Severity of the COVID-19 Pandemic: Socioeconomic, Demographic and Health Risk Factors

Ciarán Dervan

Dublin City University

Abstract

This article details several risk factors necessary for the accurate prediction of the severity of the COVID-19 pandemic globally. The main objective is to evaluate the possible effects of socioeconomic, demographic and pre-existing health conditions on COVID-19 outcomes. Utilising data from several countries, it is possible to compare the impact of COVID-19 on Ireland's population with that of other nations with different circumstances to our own.

I. INTRODUCTION

The coronavirus pandemic has greatly strained the Irish health care system this past year. Despite being among the strictest countries in Europe regarding lockdowns, the Irish Government's enforcement of stringent rules has not managed to flatten the curve of the virus, and the nation has already suffered 4,803 deaths due to the pandemic. Given the inefficacy of the continuous lockdown measures imposed on the Irish people, this paper sought to investigate whether there are underlying risk factors at play in the transmission, and fatality, of COVID-19. We compared data collected in Ireland regarding the pandemic with data from 9 other countries, all of which have experienced different degrees of severity regarding the pandemic and their response to the virus.

Previous research has identified that older populations are at greater risk of dying from the COVID-19 virus (Natale et al., 2020). Ireland is an aging population and ranks older than the other countries in this study on every metric (Median age: 38.7, life expectancy: 82.3, with 13.93% of the population at 65 years or older). As such, it was essential to test whether age was a predictor for COVID-19 fatality. Along with age, pre-existing health risk factors' influence on the fatality of COVID-19 were analyzed. This included the cardiovascular death rate, the prevalence of diabetes in the population, and the proportion of female and male smokers in a country. People with these conditions/habits have been identified as being at a higher risk of a more severe response should they contract COVID-19 (e.g., HSE, 2020). Lastly, population density is another factor that has been proposed to have contributed to the transmission and subsequent fatality of the virus (e.g., Hamidi et al., 2020). Yet, this assumption is still contentious, and some historical studies have demonstrated inverse relations between population density and mortality rates for previous outbreaks (e.g., Mills et al., 2004). Population density, however, as an individual variable does not yield as much information at the country level, as opposed to the regional/county level. As such, further predictors should be taken into consideration to form a socioeconomic predictor model. For example, Saint Kitts and Nevis is a small island in the Caribbean and has a population of just 53,192 people but resultantly had the second largest density at 212.86 per km^2. Conversely, Saudi Arabia had the biggest population with 34,813,867 residents, yet a density of only 15.32 per km^2. Based off density alone, one could assume that a virus would spread much quicker in Saint Kitts and Nevis than in Saudi Arabia, however, when population is accounted for, it is easy to see how an island nation with only 53,192 residents would be in a better position to enforce restrictions and ensure protocols are followed.

Our hypotheses centre around three distinct models focusing on 1) age and CFR; 2) health risk factors and covid deaths; 3) socioeconomic factors and case numbers. Regarding age demographics, in order to analyze if Ireland fares worse in the pandemic due to our aging population, we hypothesized that populations with a higher percentage of elderly residents (65 years old or over) would exhibit significantly higher Case Fatality Ratios (i.e., CFR: the ratio of COVID related deaths per COVID cases) than nations with younger populations. Regarding health risk factors, we hypothesized that these four factors which are indicative of the health of a population will each be statistically significant predictors for COVID-19 related deaths. Regarding socioeconomic factors we hypothesized that GDP per-capita would have a main effect on case numbers, and that population density would have a main effect on case numbers but moderated by the country's total population. Thus, wealthier, denser and more populated countries would experience greater numbers of COVID cases.

II. Methods

Data from 10 countries: Ireland, Armenia, The Bahamas, Saint Kitts and Nevis, South Sudan, Saudi Arabia, The Congo, Guinea-Bissau, Liberia and Paraguay) was collected from the website www.ourworldindata.org which spanned from the 29tt of February 2020, to the 10th of February 2021. For descriptive statistics on all variables see Appendix A.

