**Investigating COVID-19 Virus Trends**

Thesis Paper for I-492 Project

by

Debbie Cung

**Advisor**

Dr. Sridhar Ramachandran

Table of Contents

[I. Abstract 3](#_Toc212057073)

[II. Introduction 3](#_Toc212057074)

[III. Literature Review 4](#_Toc212057075)

[3.1 COVID-19 Testing Strategies and Effectiveness 4](#_Toc212057076)

[3.1.1 Timing and Scale of Testing 4](#_Toc212057077)

[3.1.2 Accuracy and Reporting Limitations 4](#_Toc212057078)

[3.1.3 Innovations in Testing and Detection 4](#_Toc212057079)

[3.1.4 Summary 5](#_Toc212057080)

[3.2 Pandemic Outcomes and Public Health Measures 5](#_Toc212057081)

[3.2.1 Hospitalization and Recovery Patterns 5](#_Toc212057082)

[3.2.2 Vaccines and Treatments 5](#_Toc212057083)

[3.2.3 Policy Interventions and Interaction with Testing 6](#_Toc212057084)

[3.2.4 Summary 6](#_Toc212057085)

[3.3 Critical Evaluation 6](#_Toc212057086)

[IV. Methods 6](#_Toc212057087)

[4.1 Dataset 7](#_Toc212057088)

[4.2 Tools and Software 7](#_Toc212057089)

[4.3 Procedures 7](#_Toc212057090)

[4.4 Reproducibility 8](#_Toc212057091)

[V. Results 9](#_Toc212057092)

[5.1 Overview of Findings 9](#_Toc212057093)

[5.2 Fixed-Effects Regression Analysis 9](#_Toc212057094)

[5.3 Robustness Checks 9](#_Toc212057095)

[5.4 Visual Summaries 10](#_Toc212057096)

[5.5 Interpretation of Results 14](#_Toc212057097)

[VI. Discussion 14](#_Toc212057098)

[VII. Conclusion 15](#_Toc212057099)

[VIII. Future Directions 16](#_Toc212057100)

[List of References 16](#_Toc212057101)

[Link to the References 18](#_Toc212057102)

# I. Abstract

This study examines whether higher COVID-19 testing rates were associated with lower numbers of active cases and fewer negative outcomes due to faster detection and more effective countermeasures, such as isolating infected individuals quickly. Using the *COVID-19 Worldwide Testing Dataset* from Kaggle, the research analyzes daily global records of testing, case counts, and recoveries from 2020 through 2022. The dataset was cleaned to remove outliers, handle missing values, and smooth out fluctuations using seven-day rolling averages. Fixed-effects regression models were then applied to estimate the relationship between lower and slower testing rates and the growth in the number of positive cases across countries and time.

When the results were reviewed, a clear pattern emerged: areas that conducted more testing earlier tended to see their case numbers rise more slowly later on. It seems that steady, early testing helped spot infections before they spread widely. To ensure the trend wasn’t random, the study ran several reliability checks, such as bootstrap samples, placebo runs, and jackknife tests — and the pattern still held up. Visual summaries, such as distributed lag charts and scatter plots, told the same story: stronger testing efforts were often followed by flatter growth in infections. This helped reinforce the research hypothesis: broad testing early on made a real difference in slowing the virus, and that lesson could help shape how future outbreaks are handled.

# II. Introduction

The COVID-19 pandemic became one of the most difficult health crises in recent history, touching nearly every country and overwhelming many healthcare systems. Governments tried different ways to slow the spread, such as lockdowns, distancing rules, and vaccines — but widespread testing turned out to be one of the most effective tools. Testing made it possible to find infections early, isolate sick individuals, and guide public-health decisions before outbreaks grew worse. Still, not every country had the same testing capacity or data reliability. Some moved quickly with large-scale testing programs, while others struggled to get them running, leading to much faster and less controlled transmission.

This study looks at how differences in testing rates may have shaped the course of COVID-19 outbreaks in different countries. Although testing is widely accepted as a key public-health measure, it isn’t always clear how much more testing actually slows the rise in active cases when viewed across global data. Gaining a clearer sense of that connection can help explain how testing strategies influence wider outcomes, including hospitalizations and deaths.

The hypothesis guiding this study is that when testing rates rise, the number of active COVID-19 cases should fall over time, primarily because infections are caught and isolated earlier. To examine this, the research uses a fixed-effects panel regression along with several reliability checks, drawing on daily worldwide data covering tests, confirmed cases, recoveries, and deaths.

