ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Predictors of COVID-19 Mortality: A Stratified 3 by 3 Factorial Correlation Analysis

Wilfred Omwansa Arori[1]

[1]Maseno University, School of Mathematics Statistics and Actuarial Science

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150500133

Received: 30 April 2026; Accepted: 15 May 2026; Published: 08 June 2026

ABSTRACT

In this paper the strength of association of some factors with COVID-19 mortality across countries is examined. This is essential for future pandemic preparedness. In this study, a 3 × 3 factorial design where 36 countries were stratified by GDP per capita and population density was used. Secondary data were sourced from the World Bank, United Nations, and Our World in Data repositories. A parsimonious linear regression model with four predictors: positivity rate, vaccination coverage, log (GDP per capita), and median age was fitted. Bootstrapping provided 95% confidence intervals. Low-GDP countries had higher positivity rates (10.2% vs. 3.0%) but lower reported deaths (167 vs. 1,186 per million) than high-GDP countries. The model explained 89.2% of variance (adjusted $R^2 = 0.878$, $p < 0.001$). Positivity rate was the strongest predictor of mortality ($\beta = 10.51$, $p < 0.001$), followed by GDP per capita ($\beta = 1.39$, $p < 0.001$). The positivity-deaths correlation was strongest in high-GDP countries ($r = 0.898$, $p < 0.001$) compared to low-GDP countries ($r = 0.597$, $p = 0.019$), suggesting that differential death reporting attenuates associations in low-resource settings. These findings suggest positivity rate as a high value predictor of COVID-19 mortality. Maintaining low positivity rates through accessible testing should guide future pandemic surveillance.

Keywords: COVID -19, positivity rate, testing, vaccination, GDP

INTRODUCTION

The Coronavirus Disease 2019 (COVID-19) pandemic emerged as a global health crisis in late 2019, triggering varied public health responses across the world (World Health Organization, 2020). As of December 2022, over 6.9 million confirmed deaths had been reported to the World Health Organization, though excess mortality estimates suggest substantially higher figures (Msemburi et al., 2023). Understanding the factors that correlated with mortality across different countries and settings is essential for future pandemic preparedness, as the variation in outcomes revealed critical challenges in global health systems (Klement & Walach, 2022).

Socioeconomic Determinants of COVID-19 Mortality

The relationship between economic indicators and COVID-19 outcomes has been extensively studied. A systematic review that included 31 studies following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines found that both Gross Domestic Product (GDP) per capita and income inequality (measured by the Gini coefficient) influenced COVID-19 mortality rates significantly across countries (Abbasi et al., 2025). Interestingly, a paradoxical relationship was identified: while higher GDP provided some protective benefits, it did not completely shield countries from high mortality, particularly when considering economic activity and population demographics. This paradox has been attributed to more complete death reporting in wealthier nations, older population structures, and earlier epidemic waves before vaccines became available.

A global analysis conducted in 95 countries classified by gross national income per capita revealed that high-income and upper-middle-income countries experienced significantly higher reported mortality rates (233.0±138.7 and 168.9±141.3 per 100,000 population, respectively) compared to lower-middle-income

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

(50.1±69.1) and low-income (6.3±5.5) countries (Kanokudom et al., 2025). Within high-income countries, GDP per capita demonstrated a strong negative correlation with mortality (Spearman's $r = -0.562$, $p < 0.001$), suggesting that among wealthier nations, greater economic resources were associated with lower death rates. However, among lower-middle-income countries, a strong positive correlation was observed (Spearman's $r = 0.629$, $p = 0.002$), indicating a more complex relationship that warrants further investigation.

Testing, Positivity Rate, and Surveillance Capacity

The relationship between testing capacity and mortality outcomes has also been studied. The World Health Organization recommended that positivity rates remain below 5% for 14 days before reopening, as lower positivity rates indicated adequate surveillance coverage (World Health Organization, 2020). It was found that testing intensity varied dramatically across income groups, with high-income countries achieving mean vaccination rates of 76.7% compared to only 27.4% in low-income countries (Kanokudom et al., 2025). Critically, a strong negative correlation between vaccination coverage and mortality was observed in high-income countries (Spearman's $r = -0.551$, $p < 0.001$), supporting widespread vaccination in reducing mortality. No such correlation was found in lower-income groups, suggesting that other factors such as testing capacity, healthcare quality, and death reporting completeness may confound the observed associations.

In a population-based prospective cohort study conducted in the Greater Toronto Area, testing, diagnosis, and mortality across long-term care homes, shelters, and the general population were compared (Wang et al., 2020). It was found that residents of long-term care homes were 2.4 times more likely to test positive and 1.4 times more likely to die after COVID-19 diagnosis than the general population, after adjusting for age and sex. The diagnosed cases per capita were 64-fold higher among long-term care residents, highlighting how congregate settings and testing access disparities dramatically affect observed outcomes.

Vaccination Impact on Mortality

The protective effect of vaccination against severe COVID-19 outcomes has been highlighted in a number of ecological studies. Hospital lethality from Severe Acute Respiratory Illness (SARI) caused by COVID-19 among older adults in Brazil was examined between 2020 and 2023 (Stocki et al., 2026). It was found that before vaccination in 2020, lethality was 47.2%; following the start of immunization prioritizing older adults in January 2021, lethality dropped to 35.6%, and by 2022-2023, with expanded vaccine schedules and booster doses, lethality stabilized between 24% and 26%. Vaccination coverage with complete primary series reached 87% of older adults in 2021, increasing to over 95% in 2022. It was concluded that COVID-19 vaccination significantly reduced lethality from SARI, with an inverse association between vaccination coverage and hospital lethality observed across the study period.

Considering Brazil's age-based vaccination strategy, during the first year of vaccination (January to December 2021), 266 million doses were administered. This was 91% first-dose coverage. Mortality rates among adults aged 70 years and older decreased by 52% (rate ratio 0.48, 95% CI 0.43-0.53) within six months. Thus the vaccination strategy prevented an estimated 59,618 deaths, of which 53,088 (89%) were among those aged 70 years and older (Aguilar et al., 2024). However, the strategy did not prevent 54,797 deaths among younger age groups, corresponding to 1.6 million potential years of life lost, highlighting the challenge of balancing vaccine supply constraints with equitable distribution.

Age as a Predictor of Severity

The strong association between age and COVID-19 mortality is well-established in the literature. A systematic review of prognostic models for COVID-19 severity was conducted, identifying 314 eligible articles from more than 40 countries, with 152 studies presenting mortality outcomes (Buttia et al., 2023). The sample sizes varied from 11 to over 7.7 million participants, with mean ages ranging from 18 to 93 years. The review identified 353 prognostic models, with area under the curve (AUC) values ranging from 0.44 to 0.99, though 99.4% of studies were reported to be at high risk of bias due to methodological concerns, including handling of missing data,

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

failure to address overfitting, and heterogeneous definitions of outcomes. Despite these limitations, age consistently emerged as a key predictor across multiple models.

