Association of Economic, Demographic, and Policy Factors on COVID-19 Deaths Per Capita and per Case.

Boris Khesin, Beckett Pilling, Jack Pada

2025-12-19

Abstract

Using early 2020 datasets compiled by Jiangsu University, the CIA, and the World Bank, we examined which economic, demographic, and policy factors were most associated with COVID-19 deaths per capita (DPC) and COVID-19 cases per death (COD) across countries. Predictors included the Human Development Index (HDI), Gross Domestic Product (GDP), Stringency Index, population density, and mean age. Nested linear regression models and F-tests identified HDI as the strongest predictor of deaths per capita: neither adjusting for GDP nor mean age improved the model (Stringency Index was found to be insignificant). The final logistic regression model, which only included HDI, suggested that a 0.1 increase in HDI corresponded to an increase of approximately 62.7% in DPC, which indicates that countries with a higher HDI experience slightly higher deaths per capita. On the other hand, population density was identified as the best predictor for deaths per case; all others had little to no significance. The final model for deaths per capita suggests that countries with a higher HDI potentially report COVID-19 with higher efficacy, although further analysis would be necessary to confirm. The final model for deaths per case found that a higher population density is significantly associated with deaths per case, with a 1% increase in population density corresponding to about a 0.16% decrease in deaths per case, on average. Overall, the HDI and population density were the most informative predictors, from those we analyzed, of early 2020 COVID-19 mortality, highlighting the importance of development-aware and demographic-aware models in pandemic crisis analysis.

Background

The COVID-19 pandemic dramatically affected all countries, be it economically, socially, or by the death toll. As many of the economic and social effects pertained to minimizing the deaths per capita of a country, insofar as shutdowns were implemented, it becomes pertinent to analyze which factors may contribute to the efficacy of a country in preventing deaths from COVID-19. For instance, does wealth per citizen predict lower death rates? Or did implementing higher shutdowns prove to be effective? These types of questions led us to the research question: What population variables are most strongly associated with a country's deaths per person? To answer this, our study applies statistical methods to examine various socio-economic indicators and policy decisions, aiming to identify how they associate with deaths per person caused by COVID-19 in early 2020.

Methods

The datasets sourced from Jiangsu University, the CIA World Factbook, and the World Bank provide country-level measures of COVID-19 outcomes (cases and deaths) alongside potential explanatory variables including Human Development Index (HDI), GDP per capita, mean age, population, population density, and a Stringency Index. We combined these sources into a single analytic dataset by aligning countries across datasets and retaining variables needed for modeling.

We studied two outcome measures designed to capture different aspects of pandemic burden. Deaths per capita (DPC) was defined as total COVID-19 deaths divided by population, and deaths per case (DPCase) was defined as total deaths divided by total cases. After excluding countries with missing values among the variables used in each analysis, the main analytic dataset contained 150 countries. Because the Stringency Index was available for far fewer countries, analyses involving stringency were limited to 33 countries, which is important to keep in mind when interpreting null results for that predictor.

Our primary analytic approach was linear regression. We first fit univariable linear regression models relating each predictor to each outcome to understand bivariate patterns and identify candidate predictors. We then fit multivariable models when multiple predictors appeared associated with the outcome. For coefficient-level inference we used standard regression tests with a significance threshold of $p < 0.05$, and for model-level comparisons we used nested-model F-tests to assess whether adding predictors meaningfully improved fit relative to a simpler model. Because both outcomes are ratios and exhibited skewness and non-constant variance in preliminary diagnostics, we evaluated standard linear model assumptions using residual-based checks (e.g., residual distribution and residuals vs fitted). When needed, we applied log-transformations to better satisfy modeling assumptions. In the final models, log(deaths per capita) was used for the HDI model, and both log(deaths per case) and log(population density) were used for the density model.

Results

We constructed linear regression models for each independent variable against both outcome variables—COVID-19 deaths per capita and deaths per case—and tested the statistical significance of each coefficient. For deaths per capita, the coefficients for GDP per capita ($p = 0.003883$), Human Development Index (HDI; $p < 0.0001$), and mean age ($p < 0.0001$) were statistically significant. The outcome variable (deaths per capita) was log-transformed to satisfy linear model assumptions. In contrast, the Stringency Index (p = 0.7630) and population density (p = 0.3277) were not statistically significant predictors. *(Figure 1)*

An F-test comparing the full model (GDP per capita + HDI + mean age) to a nested model containing only HDI indicated that GDP per capita and mean age did not significantly improve model fit (F = 0.1984, p = 0.8203). *(Figure 2, Figure 3, Figure 4)* These results support the conclusion that HDI alone sufficiently captures the explanatory power of the multivariable model.