This research matters because it offers real-world insights for improving public health readiness. By measuring how testing relates to infection trends, the study gives policymakers concrete evidence that investing in testing systems early can slow the spread of disease. The findings add to global discussions about pandemic preparedness and highlight lessons that apply not only to COVID-19 but also to future outbreaks of infectious diseases.

# III. Literature Review

## 3.1 COVID-19 Testing Strategies and Effectiveness

Testing proved to be one of the most effective ways to slow the spread and lessen the impact of COVID-19. It made it possible to detect infections early, trace contacts, and isolate cases quickly, which helped public health officials respond faster and more accurately. Most researchers agree that testing was a key factor in lowering infection rates, though they note that timing, scale, and testing methods varied widely from place to place.

### 3.1.1 Timing and Scale of Testing

Chinazzi et al. (2020) studied how early testing and travel limits shaped the course of COVID-19 in China, finding that fast, widespread testing combined with movement controls helped slow the spread of the virus. Walker et al. (2020) reached a similar conclusion in their analysis of low- and middle-income countries, showing that delays in testing weakened public health responses and led to higher death rates. Writing from the UK context, Peto (2020) argued for ongoing mass testing, noting that even less-than-perfect tests could do more to contain outbreaks than occasional lockdowns.

Although these studies differed in location and method, they shared a common finding: testing was most effective when started early and maintained at a broad scale. Chinazzi and Walker underscored the importance of timing and infrastructure, while Peto emphasized consistency and reach. Taken together, their work points to the crucial role of testing capacity and timeliness in shaping pandemic outcomes.

### 3.1.2 Accuracy and Reporting Limitations

Testing also revealed serious issues with data quality and consistency in reporting. Li et al. (2020) examined the early stages of the outbreak in Wuhan and found that limited testing capacity caused widespread underreporting of cases. Hasell et al. (2020) developed a global testing database — later incorporated into *Our World in Data* — and showed that differences in definitions and data collection practices made accurate cross-country comparisons difficult. These inconsistencies often hid the real extent of the pandemic and made global analysis more complex.

Both Li and Hasell’s work highlight that the value of testing went beyond simply identifying infections—it also relied on how accurate and transparent the data were. Their findings stress the need for standardized, high-quality datasets, such as the Kaggle COVID-19 Worldwide Testing Data, which allow for reliable comparisons across countries and over time.

### 3.1.3 Innovations in Testing and Detection

As traditional swab tests became harder to obtain, researchers began turning to new diagnostic approaches. The SIIM-FISABIO-RSNA Kaggle competition (2021) challenged data scientists to apply deep learning to chest X-rays, demonstrating that computer vision could accurately identify COVID-19 infections. Likewise, Li et al. (2020) showed that artificial intelligence was capable of distinguishing COVID-19 pneumonia from other lung conditions with impressive accuracy.

These developments reflect both innovation and inequality in diagnostic efforts. While such technologies broadened the ways infections could be detected, their use was mostly confined to hospitals with advanced imaging equipment. As a result, wealthier regions gained greater benefits from these tools, further widening global gaps in testing capacity and outcomes.

### 3.1.4 Summary

Taken together, the research on COVID-19 testing shows how essential it was in controlling the pandemic, while also revealing major inequalities. Countries that began testing early and maintained frequent screening tended to contain outbreaks more effectively, whereas those that delayed testing often faced longer and more severe waves of infection. Differences in reporting systems and access to testing further complicated international comparisons. Altogether, this body of work points to the importance of unified and standardized analyses that connect testing patterns not just to case counts, but also to hospitalizations, recoveries, and deaths — the central aim of this study.

## 3.2 Pandemic Outcomes and Public Health Measures

COVID-19 outcomes—such as hospitalizations, recoveries, and deaths—capture both how severe outbreaks became and how well different interventions worked. Previous research has tied these outcomes to factors like healthcare capacity, medical progress, and policy decisions. However, many of these studies stop short of directly linking such outcomes to testing patterns, leaving an important gap in understanding how detection efforts shaped broader results.