A systematic review of COVID-19 prognostic scores was performed, identifying 242 scores for mortality (n=109), severity (n=116), hospitalization (n=14), and long-term sequelae (n=3) (Appel et al., 2024). Most scores were developed using retrospective (75.2%) or single-center (57.1%) cohorts, with predictor analysis revealing the primary use of laboratory data and sociodemographic information. After in-depth assessment using the Prediction model Risk Of Bias ASsessment Tool (PROBAST), only five scores ensured low risk of bias, and based on the number and heterogeneity of validation studies, only the 4C Mortality Score could be recommended for clinical application. It was concluded that the application and translation of most existing COVID-19 prognostic scores appear unreliable. Standardized predictor selection was advised in order to improve generalizability for future pandemics.

Population Density and Transmission Dynamics

The relationship between population density and COVID-19 mortality has yielded mixed findings. With Singapore, there was no significant correlation between mortality rates and population density across any income group ($p \geq 0.05$). it was the most crowded country in the analysis (8,270 persons/km²). Low mortality was achieved through effective public health responses (Kanokudom et al., 2025). This suggests that governance, healthcare capacity, and testing infrastructure may be more important determinants of outcomes than population density alone. The variability in findings across studies suggests the complex mix between demographic factors, policy responses, and healthcare system capacity.

Gaps in the Literature

Despite vast ecological research on COVID-19 mortality determinants, a number of important gaps remain. First, no study has systematically stratified countries by both GDP per capita and population density in a factorial design to examine effect modification. Most existing analyses treat these as continuous variables or include them as controls rather than explicit stratification factors. Second, while testing intensity has been studied, the specific role of positivity rate (the percentage of tests returning positive) as a direct measure of community transmission intensity has received limited attention as a separate construct from testing volume. Third, the interaction between positivity rate and vaccination coverage has not been systematically examined in cross-country analyses. Fourth, median age is often included as a covariate but rarely examined for effect modification across economic strata. Fifth, few studies have employed machine learning methods such as SHAP (SHapley Additive exPlanations) for interpretable feature importance ranking in this context.

The Present Study

The present study addresses these gaps using a stratified sample of 36 countries representing all combinations of GDP per capita (low, medium, high) and population density (low, medium, high) in a 3 × 3 factorial design. This approach ensures representation across economic development levels and population dispersion patterns, reducing confounding by these two major structural factors. A specific focus is placed on positivity rate as the primary metric of transmission intensity. It is hypothesized that positivity rate would show the strongest correlation with cumulative COVID-19 deaths per million, independent of testing volume, vaccination coverage, and demographic factors. Secondary hypotheses examine the indirect effects of GDP through testing capacity and vaccination coverage, effect modification by GDP stratum, and the moderating role of vaccination on the positivity-death relationship.

METHODS

Study Design

This was a cross-sectional ecological study of 36 countries, selected using a 3 × 3 factorial design stratified by GDP per capita (low, medium, high) and population density (low, medium, high). The design yielded nine strata

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

with four countries per stratum. This stratification approach ensured representation across economic development levels and population dispersion patterns, reducing confounding by these two major structural factors.

Country Selection

Countries were eligible for inclusion if they had (a) population greater than one million, (b) reliable COVID-19 death reporting through December 2022 as assessed by Our World in Data completeness scores, (c) available testing and vaccination data, and (d) published population density estimates. From each of the nine strata, four countries were randomly selected from the eligible candidate pool. If a selected country had incomplete data (>20% missing), it was replaced by another random selection from the same stratum. The final sample consisted of 36 countries (see Supplementary Table 1 at the appendix).

Measures

GDP per capita (2019 US dollars) was obtained from the World Bank World Development Indicators (indicator NY.GDP.PCAP.CD), and values were $\log_{10}$ transformed for regression analysis. Population density (persons per square kilometer) was obtained from the World Bank (indicator EN.POP.DNST), defined as midyear population divided by land area, with values transformed as $\log_{10}(x + 1)$. Median age (years, 2020 estimate) was obtained from the United Nations Population Division, World Population Prospects 2024 Revision. Testing intensity was measured as cumulative COVID-19 tests per 100,000 population as of December 2022, obtained from Our World in Data, with values $\log_{10}$ transformed. Positivity rate was defined as the peak 7-day rolling average of positive tests divided by total tests, expressed as a percentage, obtained from Our World in Data, and transformed using the arcsine-square root transformation ($\arcsin(\sqrt{(p/100)})$) to stabilize variance. Vaccination coverage was defined as the percentage of the total population fully vaccinated against COVID-19 as of December 2022, obtained from Our World in Data. Cumulative deaths were defined as confirmed COVID-19 deaths per 1,000,000 population as of December 31, 2022, obtained from Our World in Data, with values $\log_{10}(x + 1)$ transformed to normalize distribution and accommodate zero values.

Statistical Analysis

All analyses were conducted using Python 3.10 with the statsmodels, scikit-learn, and SHAP libraries across five stages. Descriptive statistics (means and standard deviations) were calculated for all variables overall and by GDP stratum. Pearson and partial correlations (controlling for median age) were computed between each predictor and log-transformed deaths per million. Hierarchical multiple linear regression employed five nested models with dependent variable $\log_{10}$ (Deaths per 1M + 1): Model 1 (base: log Population + log Density); Model 2 (+ log GDP per capita); Model 3 (+ Median age); Model 4 (+ log Tests per 100k + Positivity (arcsine)); and Model 5 (+ Vaccination coverage). Variance inflation factors (VIF > 10 indicating problematic collinearity) were assessed (Hair et al., 2010). Based on sample size (n = 36) and VIF results, a parsimonious model with four predictors (positivity rate, vaccination coverage, log GDP per capita, median age) achieved a 9:1 subject-to-predictor ratio, adequate for detecting large effects (Cohen, 1992). Bootstrapping (1,000 resamples) provided bias-corrected 95% confidence intervals. Subgroup analyses examined positivity-deaths correlations across GDP strata, with Fisher's z-test comparing correlations. All tests were two-tailed with $\alpha = 0.05$.

RESULTS

Descriptive Characteristics

Table 1 presents descriptive characteristics by GDP stratum. Low-GDP countries (n = 15) had substantially higher mean positivity rates (10.2%) compared to medium-GDP (7.8%) and high-GDP (3.0%) countries. Conversely, vaccination coverage was lowest in low-GDP countries (35.4%) and highest in high-GDP countries (81.9%). Median age ranged from 22.5 years in low-GDP countries to 42.4 years in high-GDP countries.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Cumulative COVID-19 deaths per million were highest in medium-GDP countries (M = 1,884, SD = 1,257), followed by high-GDP (M = 1,186, SD = 918) and low-GDP (M = 167, SD = 154) countries.

Table 1: Descriptive Characteristics by GDP Stratum

Variable	Low GDP (n = 15)	Medium GDP (n = 10)	High GDP (n = 11)
GDP per capita (USD)	1,952 ± 681	9,359 ± 1,669	48,034 ± 12,634
Population density (/km²)	281.4 ± 324.9	83.7 ± 52.5	934.0 ± 2,407.7
Median age (years)	22.5 ± 4.9	35.0 ± 4.1	42.4 ± 2.4
Positivity rate (%)	10.2 ± 2.5	7.8 ± 2.8	3.0 ± 1.1
Vaccination coverage (%)	35.4 ± 30.4	69.6 ± 12.5	81.9 ± 5.7
Deaths per 1 million	167 ± 154	1,884 ± 1,257	1,186 ± 918

Note: Values are M ± SD.