Model 1: Age and CFR

To test the effect of age on mortality, we first created a computed variable that generated the CFR for all the countries using the formula below:

$$\text{Case Fatality ratio } (CFR, in\%) = \frac{\text{Number of deaths from disease}}{\text{Number of confirmed cases of disease}} \times 1$$

The variable denoting the percentage of the population aged 65 or older (coded as *"65 ≥ y/o")* was chosen as the best predictor for the model as the related variables of life expectancy and median age did not give as accurate a description of how elderly the country's population is exactly at the time of testing. We conducted multiple linear regression to test the following hypotheses:

- H_a: The **CFR** increases when **65 ≥ y/o** increases.
- H_0: The **CFR** stays the same when **65 ≥ y/o** increases.

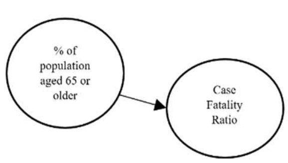

Figure 1. Model 1 representation.

This model is represented in *Fig. 1,* which shows the predicted main effect of the predictor on the outcome. The model can be written as the following equation:

$$Y = \beta_0 + \beta_1.X + \epsilon$$

Model 2: health risk factors and deaths

A linear regression was performed on the following variables: the percentage of male and female smokers of a population, the diabetes prevalence in the population, and the cardiovascular death rate. We hypothesized that as all these factors increase the total deaths of the country would increase in response. The hypotheses are thus formulated as follows:

- H_a^1: The **deaths** increase when **male smokers** increase.
- H_0^1: The **deaths** stay the same when **male smokers** increase.
- H_a^2: The **deaths** increase when **female smokers** increase.
- H_0^2: The **deaths** stay the same when **female smokers** increase.
- H_a^3: The **deaths** increase when **diabetes prevalence** increases.
- H_0^3: The **deaths** stay the same when **diabetes prevalence** increases.
- H_a^4: The **deaths** increase when **cardiovascular death rate** increases.
- H_0^4: The **deaths** stay the same when **cardiovascular death rate** increases.

This model is represented in *fig. 2,* and follows the equation below:

$$Y = \beta_0 + \beta_1.X + \beta_2.Z + \beta_3.W + \beta_4.V + \epsilon$$

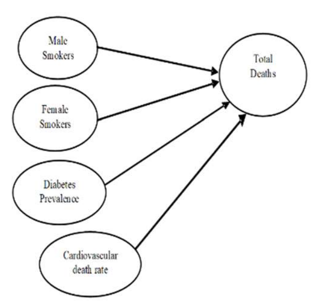

Figure 2. Model 2 representation.

To ensure the model maintained its goodness of fit the collinearity of all variables was measured as health conditions and smoking often exhibit high comorbidity, thus there was a greater risk of the predictors being too highly correlated (see *Results* for further details).

Model 3: Socioeconomic Factors and number of Cases

The effect of GDP, population density and population on the total number of cases of COVID in each country were investigated. Additionally, the interaction effect of population on density was also tested for. Multiple linear regression was conducted on the variables to test the following hypotheses:

- H_a^1: The **cases** increase when **GDP** increases.
- H_0^1: The **cases** stay the same when **GDP** increases.
- H_a^2: The **cases** increase when **population density** increases.
- H_0^2: The **cases** stay the same when **population density** increases.
- H_a^3: The **cases** increase when **population** increases.
- H_0^3: The **cases** stay the same when **population** increases.
- H_a^4: The **cases** increase when **population density * population** increases.
- H_0^4: The **cases** stay the same when **population density * population** increases.