### 3.2.1 Hospitalization and Recovery Patterns

Docherty et al. (2020) examined more than 20,000 patients in the United Kingdom to identify factors linked to severe illness and death, finding that age and preexisting health conditions were the strongest predictors. Richardson et al. (2020) analyzed 5,700 hospitalized patients in New York City and reached similar conclusions, emphasizing the influence of hospital capacity on patient outcomes. Grasselli et al. (2020) found that mortality rates in Lombardy, Italy, increased sharply when intensive care units became overwhelmed.

Together, these studies show that while individual health characteristics were crucial, the stability and capacity of healthcare systems often determined survival. This highlights the importance of large, standardized datasets that can capture both personal and structural factors shaping health outcomes across countries.

### 3.2.2 Vaccines and Treatments

Breakthroughs in vaccines and treatments changed the pandemic faster than almost anyone expected. Polack and colleagues (2020) reported that the Pfizer-BioNTech vaccine was roughly 95 percent effective, and results from the UK’s RECOVERY trial (2020) showed that dexamethasone helped reduce deaths in patients who were severely ill. These discoveries clearly saved lives, yet not everyone benefited at the same time. Wealthier countries secured early access to vaccines and medicine, while many others had to wait months before supplies arrived. That uneven rollout made it harder to compare results across nations and showed why consistent, shared data are so important for studying global health outcomes.

### 3.2.3 Policy Interventions and Interaction with Testing

Flaxman et al. (2020) estimated that lockdowns across Europe prevented millions of deaths, and Hsiang et al. (2020) expanded this analysis worldwide, showing that broad containment measures slowed the growth of COVID-19 cases. Both studies noted that such policies were most effective when paired with large-scale testing efforts. Early detection allowed governments to focus restrictions more precisely and lift them sooner. The World Health Organization (2020) added real-world context by releasing frequent situation reports that documented when and how these policies were put in place.

Together, these studies show that testing and policy interventions worked hand in hand. However, much of the existing research examines them separately — focusing on testing or restrictions in isolation. This study brings the two together to better understand how testing shaped the impact and timing of policy responses.

### 3.2.4 Summary

Research on pandemic outcomes has shown how healthcare systems, medical progress, and public policies shaped global responses. Yet, many studies have paid less attention to the direct influence of testing on these results. Without connecting testing data to hospitalizations and recoveries, it is difficult to see the full picture. This study helps fill that gap by examining how testing rates relate to hospitalization, recovery, and mortality trends across countries.

## 3.3 Critical Evaluation

The existing research has greatly expanded what is known about COVID-19 and its global impact. Across most studies, testing stands out as one of the main factors behind slowing transmission and lowering death rates. Early and widespread testing is linked to better outcomes, while the capacity of national healthcare systems often determines how well severe cases are managed. Progress in vaccines, treatments, and policy measures also reshaped the course of the pandemic, showing how coordinated scientific and government action can save lives.

Even so, several gaps remain. Inconsistent reporting across countries makes comparisons difficult, and many studies examine only one aspect—such as testing, hospitalization, or policy—without considering how these elements interact. Research has also leaned heavily toward high-income nations, leaving inequalities underexplored.

This study aims to fill those gaps by using standardized data from the COVID-19 Worldwide Testing Dataset to explore how testing patterns relate to major pandemic outcomes. Through cross-country panel analysis, it offers a more complete view of how testing influenced recovery, hospitalization, and mortality over time. In doing so, the research contributes both to academic understanding and to practical efforts to strengthen future public health responses.

# IV. Methods

## 4.1 Dataset

This study draws on the COVID-19 Worldwide Testing Dataset from Kaggle, which compiles country-level daily data on testing, confirmed cases, recoveries, and deaths. The dataset spans the early months of the pandemic through late 2022 and includes information from both developed and developing countries. It brings together data from national health ministries, the World Health Organization (WHO), and *Our World in Data*, making it one of the most complete global sources for tracking pandemic indicators.

Data preprocessing played a key role in improving accuracy and consistency. Records with missing or inconsistent values were cleaned using a forward–backward imputation approach, filling gaps based on surrounding trends within each country’s timeline. Entries containing negative numbers, often the result of data corrections or reporting errors, were removed. To smooth out daily fluctuations in reporting, seven-day rolling averages were applied to core variables such as tests performed, new positive cases, and active infections. The final cleaned dataset offers a stable and reliable picture of pandemic dynamics across regions and over time.