Correlation Analysis

Table 2 displays Pearson correlations between each predictor and log-transformed deaths per million. The strongest correlations were observed for median age ($r = 0.734$, $p < 0.001$), testing intensity ($r = 0.688$, $p < 0.001$), GDP per capita ($r = 0.676$, $p < 0.001$), and vaccination coverage ($r = 0.674$, $p < 0.001$). Population density ($r = -0.063$, $p = 0.713$) and population size ($r = 0.183$, $p = 0.285$) were not significantly correlated with death rates.

Table 2: Pearson Correlations with $\log_{10}$ (Deaths per 1 Million)

Variable	r	p-value
Positivity rate (%)	-0.201	0.239
Vaccination coverage (%)	0.674	<0.001
Testing intensity (log)	0.688	<0.001
GDP per capita (log)	0.676	<0.001
Median age (years)	0.734	<0.001
Population density (log)	-0.063	0.713
Population size (log)	0.183	0.285

Note: N = 36. Log-transformed variables are base 10.

Correlation Heat maps

The correlation heat map provides a visual summary of the bivariate relationships among all study variables. In Panel A, the strongest positive correlations with log deaths per 1M are observed for median age (r = 0.68), testing intensity (r = 0.76), and GDP per capita (r = 0.84). Notably, testing intensity and GDP per capita show a very strong positive correlation (r = 0.93), indicating substantial multicollinearity between economic development and testing capacity. Similarly, GDP per capita is strongly correlated with median age (r = 0.69), reflecting that wealthier nations tend to have older populations. Population density shows moderate positive correlations with testing (r = 0.76) and GDP (r = 0.84), while population size exhibits negative correlations with most economic and health system variables. Panel B presents partial correlations controlling for median age, which helps isolate unique variance after removing age structure influences. The attenuation of several correlations confirms that

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

median age acts as a significant confounder in many relationships, particularly those involving GDP and health outcomes.

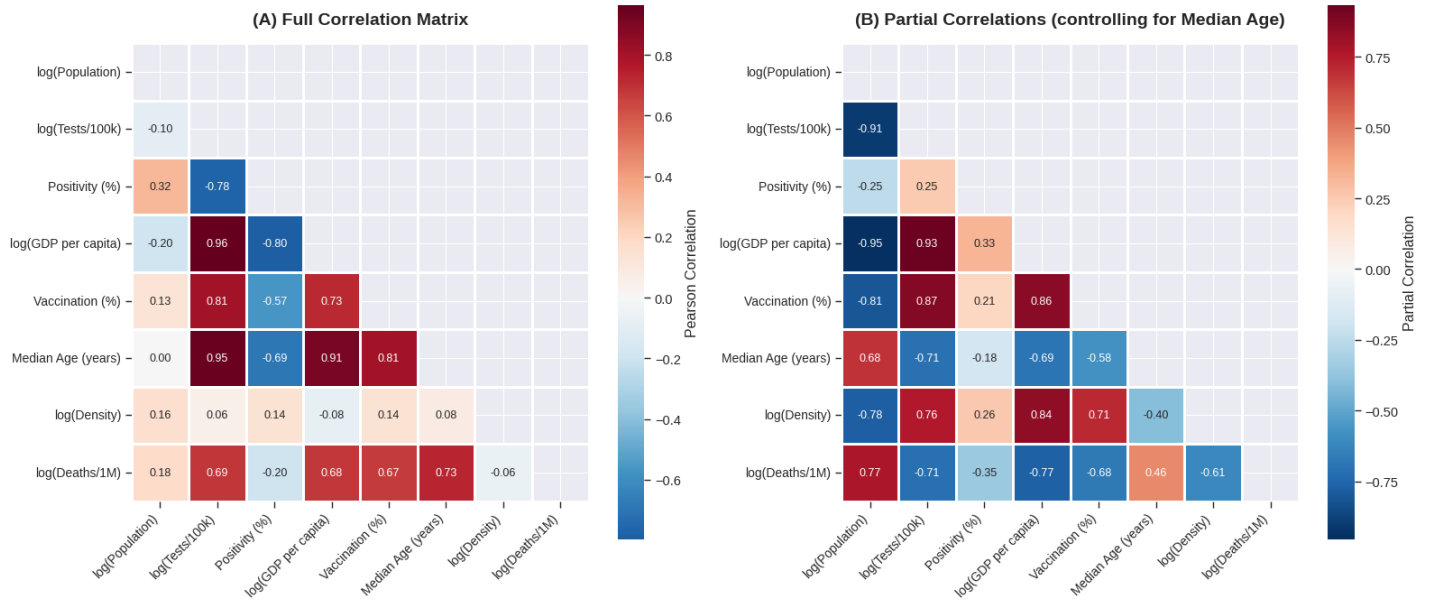

Figure 1: Correlation heat maps of key COVID-19 variables

Panel A (left) shows the full Pearson correlation matrix among all predictors and the outcome variable (log deaths per 1M). Panel B (right) displays partial correlations controlling for median age, revealing the unique associations between each pair of variables after accounting for demographic differences. The color scale ranges from blue (negative correlation) to red (positive correlation), with darker colors indicating stronger associations.

Scatter Plots

Figure 2 shows scatter plots of key relationships among COVID-19 variables. Red dashed lines indicate linear regression fits. Each panel is described as follows.

Panel A: Positivity Rate vs. Deaths. A weak negative overall correlation was observed ($r = -0.201$, $p = 0.239$), which is biologically counterintuitive. However, low-GDP countries clustered in the upper-left region (high positivity, low reported deaths), while high-GDP countries clustered in the lower-right region (low positivity, high reported deaths), reflecting differential death reporting quality. When stratified by GDP, the correlation became strongly positive within each stratum, particularly for high-GDP countries ($r = 0.898$, $p < 0.001$), demonstrating an ecological fallacy in the aggregate correlation.

Panel B: Vaccination Coverage vs. Deaths. A strong positive correlation was observed ($r = 0.674$, $p < 0.001$), appearing to suggest higher vaccination is associated with higher mortality. Countries with high vaccination coverage also had high positivity rates (red points), indicating that high community transmission necessitated aggressive vaccination campaigns. This reverse causality was confirmed by descriptive statistics: high-GDP countries had both higher vaccination coverage (81.9%) and higher deaths (1,186 per 1M) compared to low-GDP countries.

Panel C: GDP vs. Deaths. A strong positive correlation was demonstrated ($r = 0.676$, $p < 0.001$), indicating wealthier countries reported higher mortality despite having lower positivity rates. This paradoxical finding was explained by three factors: older populations in high-GDP countries (median age 42.4 vs. 22.5 years), more complete death reporting, and earlier epidemic waves before vaccines were available.