We hypothesized that these factors would contribute to an increase in the number of cases, but that population would have a moderating effect on density, meaning that the estimate result for the interaction of **population density * population** would be far less than the estimate of the **population density predictor** alone. The model has been represented in *Fig. 3,* and follows this equation:

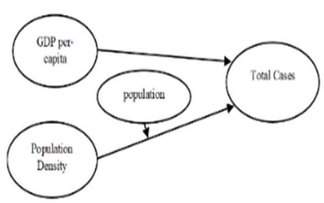

Figure 3. Model 3 representation

$$Y = \beta_0 + \beta_1 . X + \beta_2 . Z + \beta_3 . W + \beta_4 . W*Z + \epsilon$$

III. RESULTS

Model 1 exhibited statistically significant predictive capability ($R^2 = 0.002$, F(1,2831) = 61.0, P = <.001.), thus the model is better than one without predictors. Regarding the effect of the percentage of the population aged 65 and over on the CFR, H_0 can thus be rejected (β = 0.08, 95\% CI [0.06,0.1], t(30004) = 7.81, p= <.001). See Fig 4. For the Q-Q plot of residuals.

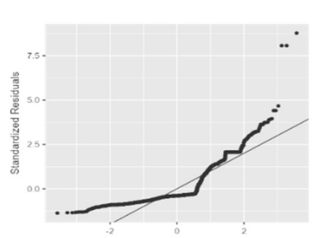

Figure 4. Q-Q plot for Model 1

Model 2 was statistically significant, meaning its 4 predictors were better at explaining the variance in the sample than a model with no health-risk factor predictors ($R^2 = 0.266$, F(4,2375) = 215, P = <.001.) However, on closer examination this model failed to meet the collinearity assumption as the cardiovascular death-rate predictor was too highly correlated with the other predictors (VIF= 25.07, see *Appendix B* for the correlation matrix) and so had to be excluded from the model, its inclusion would have also meant that $\mathbf{H_a^1}$ for this model would have been rejected, as the percentage of male smokers exhibited an inverse relationship with COVID deaths when cardiovascular death-rate was in the model (β = -22.1, 95\% CI [-33.24, -11], t(2371) = -3.91, p= <.001.). After removing the cardiovascular death-rate predictor, the model retained its statistical significance, losing only 1% of the variability it could account for ($R^2 = 0.256$, F(3,2375) = 272, P = <.001.), the collinearity returned to an acceptable amount between predictors also (*see Table 1)*. $\mathbf{H_0^1}$ can be rejected with the revised three-level model (β = 7.85, 95\% CI [3.7212.0], t(2371) = 3.73, p= <.001.). $\mathbf{H_0^2}$ can also be rejected as the female smokers predictor was statistically significant (β =82.32, 95\% CI [73.93, 90.7], t(2371) = 19.26, p= <.001.). $\mathbf{H_0^3}$ can also be rejected as diabetes prevalence was statistically significant, and boasts a very high estimated coefficient (β =116.57, 95\% CI [154.49, 178.6], t(3326) = 27.05, p= <.001.). $\mathbf{H_0^4}$ must be accepted due to the exclusion of that predictor.

Finally, Model 3 exhibited a statistically significant effect also ($R^2 = 0.56$, F(4,3331) = 1056, P = <.001.). GDP exhibited a significant main effect on the number of cases (β = 0.49, 95\% CI [0.38, 0.59], t(33326) = 8.93, p= <.001) and so $\mathbf{H_0^1}$ can be rejected. Population density exhibited a significant effect on the number of cases also (β = 158.07, 95\% CI [117.8, 198.33], t(3326) = 7.70, p= <.001) we can resultantly reject $\mathbf{H_0^2}$. $\mathbf{H_0^3}$ can be rejected as population exhibited a small yet significant effect on the total cases (β = 0.005, 95\% CI [0.004, 0.005], t(3326) = 29.53, p= <.001). Finally, population had a moderating effect on population density (β =1.01, 95\% CI [8.04,1.22], t(3326) = 9.46, p= <.001), as this interaction effect was *Fig. 5*).

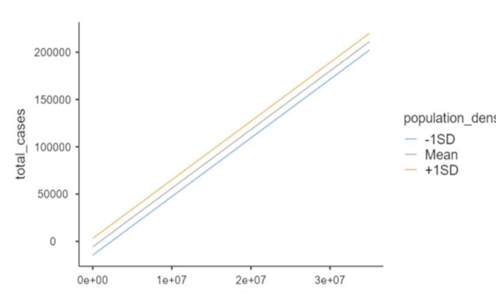

Figure 5. Estimated marginal means for population density*population interaction.