## 4.2 Tools and Software

All analyses were conducted using **RStudio (version 4.3.1)**, a statistical programming environment widely used for data analysis and visualization. The following R packages were utilized:

* **tidyverse** – for data wrangling and visualization
* **lubridate** – for managing and formatting date variables
* **zoo** – for rolling averages and missing-value imputation
* **fixest** – for panel data regression with fixed effects
* **modelsummary** and **gt** – for generating regression tables and formatted output
* **boot** – for bootstrap resampling and confidence interval estimation
* **ggplot2** – for producing visualizations such as histograms, scatter plots, and distributed-lag graphs

These tools were chosen for their reproducibility, open-source accessibility, and suitability for large-scale time-series panel data analysis.

## 4.3 Procedures

The data processing and analysis followed four main steps:

1. **Data Cleaning and Structuring**

The dataset was filtered to include only national totals (excluding state- or province-level data) to maintain comparability across countries. Two columns with excessive missing data—*hospitalized* and *hospitalized current*—were removed. Missing values for numeric variables were imputed using forward and backward filling within each country group.

1. **Variable Construction**

To measure dynamic relationships, several new variables were created:

* + **7-day rolling averages** for testing, positive cases, and active cases
  + **Growth rates** for active and positive cases, computed as the daily percentage change
  + **Lagged variables** (e.g., testing lagged by seven days) to analyze delayed effects of testing
  + **Death and recovery growth rates** to study downstream health outcomes

1. **Model Estimation**

Fixed-effects panel regressions were used to control for unobserved country-specific and time-specific differences. This model isolates the within-country relationship between testing and case growth over time while holding constant broader global shocks. The key model took the form:

\text{Growth}{it} = \beta\_1 \text{Lagged Testing}{it-7} + \alpha\_i + \gamma\_t + \epsilon\_{it}

where \alpha\_i and \gamma\_t represent country and time fixed effects, respectively.

Robustness checks included:

* + **Bootstrap resampling** (1,000 draws) to estimate confidence intervals around regression coefficients.
  + **Jackknife analysis**, which re-estimated the model after removing one country at a time, ensuring results were not driven by any single observation.
  + **Placebo tests**, substituting future testing rates for lagged values to confirm that causality flowed from testing to case outcomes—not vice versa.

1. **Visualization and Interpretation**

Statistical results were complemented by several visual analyses, including distributed-lag plots, binned scatter plots, and confidence interval (CI) bands. These visualizations made it possible to observe the temporal relationship between testing rates and case growth while confirming regression-based findings.

## 4.4 Reproducibility

All procedures in this research were designed for reproducibility. The full R script, titled *Investigating COVID-19 Virus Trends.Rmd*, includes code for data import, cleaning, transformation, modeling, and visualization. Because the Kaggle dataset and all R packages used are publicly available, another researcher could fully replicate this analysis by downloading the same dataset and running the provided R Markdown file. The output includes standardized tables, summary statistics, and figures, ensuring transparency and consistency in the analytical process.

# V. Results

## 5.1 Overview of Findings

The results show clear evidence that higher testing rates were linked to slower growth in active COVID-19 cases over time. Findings from the fixed-effects models, supported by several robustness checks, align with the idea that early and widespread testing helped limit transmission. Countries that kept testing regularly tended to see fewer new infections and faster drops in active cases, while those that delayed or limited testing often faced longer periods of growth.

## 5.2 Fixed-Effects Regression Analysis

The regression analysis showed a clear negative and statistically significant link between testing levels and the growth of active COVID-19 cases across several model setups. In the baseline model, which used a one-day lag, higher daily testing was linked to measurable declines in active case growth. The relationship grew stronger with a seven-day lag, suggesting that the impact of testing becomes visible within about one to two weeks—a window that aligns with the virus’s average incubation and reporting period.

Models using seven-day rolling averages, which help smooth short-term fluctuations, produced nearly identical results. This consistency shows that the findings were not simply the result of temporary reporting changes or daily anomalies. When the models were extended to include new positive cases, deaths, and recoveries, the results showed:

* **Active Case Growth:** steady declines following higher lagged testing rates
* **New Case Growth:** slower increases after periods of expanded testing
* **Death Growth:** a negative link, suggesting fewer fatalities after testing surges
* **Recovery Growth:** a positive link, implying that earlier detection and isolation supported better recovery outcomes

Taken together, these results strengthen the view that widespread testing played a major role in slowing transmission and improving overall health outcomes.