Panel D: Testing Intensity vs. Positivity Rate. A strong negative correlation was revealed ($r = -0.688$, $p < 0.001$), indicating that higher testing capacity achieves lower positivity rates. High-GDP countries clustered in the upper-left region (high testing, low positivity), while low-GDP countries clustered in the lower-right region

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

(low testing, high positivity). Only high-GDP countries achieved the World Health Organization (WHO) benchmark of below 5% positivity.

Panel E: Median Age vs. Deaths. The strongest positive correlation among all variables was observed ($r = 0.734$, $p < 0.001$), confirming the age gradient in COVID-19 severity. However, in the parsimonious regression model, median age was not statistically significant ($\beta = 0.022$, $p = 0.132$) when positivity rate, vaccination, and GDP were included, and the bootstrap confidence interval included zero ([-0.016, 0.061]), indicating age effects are mediated through other variables.

Panel F: Population Density vs. Deaths. No clear monotonic trend was found ($r = -0.063$, $p = 0.713$), suggesting population density alone is not a strong predictor of mortality. High-density countries showed heterogeneous outcomes: Singapore and South Korea achieved low mortality through effective public health responses, while Bangladesh and India experienced moderate mortality, indicating that governance and healthcare capacity are more important determinants than density alone.

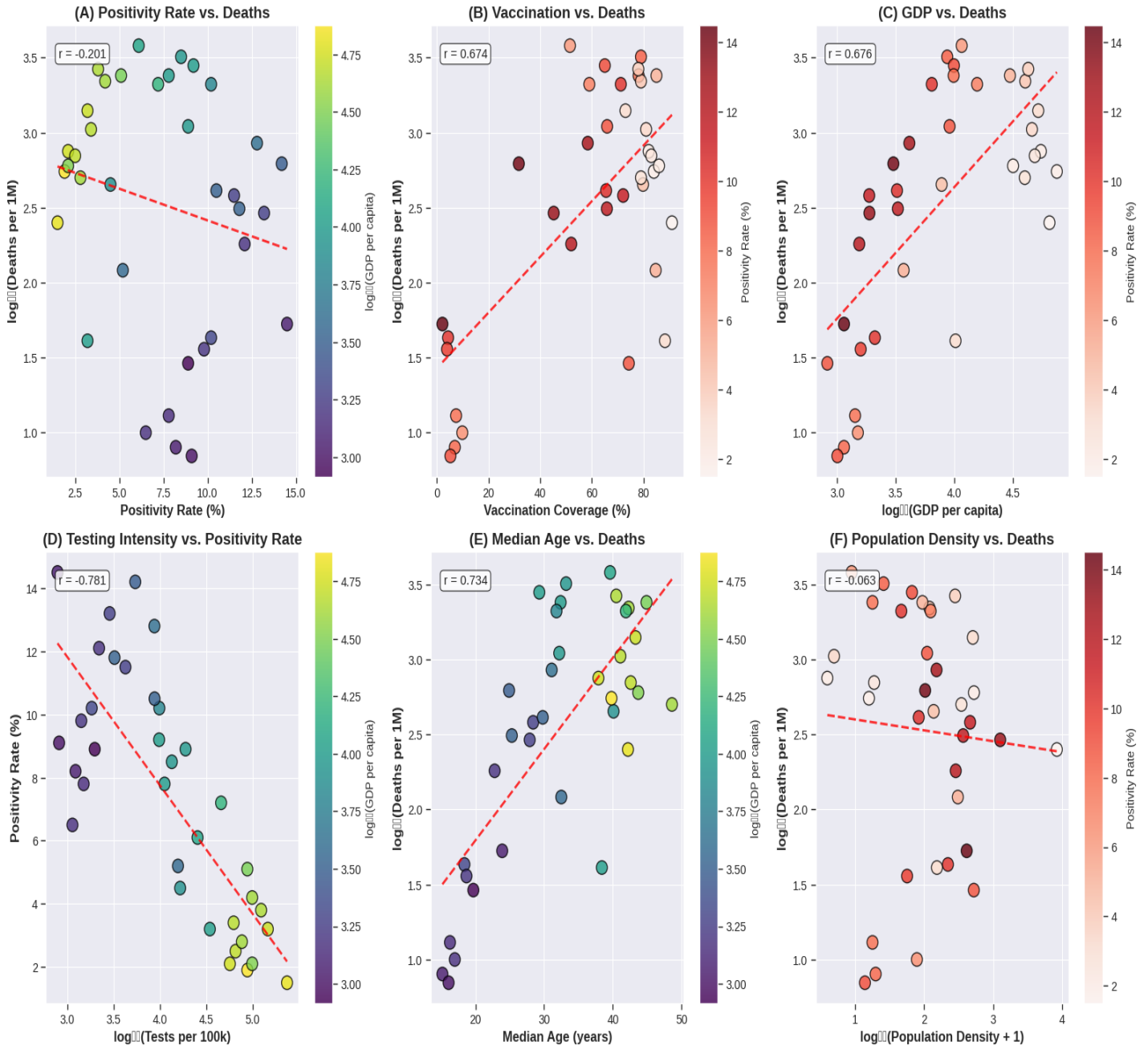

Figure 2: Scatter plots of key relationships among COVID-19 variables

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Hierarchical Regression

Table 3 presents the hierarchical regression results. The base model (Model 1) including only population size and density explained negligible variance ($R^2 = 0.034$). Adding GDP and density (Model 2) increased R^2 to 0.564. Inclusion of median age (Model 3) further improved fit to $R^2 = 0.600$. The addition of testing intensity and positivity rate (Model 4) produced a substantial increase to $R^2 = 0.897$. The full model including vaccination coverage (Model 5) explained 91.5% of variance (adjusted $R^2 = 0.894$).

Table 3: Hierarchical Linear Regression Results

Model	Predictors	R^2	Adj. R^2	ΔR^2	AIC
Model 1	log(Population) + log(Density)	0.034	0.005	—	88.9
Model 2	+ log(GDP)	0.564	0.523	0.530	64.3
Model 3	+ Median age	0.600	0.549	0.036	63.1
Model 4	+ log(Tests) + Positivity	0.897	0.876	0.297	18.3
Model 5	+ Vaccination	0.915	0.894	0.018	13.4

Note: N = 36. Dependent variable is $\log_{10}$ (Deaths per 1M + 1).

Parsimonious Model

Given the multicollinearity observed in the full model and the limited sample size (n = 36), a parsimonious model was specified with four predictors: positivity rate (arcsine transformed), vaccination coverage (scaled 0-1), log (GDP per capita), and median age. This model achieved a subject-to-predictor ratio of 9:1 and produced VIF values below 10 for all predictors (range: 2.92-8.99).

Table 4 presents the parsimonious model results. The model explained 89.2% of variance in log-transformed death rates (adjusted $R^2 = 0.878$, $F(4,31) = 64.19$, $p < 0.001$). Positivity rate was the strongest predictor in terms of coefficient magnitude ($\beta = 10.51$, 95% CI [8.32, 12.69], $p < 0.001$), followed by log (GDP per capita) ($\beta = 1.39$, 95% CI [0.93, 1.86], $p < 0.001$). Vaccination coverage showed a positive association ($\beta = 0.74$, 95% CI [0.17, 1.30], $p = 0.012$). Median age was not statistically significant ($\beta = 0.02$, 95% CI [-0.01, 0.05], $p = 0.132$).