The reduced model including only HDI estimates a coefficient of 6.27, implying that a 0.1 increase in HDI is associated with an approximately 87% increase in COVID-19 deaths per capita. The intercept (-14.485) lies outside the range of observed HDI values and therefore has no practical interpretation. *(Figure 5, Figure 6)*

For the outcome deaths per case, population density was the only statistically significant predictor (p = 0.00117). The Stringency Index (p = 0.387), GDP per capita (p = 0.173), mean age (p = 0.9667), and HDI (p = 0.421) were not statistically significant. Both deaths per case and population density were log-transformed to meet linear model assumptions. The estimated coefficient of population density is -0.15572, suggesting that a 1% increase in population density is associated with a 0.15572% decrease in COVID-19 deaths per case. Because no country can reasonably have a population density of zero, the intercept (-3.37239) has no meaningful interpretation. *(Figure 7, Figure 8, Figure 9)*

Discussion

This analysis found two main cross-national relationships: (1) HDI is positively associated with deaths per capita, and (2) population density is negatively associated with deaths per case. These findings should be interpreted as associations rather than causal effects. Country-level COVID outcomes are shaped by many processes simultaneously—true disease burden, timing of outbreak waves, healthcare capacity, and measurement/reporting practices—and the regressions here cannot fully separate these mechanisms.

One plausible interpretation of the HDI result is that HDI is acting less as a biological driver of mortality and more as a proxy for institutional capacity and measurement quality. Higher-HDI countries often have more complete vital registration systems, more consistent cause-of-death attribution, and greater testing and surveillance. If higher-HDI settings capture deaths more completely, reported deaths per capita would rise even if underlying mortality risk were similar. Meanwhile, if case detection also improves with development, the denominator of deaths-per-case increases (more detected infections), which can reduce deaths per case without implying that outcomes are biologically better. This general logic is consistent with the pattern that HDI was strongly associated with deaths per capita but not clearly associated with deaths per case. The population density finding can also be interpreted through an institutional/measurement lens. More densely populated countries are often more urbanized and may have more centralized healthcare access and higher testing availability, which can lead to more confirmed infections. An increase in detected cases can lower the deaths-per-case ratio mechanically, even if the absolute number of deaths does not fall. Thus, population density itself is unlikely to be “protective” in a biological sense; instead, it may proxy for features correlated with density, such as testing intensity, healthcare infrastructure, or how outbreaks were recorded and managed in different contexts. At the same time, it remains possible that density relates to other unmeasured factors (e.g., age structure, household composition, or timing of spread) that could influence deaths per case.

Several limitations constrain interpretation. First, missingness reduced the dataset from the original coverage to 150 countries, which may introduce selection bias if countries excluded due to missing data differ systematically in reporting or infrastructure. Second, mean age estimates were not matched exactly to the pandemic year, creating potential measurement error that could weaken or distort estimated age effects. Third, early COVID case and death data vary widely in reliability across countries due to differences in testing, diagnostic criteria, and death attribution. Fourth, the Stringency Index analysis had low power (n = 33), so a non-significant result should not be read as evidence of no policy effect. Finally, the analysis is cross-sectional and observational, so unmeasured confounding (e.g., healthcare capacity, comorbidity prevalence, epidemic timing, testing rates) may drive the observed associations. Future work would benefit from incorporating longitudinal outcomes, explicit measures of testing intensity and health system capacity, and model designs better suited to causal questions.

Figures

Figure 1: HDI vs Log(Deaths per Capita)

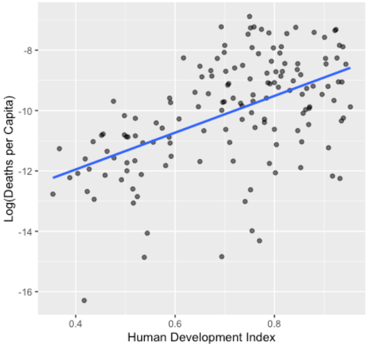


Figure 2: Residuals vs Fitted (Full Model)