## 5.3 Robustness Checks

To check the reliability of the results, several robustness and validation procedures were carried out:

* **Bootstrap Resampling:**

A 1,000-draw bootstrap was used to estimate the distribution of the main coefficient for lagged testing. The average estimate stayed negative, and the 95 percent confidence interval did not include zero. This consistency supports the conclusion that testing and case growth are reliably and significantly related.

* **Jackknife Stability Test:**

The model was re-estimated while removing one country at a time. The coefficients changed very little across these runs, showing that no single country had an outsized influence on the overall findings.

* **Placebo Test:**

Lagged testing was replaced with a “placebo” variable using future testing rates. This substitution produced no significant relationship with case growth, reinforcing the idea the testing helps reduce cases rather than simply reacting to them.

* **Confidence Interval (CI) Bands:**

A global mean growth plot with 95 percent confidence intervals showed a clear decline in active-case growth as testing intensity increased, providing additional visual confirmation of the regression results.

## 5.4 Visual Summaries

Although the full visual outputs were generated in R, their key insights are summarized below:

* **Figure 1: Regression Summary Table**

A close-up of a report

AI-generated content may be incorrect.

When looking at the regression output, the main pattern is hard to miss. Countries that carried out more testing generally saw slower growth in both active and new COVID-19 cases. This trend stayed the same across different model setups, including those that used seven-day rolling averages to smooth noisy data. The estimated effects were modest but steady, suggesting that regular testing helped limit, though not erase, transmission. By controlling for differences between countries and across time, the fixed-effects approach made the results more trustworthy. The R² values, roughly between 0.28 and 0.45, show that the models explained a fair share of the variation. Overall, nations that kept testing consistently seemed to manage outbreaks better than those that didn’t keep up.

* **Figure 2: Bootstrap Distribution**

A graph of a graph showing a number of test results with Willis Tower in the background

AI-generated content may be incorrect.

The bootstrap results give a clear picture of how stable the relationship is between lagged testing and active case growth. Using 1,000 resamples, the distribution centers on a slightly negative average, showing that higher testing rates are consistently linked to slower growth in active cases, even when the model is rerun on different parts of the data. The shape of the histogram is fairly smooth and balanced, which suggests the results aren’t being skewed by outliers or a few influential countries. Taken together, the bootstrap analysis supports the reliability of the findings and shows that the negative link between testing and case growth is unlikely to be a random outcome.

* **Figure 3: Jackknife Distribution**

A graph of a graph

AI-generated content may be incorrect.

The jackknife results show how stable the estimated testing effect remains when one country is removed from the analysis at a time. The coefficients cluster closely around zero, which means that no single country is responsible for driving the overall link between lagged testing and active case growth. This consistency across many national samples suggests the relationship holds broadly, not just within a specific region. The compact shape of the distribution also implies that differences in testing systems or the timing of policy responses did not greatly affect the outcome. Overall, the jackknife test supports the idea that the model’s findings are reliable across countries and not dependent on a few outliers.

* **Figure 4: Distributed-Lag Effects**

A graph with lines and dots

AI-generated content may be incorrect.

The distributed-lag plot tracks how the impact of testing on active-case growth changes over time. The pattern shows that higher testing rates have their strongest effect about one to two weeks later, roughly matching the known infection and reporting cycles. In the early lags, the effects are small and somewhat uneven as new tests pick up recent infections. After about a week, the coefficients settle and turn clearly negative, showing that testing helps identify and isolate cases before wider spread occurs. The confidence intervals widen at longer lags, which is expected as uncertainty increases farther from the testing date. Overall, the plot suggests that testing does slow case growth, but its influence becomes most visible after a short delay.

* **Figure 5: Within-Country Binned Scatter Plot**

A graph with a line and a dotted line

AI-generated content may be incorrect.

The within-country binned scatter plot shows a clear negative link between testing rates and active-case growth after adjusting for country and time effects. Each point represents an average value within equal-sized bins of lagged testing, making it easier to see how higher testing levels align with lower growth in active cases. The fitted line slopes downward, indicating that once national context and time trends are accounted for, greater testing intensity consistently relates to slower epidemic spread. This pattern points to testing’s main advantage coming from early detection and isolation within countries, rather than from differences between them. Overall, the figure offers straightforward visual support for the regression and robustness results, showing that expanded testing directly helps slow transmission.