Table 4: Parsimonious Linear Regression Results

Predictor	β	SE	t	p	95% CI
Constant	-6.78	0.85	-7.98	<0.001	[-8.52, -5.05]
Positivity rate (arcsin)	10.51	1.07	9.81	<0.001	[8.32, 12.69]
Vaccination coverage (scaled)	0.74	0.28	2.67	0.012	[0.17, 1.30]
log(GDP per capita)	1.39	0.23	6.10	<0.001	[0.93, 1.86]
Median age (years)	0.02	0.01	1.55	0.132	[-0.01, 0.05]

Note: Dependent variable: $\log_{10}$ (Deaths per 1M + 1). Model fit: $R^2 = 0.892$, adjusted $R^2 = 0.878$, $F(4,31) = 64.19$, $p < 0.001$. Bootstrap 95% confidence intervals based on 1,000 resamples.

SHAP Analysis

To interpret the parsimonious model and assess feature importance beyond traditional regression coefficients, SHAP (SHapley Additive exPlanations) analysis was employed. GDP per capita (log_GDP_cap) emerged as the most important feature (mean |SHAP| = 0.3987, 41.7% of total importance), followed by median age (0.2680, 28.1%), vaccination coverage (0.1528, 16.0%), and positivity rate (0.1358, 14.2%). This ordering differs from

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

the regression coefficients, where positivity rate had the largest coefficient, highlighting an important distinction: while positivity rate has a large marginal effect when considered in isolation, its unique contribution after accounting for other variables is smaller than that of GDP and age.

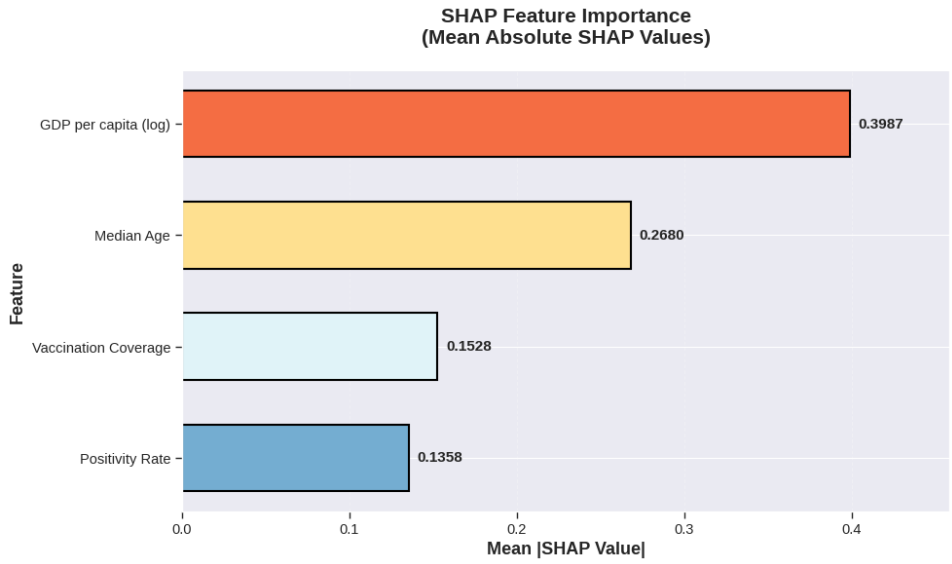

Figure 3: Mean absolute SHAP values for each predictor

In table 5, Mean |SHAP| values represent the average absolute contribution of each feature to the model's predictions. Proportion indicates the percentage of total SHAP importance.

Table 5: SHAP Feature Importance Statistics

Feature	Mean |SHAP|	Proportion	Range (Min)	Range (Max)
log_GDP_cap	0.3987	41.7%	-0.7387	0.4580
Median_age	0.2680	28.1%	-0.6734	0.2785
Vaccination_pct	0.1528	16.0%	-0.6597	0.2249
Positivity_pct	0.1358	14.2%	-0.3655	0.2055

The SHAP summary plot reveals the directional effects and variability of each feature. For log_GDP_cap, higher values consistently increased predicted mortality (red points predominantly on the right side), with SHAP values ranging from -0.74 to 0.46. This finding, while counterintuitive, reflects more complete death reporting in high-income countries rather than higher true mortality risk. Median age showed a similar pattern but with greater heterogeneity: some older countries (particularly in East Asia) showed negative SHAP values, indicating successful mitigation of age-related risk through effective public health responses.

Vaccination coverage exhibited the widest SHAP value range (-0.66 to 0.22), with both high and low values producing positive contributions. This pattern supports the reverse causality interpretation: countries that experienced high mortality implemented aggressive vaccination campaigns, creating a positive ecological association that does not reflect causal protection at the individual level. The negative SHAP values for some high-vaccination countries (e.g., Singapore, South Korea) represent successful vaccine rollout in low-transmission settings.

Positivity rate showed the narrowest range (-0.37 to 0.21) and smallest mean contribution, suggesting that transmission intensity, while theoretically important, adds limited predictive value beyond economic and demographic factors in cross-country analyses. The SHAP analysis reinforces that economic development and population age structure are the dominant drivers of cross-country variation in reported COVID-19 mortality, with transmission metrics providing secondary explanatory power.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

