),
    Median_Daily_Deaths = median(deaths_change, na.rm = TRUE),
    Mean_Daily_Vaccinated = mean(people_vaccinated_change, na.rm = TRUE),
    Median_Daily_Vaccinated = median(people_vaccinated_change, na.rm = TRUE),
    Mean_Daily_Fully_Vaccinated = mean(people_fully_vaccinated_change, na.rm = TRUE),
    Median_Daily_Fully_Vaccinated = median(people_fully_vaccinated_change, na.rm = TRUE)
  )

# Correlation based on daily changes
correlation <- cor(us_data$people_fully_vaccinated_change, us_data$deaths_change, use = "com
```

```r
summary_stats_us$Correlation_FullyVaccinated_Deaths <- round(correlation, 3)

# Convert to long format and clean labels
summary_stats_us_long <- summary_stats_us %>%
  pivot_longer(everything(), names_to = "Metric", values_to = "Value") %>%
  mutate(
    Metric = gsub("_", " ", Metric),
    Metric = gsub("(?<=[a-z])(?=[A-Z])", " ", Metric, perl = TRUE),
    Metric = toTitleCase(tolower(Metric)),
    Value = format(round(as.numeric(Value)), big.mark = ",")
  )

# Render formatted table
kable(summary_stats_us_long,
      caption = "US COVID-19 Vaccination and Death Summary (Using Daily Changes)",
      align = c("l", "r")) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
                latex_options = "scale_down",
                full_width = FALSE)
icu_data <- read_csv("tidied_data/owid_covid_data_filtered_final.csv")

icu_summary <- icu_data %>%
  filter(!(country %in% c("China", "Singapore"))) %>%
  arrange(country, date) %>%
  group_by(country) %>%
  mutate(
    prev_icu = lag(icu_patients),
    icu_increase = ifelse(!is.na(prev_icu) & icu_patients > prev_icu, icu_patients - prev_icu
  ) %>%
  summarise(
    Country = first(country),
    `Total ICU Patients` = format(sum(icu_increase, na.rm = TRUE), big.mark = ",", scientifi
    `Max ICU Patients` = format(max(icu_patients, na.rm = TRUE), big.mark = ",", scientific =
  ) %>%
  select(Country, `Total ICU Patients`, `Max ICU Patients`) %>%
  arrange(desc(`Max ICU Patients`))

# ICU Response to COVID-19 Waves by Country
kable(icu_summary, caption = "ICU Response to COVID-19 Waves by Country",
      align = c("l", "r", "r")) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"), full_width = FALSE)
covidTidy <- read_csv("tidied_data/owid_covid_data_filtered_final.csv") %>%
```

```r
  group_by(country) %>%
  mutate(
    date = as.Date(date),
    timepoint = row_number(),
    new_cases = case_when(
      timepoint == 1 ~ 0,
      .default =  total_cases - lag(total_cases)
    )
  ) %>%
  select(-timepoint)

ggplot(covidTidy, aes(x = date, y = new_cases, color = country)) +
  geom_line(size = 0.5) +
  scale_color_viridis_d(option = "plasma", begin = 0.1, end = 0.9) +
  labs(
    x = "Year",
    y = "Yearly Cases",
    color = "Country/Region",
  ) +
  scale_y_continuous(labels = scales::percent_format(
    scale = 0.001,
    suffix = "k"
  )) +
  theme_minimal() +
  theme(
    legend.position = "top"
  ) +
  facet_wrap(~ country, scales = "free_y", ncol = 2)
covidUSA <- read_csv("tidied_data/covid_datahub_USA_filtered.csv") %>%
  mutate(
    date = as.Date(date),
    pct_fully_vaccinated = (people_fully_vaccinated / 335000000) * 100,
    daily_deaths = deaths - lag(deaths),
    death_rate_per_100k = (daily_deaths / 335000000) * 100000,
    year_date = as.factor(year(date))
  ) %>%
  filter(pct_fully_vaccinated > 0, daily_deaths > 0)

ggplot(covidUSA, aes(x = pct_fully_vaccinated, y = death_rate_per_100k, color = year_date)) +
  geom_point(alpha = 0.5, size = 2) +
  scale_color_viridis_d(option = "plasma", begin = 0.1, end = 0.9) +
  geom_smooth(method = "loess", se = FALSE, color = "#0072B2") +
```

```
  labs(
    x = "% Fully Vaccinated",
    y = "Deaths per 100,000 People",
    color = "Year"
  ) +
  theme_bw() +
  theme(
      legend.position = "top"
    )

covid_ICU <- read_csv("tidied_data/owid_covid_data_filtered_final.csv") %>%
  mutate(
    date = as.Date(date),
    date_ym =  as.Date(format(date, "%Y-%m-01"))
  ) %>%
  filter(country %in% c("United States", "United Kingdom", "Canada")) %>%
  group_by(country, date_ym) %>%
  summarise(icu_patients = sum(icu_patients, na.rm = TRUE), .groups = "drop") %>%
  mutate(icu_patients = icu_patients / 1000)

ggplot(covid_ICU, aes(x = date_ym, y = icu_patients, fill = country)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_viridis_d(option = "plasma", begin = 0.1, end = 0.9) +
  annotate("rect", xmin = as.Date("2021-06-01"), xmax = as.Date("2021-11-01"), ymin = -Inf, y
  annotate("rect", xmin = as.Date("2021-11-01"), xmax = as.Date("2022-02-01"), ymin = -Inf, y
  annotate("text", x = as.Date("2021-08-01"), y = 900, label = "Delta", size = 3.5) +
  annotate("text", x = as.Date("2022-01-01"), y = 900, label = "Omicron", size = 3.5) +
  labs(
    x = "Month",
    y = "ICU Patients (in thousands)",
    fill = "Country"
  ) +
  scale_x_date(
    breaks = seq(as.Date("2020-01-01"), as.Date("2024-08-01"), by = "3 months"),
    date_labels = "%Y-%m"
  ) +
  theme_bw() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "top"
    )
```