## 5.5 Interpretation of Results

Taken together, the analyses point to a clear trend: when testing increases, pandemic growth slows. The lag in the data suggests it takes roughly one to two weeks for the effects of testing to appear in case numbers. By catching infections earlier, countries could isolate positive cases sooner and break transmission chains before they spread further.

The combination of statistical modeling and visual evidence makes these results both convincing and reliable. The consistency across the bootstrap, placebo, and jackknife tests also shows that the relationships are not the result of quirks in the data or model setup. Overall, the findings underline that testing works best as a preventive measure, allowing governments to act early and reduce the severity of outbreaks.

# VI. Discussion

The results of this study point to a clear conclusion: countries that tested more frequently saw slower growth in active COVID-19 cases. This supports the idea that early and widespread testing helps slow transmission by finding infections sooner and allowing quicker isolation. Across all model versions, the coefficients for lagged testing stayed negative and consistent, showing a stable link between higher testing volume and lower case growth. The bootstrap and jackknife tests backed up these findings, confirming that the relationship held under repeated resampling and across different country groups. Taken together, the evidence suggests that testing’s impact on controlling outbreaks is both statistically reliable and broadly applicable across national contexts.

When compared with earlier studies, these results line up well with the work of Chinazzi et al. (2020) and Walker et al. (2020), who found that early, large-scale testing was key to reducing transmission. Like those studies, this analysis shows that testing success depends not only on how much testing is done but also on when it happens. What this study adds is a clearer look at timing—it measures the lag between testing and changes in case growth, showing that the strongest effects appear about one to two weeks later, which fits known infection and reporting cycles. While other research, such as Li et al. (2020) and Hasell et al. (2020), focused on underreporting and data consistency, this work demonstrates that standardized global datasets can still produce stable, interpretable results even with cross-country variation.

This study also helps connect testing trends more directly with epidemiological outcomes instead of treating them as separate topics. That link complements findings from research on hospitalizations (Docherty et al., 2020) and public-health policies (Flaxman et al., 2020), suggesting that widespread testing supports and strengthens other interventions. Countries that maintained high testing rates likely eased pressure on their healthcare systems by catching infections earlier in the transmission chain.

Of course, several limits should be noted. The analysis relies on observational data, so strong causal claims aren’t possible — countries that test more might also enforce other policies that reduce spread. Testing definitions may still vary across regions, even in standardized datasets. And while the panel regressions with fixed effects help control for unobserved differences, they can’t fully capture variations in healthcare infrastructure or public compliance. Future work could combine testing data with more detailed hospitalization or vaccination records and use causal methods such as instrumental variables or structural models to tease out direct effects.

In short, the evidence strongly supports the main hypothesis: greater and timelier testing is linked to slower pandemic growth. These results reinforce that testing should remain a core element of pandemic preparedness and offer practical guidance for designing faster, data-driven responses in future global health crises.

# VII. Conclusion

This study set out to test whether higher COVID-19 testing rates helped slow the growth of active cases across countries, using a large and comprehensive global dataset. The results strongly support that idea: countries that tested more, and did so earlier, generally saw smaller increases in case numbers. Across several regression models and robustness checks—including bootstrap, placebo, and jackknife tests—the findings held steady. The evidence shows that expanded testing helps catch infections sooner, allowing quicker isolation and reducing the speed of transmission. The distributed-lag and binned scatter plots also showed that these effects usually appear after about one to two weeks, matching known infection and reporting timelines.

The broader message goes beyond COVID-19. These results underline the value of widespread testing as a proactive tool for controlling infectious disease outbreaks. Effective testing does more than just find cases; it helps governments target restrictions, manage hospital capacity, and avoid the deeper social and economic costs of long lockdowns. The findings echo earlier work by Chinazzi et al. (2020) and Walker et al. (2020) but go further by measuring testing’s impact over time and confirming its stability with updated statistical approaches.

By linking testing data directly to public health outcomes, this study provides a foundation for future research and decision-making. The results make clear that investing in diagnostic capacity and real-time data systems is not just a reaction to crisis—it is an essential part of readiness. As countries prepare for the next pandemic, early and coordinated testing efforts will remain one of the most effective ways to achieve faster, fairer, and more sustainable control of disease spread.