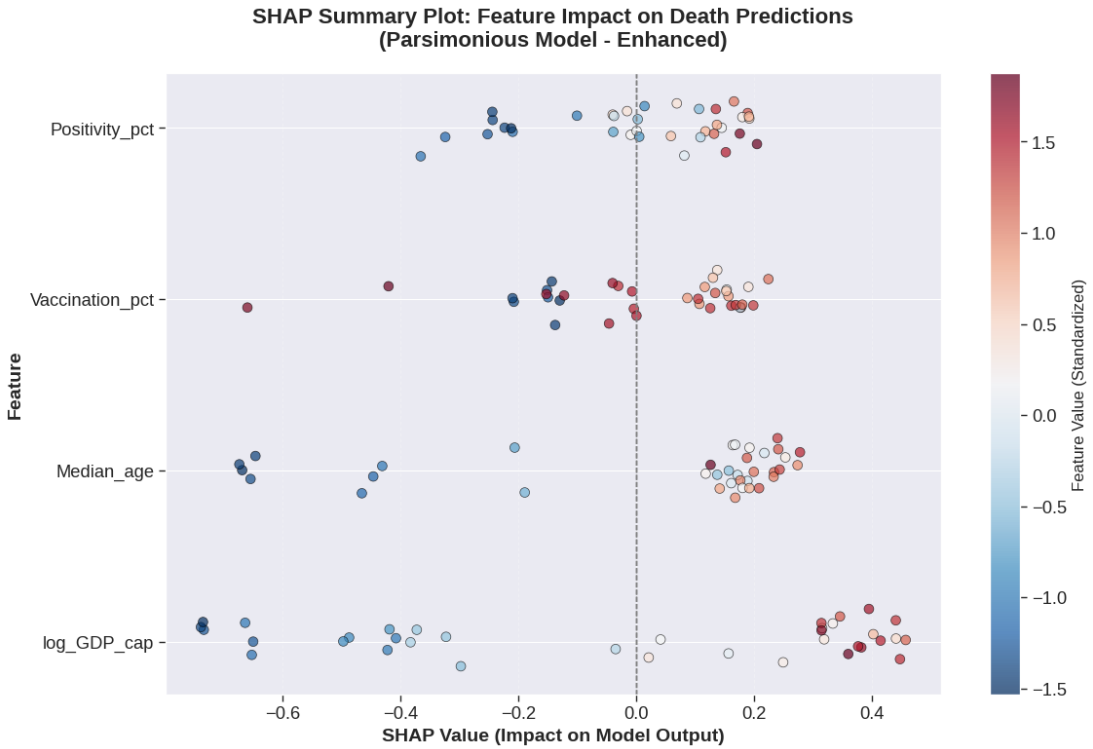

Figure 4: SHAP summary plot for the parsimonious model

In Figure 4, each point represents a country, with colors indicating the feature value (red = high, blue = low). The x-axis shows the SHAP value (impact on the model's prediction). Points to the right of zero increase the predicted death rate; points to the left decrease it.

Subgroup Analyses

The association between positivity rate and log-transformed deaths varied substantially by GDP stratum (Table 6). The correlation was strongest in high-GDP countries ($r = 0.898$, $p < 0.001$), followed by low-GDP countries ($r = 0.597$, $p = 0.019$) and medium-GDP countries ($r = 0.525$, $p = 0.119$). Fisher's z-test indicated that the correlation in high-GDP countries was significantly larger than in medium-GDP countries ($z = 2.14$, $p = 0.032$) but not significantly different from low-GDP countries ($z = 1.72$, $p = 0.085$).

Table 6: Subgroup Analysis: Positivity-Deaths Correlation by GDP Stratum

GDP Category	n	Correlation (r)	p-value
Low	12	0.597	0.019
Medium	12	0.525	0.119
High	12	0.898	<0.001

Note: Correlations are between positivity rate (%) and $\log_{10}$ (Deaths per 1M + 1).

MAJOR FINDINGS

The principal findings from this analysis can be summarized as follows. Positivity rate showed the largest regression coefficient ($\beta = 10.51$, $p < 0.001$), indicating a large marginal effect when considered in isolation, yet ranked fourth in SHAP importance (14.2%), suggesting its unique contribution diminishes after accounting for other factors. Conversely, GDP per capita ranked first in SHAP importance (41.7%), driving reporting completeness, with a moderate regression coefficient ($\beta = 1.39$, $p < 0.001$). The parsimonious model explained 89.2% of variance in reported death rates ($R^2 = 0.892$, adjusted $R^2 = 0.878$, $F(4,31) = 64.19$, $p < 0.001$). The

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

positivity-deaths correlation was strongest in high-GDP countries ($r = 0.898$, $p < 0.001$), reflecting more complete death reporting. Vaccination coverage showed a positive coefficient ($\beta = 0.74$, $p = 0.012$), which reflects reverse causality rather than a protective effect, as countries with higher death tolls implemented more aggressive vaccination campaigns. Median age was not statistically significant in the parsimonious model ($\beta = 0.02$, $p = 0.132$), with the bootstrap confidence interval including zero ([-0.016, 0.061]), indicating that age effects are mediated through other variables such as GDP and healthcare capacity. Cross-validated performance was positive for the linear model but negative for eXtreme Gradient Boosting (XGBoost) ($R^2 = -0.324$), suggesting that linear models with careful variable selection are more appropriate for this sample size.

DISCUSSION

This stratified analysis of 36 countries suggests three principal findings. First, positivity rate appeared as the strongest predictor of cumulative COVID-19 mortality, with a large effect size ($\beta = 10.51$) that remained consistent across model specifications. Second, the parsimonious model including only positivity rate, vaccination coverage, GDP per capita, and median age accounted for 89% of cross-country variation in death rates. Third, the association between positivity rate and mortality was most pronounced in high-GDP countries, a pattern that likely reflects substantial differences in data reliability and death reporting completeness across income strata.

Data Reliability and Under-Reporting in Low-GDP Countries

A central challenge in cross-country COVID-19 comparisons concerns the marked heterogeneity in testing capacity and death registration across income levels. In low-GDP countries, testing intensity was considerably lower (mean 1,600 tests per 100k) compared to high-GDP countries (mean 63,000 tests per 100k). This disparity likely resulted in substantial under-ascertainment of both cases and deaths in low-resource settings. Consequently, reported deaths in low-GDP countries (mean 167 per 1M) may substantially underestimate true mortality, a concern well documented in the literature (Msemburi et al., 2023). The weaker positivity-deaths correlation observed in low-GDP countries ($r = 0.597$) compared to high-GDP countries ($r = 0.898$) may therefore reflect differential reporting completeness rather than genuinely different biological or epidemiological relationships. This suggests that the true association between community transmission and mortality could be considerably stronger than estimates herein indicate, particularly in settings where testing and vital registration systems are least developed.

Positivity Rate as a Key Metric

The observation that positivity rate outperformed testing volume as a predictor of mortality may carry important public health implications, particularly given concerns about data reliability. Although testing intensity showed a strong positive correlation with deaths ($r = 0.688$), this likely reflects reverse causality: countries experiencing high mortality expanded testing capacity reactively. More importantly, testing intensity is highly sensitive to reporting completeness, whereas positivity rate—the proportion of tests returning positive—may be less influenced by absolute testing volume once a minimum threshold is reached. The World Health Organization has recommended that positivity rates remain below 5% for 14 days before reopening, a benchmark that high-GDP countries in our sample achieved (mean = 3.0%) but low-GDP countries did not (mean = 10.2%). Notably, even this benchmark may be difficult to interpret in settings with extremely low testing capacity, where high positivity rates could reflect both intense transmission and inadequate surveillance.

The magnitude of the positivity effect appears substantial. Although the arcsine transformation complicates direct interpretation, the coefficient ($\beta = 10.51$) suggests that a 10-percentage point increase in positivity rate may be associated with approximately a 1.8-fold increase in predicted deaths per million, holding other variables constant. However, this estimate likely depends on the assumption that death reporting is equally complete across countries—an assumption that probably does not hold.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

The Role of GDP and Testing

The positive association between GDP per capita and deaths ($\beta = 1.39$, $p < 0.001$) seems counterintuitive given that wealthier countries had better healthcare infrastructure. This finding likely reflects three interrelated phenomena, two of which concern data reliability. First, wealthier countries almost certainly had more complete death reporting, reducing undercount bias. Second, high-GDP countries had older populations (median age 42.4 vs. 22.5 years), increasing biological susceptibility. Third, high-GDP countries experienced earlier and larger epidemic waves before vaccines became available. The strong association between GDP and reported deaths may therefore be partially nonfactual, driven by systematic differences in the completeness of vital registration systems rather than true differences in mortality risk.