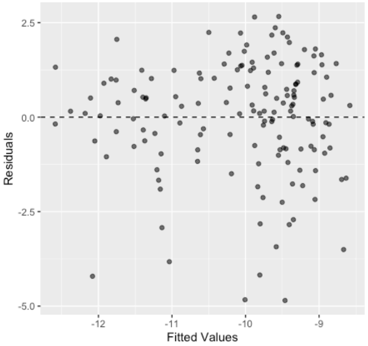


Figure 3: Q–Q Plot (Full Model)

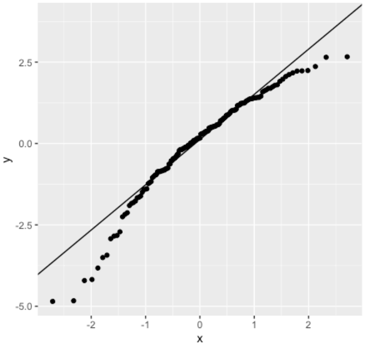


Figure 4: ANOVA Model Comparison

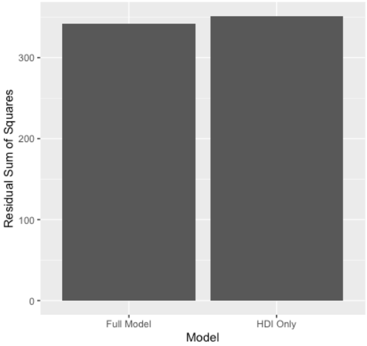


Figure 5: Residuals vs Fitted (HDI-Only Model)

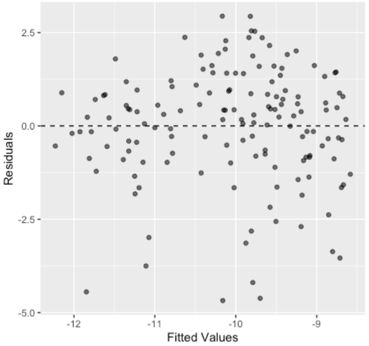


Figure 6: Q–Q Plot (HDI-Only Model)

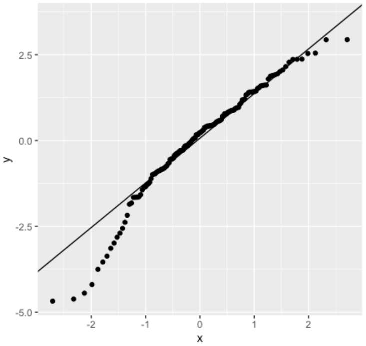


Figure 7: Log(P. Density) vs Log(Deaths per Case)

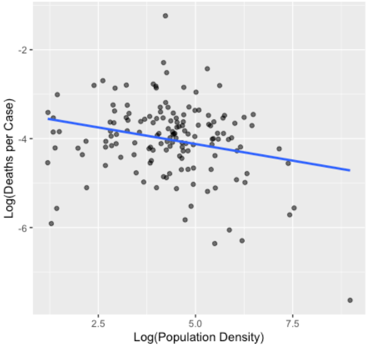


Figure 8: Residuals vs Fitted (Deaths per Case Mod

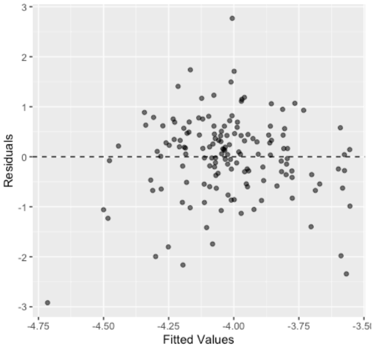
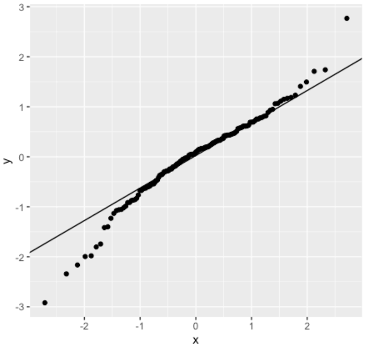


Figure 9: Q–Q Plot (Deaths per Case Model)



References

- *Vitenu-Sackey, Prince Asare (2020), "The Impact of Covid-19 Pandemic on the Global Economy: Emphasis on Poverty Alleviation and Economic Growth", Mendeley Data, V1, doi: 10.17632/b2wvnbnpj9.1*

- *Central Intelligence Agency. (2025). Median age — Country comparison. In The World Factbook.*
- *The World Bank. (2025). Population density (people per sq. km of land area) (Indicator EN.POP.DNST). World Bank Open Data.*