# VIII. Future Directions

Building on the results of this study, future research can move in several directions to better understand how testing shapes pandemic outcomes. First, expanding the dataset to include indicators of healthcare capacity, vaccination rates, and government interventions would provide a fuller view of how testing interacts with other public health strategies. Adding hospitalization and mortality data at more detailed time intervals could also clarify the pathways linking testing activity to clinical results.

Second, researchers should make greater use of causal inference techniques such as instrumental-variable models, synthetic control designs, or structural equation approaches. These methods would help separate the direct effects of testing from those of related policy measures, offering clearer guidance for decision-makers.

Third, future studies could strengthen cross-regional comparisons by including social and demographic variables. Understanding how factors like income inequality, population density, and access to healthcare shape testing outcomes would help explain why some countries benefited more than others.

Finally, advances in data collection and digital infrastructure open the door to real-time monitoring of testing effectiveness. Combining high-frequency epidemiological data with predictive modeling could allow governments to track and adjust testing strategies dynamically during future outbreaks. By applying these lessons, researchers and policymakers can turn the experience of COVID-19 into a blueprint for faster, more coordinated responses to the next global health emergency.

# List of References

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., … & Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science, 368(6489), 395–400. <https://doi.org/10.1126/science.aba9757>

Docherty, A. B., Harrison, E. M., Green, C. A., Hardwick, H. E., Pius, R., Norman, L., … & Semple, M. G. (2020). Features of 20,133 UK patients in hospital with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol: Prospective observational cohort study. BMJ, 369, m1985. <https://doi.org/10.1136/bmj.m1985>

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., … & Bhatt, S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature, 584(7820), 257–261. <https://doi.org/10.1038/s41586-020-2405-7>

Grasselli, G., Zangrillo, A., Zanella, A., Antonelli, M., Cabrini, L., Castelli, A., … & Pesenti, A. (2020). Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy region, Italy. JAMA, 323(16), 1574–1581. <https://doi.org/10.1001/jama.2020.5394>

Hasell, J., Mathieu, E., Beltekian, D., Macdonald, B., Giattino, C., Ortiz-Ospina, E., … & Roser, M. (2020). A cross-country database of COVID-19 testing. Scientific Data, 7(1), 345. <https://doi.org/10.1038/s41597-020-00688-8>

Hsiang, S., Allen, D., Annan-Phan, S., Bell, K., Bolliger, I., Chong, T., … & Wu, T. (2020). The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature, 584(7820), 262–267. <https://doi.org/10.1038/s41586-020-2404-8>

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., & Shaman, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science, 368(6490), 489–493. <https://doi.org/10.1126/science.abb3221>

Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., … & Xia, J. (2020). Artificial intelligence distinguishes COVID-19 from community-acquired pneumonia on chest CT. Radiology, 296(2), E65–E71. <https://doi.org/10.1148/radiol.2020200905>

Our World in Data. (2021). Coronavirus (COVID-19) testing dataset. <https://ourworldindata.org/coronavirus-testing>

Peto, J. (2020). Covid-19 mass testing facilities could end the epidemic rapidly. BMJ, 368, m1163. <https://doi.org/10.1136/bmj.m1163>

Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., … & Gruber, W. C. (2020). Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. New England Journal of Medicine, 383(27), 2603–2615. <https://doi.org/10.1056/NEJMoa2034577>

RECOVERY Collaborative Group. (2020). Dexamethasone in hospitalized patients with Covid-19 — Preliminary report. New England Journal of Medicine, 384(8), 693–704. <https://doi.org/10.1056/NEJMoa2021436>

Richardson, S., Hirsch, J. S., Narasimhan, M., Crawford, J. M., McGinn, T., Davidson, K. W., … & Northwell COVID-19 Research Consortium. (2020). Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA, 323(20), 2052–2059. <https://doi.org/10.1001/jama.2020.6775>

Walker, P. G. T., Whittaker, C., Watson, O. J., Baguelin, M., Ainslie, K. E. C., Bhatia, S., … & Ferguson, N. M. (2020). The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. Science, 369(6502), 413–422. <https://doi.org/10.1126/science.abc0035>

World Health Organization. (2020). Coronavirus disease (COVID-19) situation reports. World Health Organization. [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports?utm_source=chatgpt.com)

# Link to the References

[Investigating COVID-19 Virus Trends](https://indiana-my.sharepoint.com/:f:/r/personal/dzcung_iu_edu/Documents/Investigating%20COVID-19%20Virus%20Trends?csf=1&web=1&e=wQJdjj)