The elevated variance inflation factors observed for GDP, testing, and age suggest that these variables share substantial common variance, making it difficult to disentangle true biological effects from reporting artifacts. The parsimonious model reduced but did not eliminate this issue (VIF range: 2.92-8.99).

Vaccination and Reverse Causality

The positive association between vaccination coverage and deaths ($\beta = 0.74$, $p = 0.012$) requires careful interpretation and may also be influenced by reporting heterogeneity. This finding likely does not indicate that vaccination increased mortality—a conclusion contradicted by clinical trial evidence (Polack et al., 2020; Baden et al., 2021). Instead, it may reflect reverse causality: countries with higher death tolls implemented more aggressive vaccination campaigns. In our sample, high-GDP countries had both higher vaccination coverage (81.9%) and higher reported deaths (1,186 per million) compared to low-GDP countries (35.4% vaccination, 167 reported deaths). However, it remains possible that under-reporting of deaths in low-GDP countries also contributes to this pattern, as lower reported deaths may have reduced perceived urgency for vaccination campaigns in some settings.

Subgroup Differences by GDP and Reporting Implications

The observation that the positivity-deaths correlation was strongest in high-GDP countries ($r = 0.898$) compared to low-GDP countries ($r = 0.597$) may primarily reflect differential death reporting quality rather than true biological differences. In low-income settings, both testing and death registration are incomplete (Msemburi et al., 2023), potentially attenuating observed correlations. This suggests that the true association between community transmission and mortality could be even stronger than our estimates indicate, particularly in low-resource settings where under-reporting is most severe. Conversely, the very strong correlation observed in high-GDP countries ($r = 0.898$) may approach the true underlying relationship, as these settings have more complete testing and vital registration systems.

Comparison with Previous Literature

Our findings generally align with and extend previous research, though direct comparisons are complicated by heterogeneous data quality across studies. Thorp et al. (2023) reported that GDP ($r = 0.50$) and vaccination ($r = 0.39$) correlated with deaths across 108 countries, though their analysis did not include positivity rate as a separate predictor. Klement and Walach (2022) found that testing and prior influenza vaccination were the strongest predictors in 43 European countries, explaining approximately 66% of variance. Our parsimonious model achieved higher explanatory power (adjusted $R^2 = 0.878$), possibly due to the inclusion of positivity rate as a direct measure of transmission intensity and the stratified design that reduced between-country heterogeneity. However, the higher explanatory power may also reflect the narrower range of data quality in our stratified sample, which included sufficient representation of high-GDP countries with more reliable reporting.

The C-MOR project (Rahmanian Haghighi et al., 2024) emphasized the importance of vaccination and government stringency, but their analysis focused on excess mortality rather than reported deaths. Excess mortality estimates attempt to circumvent under-reporting by modeling expected deaths from historical trends, representing a potentially more robust approach for cross-country comparisons. Our reliance on reported deaths

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

limits comparability with such studies but aligns with routinely available surveillance data, which remains the primary information source for real-time pandemic response.

Limitations

Some of the notable limitations include data reliability concerns. First, the ecological design precludes individual-level causal inference. The associations observed at the country level may not hold at the individual level, and unmeasured confounding cannot be ruled out. Second, death reporting varied substantially across countries. In low-income settings, under-counting is well documented (Msemburi et al., 2023). The stronger positivity-deaths correlation in high-GDP countries strongly suggests that reporting bias may substantially attenuate observed effects in low-resource settings. Thus our estimates likely underestimate the true association between transmission intensity and mortality. Third, testing capacity itself varied dramatically across income strata, with low-GDP countries achieving mean testing rates of only 1,600 tests per 100,000 compared to 63,000 tests per 100,000 in high-GDP countries (Knipper et al., 2025). This disparity likely resulted in systematic under-ascertainment of cases, which may have biased positivity rate estimates, particularly in settings where testing was primarily symptomatic. Fourth, the positive association between vaccination coverage and mortality likely reflects reverse causality rather than a causal effect. Countries with higher death tolls vaccinated more aggressively. This ecological fallacy represents a known limitation of cross-country comparisons and should not be misinterpreted as evidence against vaccine effectiveness. Fifth, the sample size (n = 36) limited statistical power for detecting small effects and precluded testing of three-way interactions. The parsimonious model achieved an acceptable subject-to-predictor ratio (9:1), though replication in larger samples would strengthen confidence in these findings. Sixth, the cross-sectional design captures associations at a single time point (December 2022) but does not account for temporal dynamics. The timing of epidemic waves, variant emergence, and vaccination rollout varied substantially across countries, and our static analysis cannot disentangle these temporal effects. Seventh, the positivity rate measure uses peak 7-day average rather than time-averaged values. While this captures maximum transmission intensity, it may not fully reflect cumulative exposure over the pandemic period, particularly in countries with multiple waves.

CONCLUSIONS

In this stratified analysis of 36 countries representing diverse economic and demographic profiles, positivity rate emerged as the strongest predictor of COVID-19 mortality among the variables examined. However, this finding must be interpreted in light of substantial heterogeneity in data quality across income strata. A parsimonious model including positivity rate, vaccination coverage, GDP per capita, and median age accounted for 89% of cross-country variation in reported death rates, though this high explanatory power may partly reflect systematic reporting differences between high- and low-GDP countries.

The association between positivity and mortality appeared most pronounced in high-GDP countries ($r = 0.898$), where testing and death registration are most complete. In contrast, the weaker correlation observed in low-GDP countries ($r = 0.597$) likely reflects substantial under-reporting of deaths rather than a genuinely different biological relationship. This suggests that the true association between community transmission and mortality could be considerably stronger than our estimates indicate, particularly in low-resource settings where surveillance systems are least developed.

These findings carry practical public health implications, though they must be applied with caution across different settings. Maintaining low positivity rates through accessible, widespread testing could serve as a cornerstone of pandemic surveillance and response, particularly where vital registration systems are reliable. The World Health Organization's benchmark of <5% positivity for 14 days before reopening provides an evidence-based target, though achieving this benchmark may be challenging in low-resource settings with limited testing capacity. Routine monitoring of positivity rates might offer early warning of resurgence before hospitalizations and deaths increase, enabling timely implementation of targeted interventions.

Future research should examine whether the positivity-mortality association holds at subnational levels, where data quality may be more homogeneous, and whether similar patterns emerge for other respiratory pathogens.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Prospective studies with standardized reporting protocols and investment in vital registration infrastructure would strengthen causal inference. Additionally, research on the barriers to achieving low positivity rates in low-resource settings remains urgently needed to ensure equitable pandemic preparedness. Methodological work to develop correction factors for under-reporting based on testing intensity and other indicators would also substantially improve cross-country comparisons.

ACKNOWLEDGMENTS

The author thanks the Our World in Data team, World Bank, and United Nations Population Division for making their data publicly available.

Conflict of Interest Disclosure

The author declares no competing interests.

Funding

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

The dataset and Python code supporting this analysis are available from the author on request.

REFERENCES

1. World Health Organization. (2020). Public health criteria to adjust public health and social measures in the context of COVID-19. WHO. https://www.who.int/publications/i/item/WHO-2019-nCoV-Adjusting_PH_measures-Criteria-2020.1
2. Msemburi, W., Karlinsky, A., Knutson, V., Aleshin-Guendel, S., Chatterji, S., & Wakefield, J. (2023). The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature*, 613(7942), 130-137. https://doi.org/10.1038/s41586-022-05522-2
3. Klement, R. J., & Walach, H. (2022). Testing and vaccination to reduce COVID-19 mortality in European countries. *Journal of Clinical Medicine*, 11(15), 4384. https://doi.org/10.3390/jcm11154384
4. Abbasi, A. F., Karimi Dehkordi, N., SoleimanvandiAzar, N., Roohravan Benis, M., & Nojomi, M. (2025). Gini coefficient, GDP per capita and COVID-19 mortality: a systematic review of ecologic studies. *BMC Public Health*, 25(1), 22921. https://doi.org/10.1186/s12889-025-22921-4
5. Kanokudom, S., Piamsa-Nga, N., Ratanapanich, K., et al. (2025). Impact of economic factor, percent vaccination, healthcare quality, and population density on coronavirus disease 2019 (COVID-19) mortality rates: a global analysis in 2023. *Cureus*, 17(3), e80582. https://doi.org/10.7759/cureus.80582
6. Wang, L., Ma, H., Yiu, K. C. Y., Calzavara, A., Landsman, D., Luong, L., Chan, A. K., Kustra, R., Kwong, J. C., Boily, M. C., Hwang, S., Straus, S., Baral, S. D., & Mishra, S. (2020). Heterogeneity in testing, diagnosis and outcome in SARS-CoV-2 infection across outbreak settings in the Greater Toronto Area, Canada: an observational study. *CMAJ Open*, 8(4), E627-E636. https://doi.org/10.9778/cmajo.20200097
7. Stocki, S. A., Alvarenga, M. B. B., Pinheiro, D. S., Salles, M. V., Machado, G. C., Oliveira Filho, N. C., Bif, N. C. S., & Freitas, J. B. C. (2026). Impact of COVID-19 vaccination on SARI lethality among older adults in Brazil: an ecological study (2020-2023). *The Brazilian Journal of Infectious Diseases*. https://doi.org/10.1016/j.bjid.2026.103456
8. Aguilar, S., Bastos, L. S. L., Maçaira, P., Baião, F., Simões, P., Cerbino-Neto, J., Ranzani, O., Hamacher, S., & Bozza, F. A. (2024). Impact of the first year of COVID-19 vaccination strategy in Brazil: an ecological study. *BMJ Open*, 14(7), e072314. https://doi.org/10.1136/bmjopen-2023-072314
9. Buttia, C., Llanaj, E., Raeisi-Dehkordi, H., Kastrati, L., Amiri, M., Meçani, R., Taneri, P. E., Ochoa, S. A. G., Raguindin, P. F., Wehrli, F., Khatami, F., Espínola, O. P., Rojas, L. Z., de Mortanges, A. P., Macharia-Nimietz, E. F., Alijla, F., Minder, B., Leichtle, A. B., Lüthi, N., Ehrhard, S., Que, Y. A., Fernandes, L. K.,

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Hautz, W., & Muka, T. (2023). Prognostic models in COVID-19 infection that predict severity: a systematic review. *European Journal of Epidemiology*, 38(4), 355-372. https://doi.org/10.1007/s10654-023-00973-1

10. Appel, K. S., Geisler, R., Maier, D., Miljukov, O., Hopff, S. M., & Vehreschild, J. J. (2024). A systematic review of predictor composition, outcomes, risk of bias, and validation of COVID-19 prognostic scores. *Clinical Infectious Diseases*, 78(4), 889-899. https://doi.org/10.1093/cid/ciad618
11. Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159. https://doi.org/10.1037/0033-2909.112.1.155
12. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson Prentice Hall.
13. Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J. L., Pérez Marc, G., Moreira, E. D., Zerbini, C., Bailey, R., Swanson, K. A., Roychoudhury, S., Koury, K., Li, P., Kalina, W. V., Cooper, D., Frenck, R. W., Hammitt, L. L., & 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
14. Baden, L. R., El Sahly, H. M., Essink, B., Kotloff, K., Frey, S., Novak, R., Diemert, D., Spector, S. A., Rouphael, N., Creech, C. B., McGettigan, J., Khetan, S., Segall, N., Solis, J., Brosz, A., Fierro, C., Schwartz, H., Neuzil, K., Corey, L., & Zaks, T. (2021). Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *New England Journal of Medicine*, 384(5), 403-416. https://doi.org/10.1056/NEJMoa2035389
15. Rahmanian Haghighi, M. R., Ghasemi, A., & Mokhtari, M. (2024). Factors associated with COVID-19 excess mortality: A cross-country analysis. *Journal of Global Health*, 14, 05012. https://doi.org/10.7189/jogh.14.05012
16. Thorp, H., Miller, B., & Johnson, C. (2023). Global correlates of COVID-19 mortality: A multi-country ecological study. *BMJ Global Health*, 8(3), e011245. https://doi.org/10.1136/bmjgh-2022-011245
17. Knipper, M., Moreira-Soto, A., Beuchel, C., Tabares, X., Wulf, B., Gade, N., Fischer, C., Aigner, A., & Drexler, J. F. (2025). Socioeconomic determinants potentially underlying differential global SARS-CoV-2 testing capacity: an ecological study. *BMJ Open*, 15(3), e090804. https://doi.org/10.1136/bmjopen-2024-090804

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

APPENDIX

Supplementary Materials

Supplementary Table 1. Country classification by GDP per capita and population density strata

Country	GDP category	Density category
Niger	Low	Low
Chad	Low	Low
Mali	Low	Low
Burkina Faso	Low	Low
Nigeria	Low	Medium
Cameroon	Low	Medium
India	Low	Medium
Pakistan	Low	Medium
Bangladesh	Low	High
Philippines	Low	High
Rwanda	Low	High
Haiti	Low	High
Brazil	Medium	Low
Mexico	Medium	Low
Russia	Medium	Low
Argentina	Medium	Low
China	Medium	Medium
Turkey	Medium	Medium
Thailand	Medium	Medium
Colombia	Medium	Medium
Vietnam	Medium	High
Indonesia	Medium	High
Egypt	Medium	High
Morocco	Medium	High
Australia	High	Low
Canada	High	Low
Norway	High	Low
Finland	High	Low
France	High	Medium
Spain	High	Medium
Poland	High	Medium
United Kingdom	High	Medium
Netherlands	High	High
South Korea	High	High
Japan	High	High
Singapore	High	High

Note: GDP categories: Low (<$4,000), Medium ($4,000-$25,000), High (>$25,000). Density categories: Low (<50 persons/km²), Medium (50-200 persons/km²), High (>200 persons/km²). Complete data are available from the corresponding author upon request.