IV. Discussion

While the first model for age and CFR was significant, it had a small effect size, and the model is only capable of explaining around 2% of the variance in total. As evident in *fig. 4,* the regression line does not account for much of the variability at all. It is likely that an assumption of the regression model was not met, presumably normality as CFR is the product of case numbers that rise and fall over time and deaths which continually slowly increase the outcome was likely skewed in its distribution. The model only accounts for 2% of the overall variance in the CFR outcome which is much lower than expected. Ireland has the oldest population out of this sample, yet the model would suggest that that is not what is causing us to be hit so badly by the pandemic. It is likely that this is an issue with the goodness-of-fit of the model and that the actual effect of old age on the fatality of cases is much higher, and so future research should attempt to utilize alternative designs that better capture the reality of the data.

The health risk factors model has a 25.6% probability of accounting for the variance around the mean, again we see in this model that the regression line is not as well fitting as we would like, however, the effect is still significant. Cardiovascular death-rate was excluded from the final model due to its high collinearity skewing the predictive effect, removing this variable returned the collinearity to normal levels (see *Table 1)*. The percentage of male smokers had a smaller effect on the total cases than the percentage of female smokers. This is an interesting observation that warrants further study. The HSE has advised smokers that they are putting themselves at greater risk of fatality from COVID, however, if this risk is 10 times greater for females, it would be worth targeting female smokers with ads about this information in the hope of preventing them from death should they contract the disease. Diabetes prevalence was a very strong predictor of COVID deaths in a country, this is unsurprising as diabetes is heavily linked to obesity, and those who suffer from obesity are at great risk of dying from COVID as their bodies cannot deal with the strain on their respiratory systems. This model informs us that the Irish people need to get fitter and break bad health habits as we have a diabetes prevalence of 3.28, the highest percentage of female smokers from the sample of countries (23%) and an almost equal percentage of male smokers (25.7%). Stopping these habits might help ease the strain on the health care system and ensure we do not get overwhelmed by new, more aggressive strains of the coronavirus.

	VIF	Tol.
Male smokers	1.11	0.903
Female smokers	1.26	0.794
Diabetes prevalence	1.31	0.762

Table 1. Collinearity Stats. For Model 2

Case numbers were found to be effected by GDP, population density and population. The GDP effect is not very surprising when you consider it, higher GDP per-capita generally means that the economy is busier, so there are more face-to-face interactions, there is a higher level of tourism so it is harder to control the border – an issue we have faced in Ireland to the extreme. Density exhibited a huge effect on the total cases, for every additional unit increase in density the cases rise by 158.07 units. However, this effect was proven to be moderated by overall population, when taken into account this interaction effect sees cases rise by only 1.01 units when it rises. The conclusion to be taken here is that it is not just how many people live in proximity to one another that contributes to the spread, but instead, whether or not these people follow distancing and isolating measures.

References

HSE. *People at higher risk from covid-19*. [Accessed 8/04/2021]. Available at: https://www2.hse.ie/conditions/coronavirus/people-at-higher-risk.html#:~:text=have%20severe%20respiratory%20conditions%20including,have%20uncontrolled%20diabetes.

Natale A, Ghio D, Tarchi D, Goujon A, & Conte A. *COVID-19 cases and case fatality rate by age. Knowledge for policy.* 2020. [Accessed 7/04/2021]. Available at: https://ec.europa.eu/knowledge4policy/publication/covid-19-cases-case-fatality-rate-age_en.

Hamidi, S., Sabouri, S., & Ewing, R. (2020). Does Density Aggrevate the COVID-19 Pandemic? *Journal of the American Planning Assoc.,* 86(4), 459-509,

Mills, C. E., Robins, J. M., & Lipsitch, M. (2004). Transmissibility of 1918 pandemic influenza. *Nature,* 432(7019), 904–906.

APPENDICES

Appendix A. *Descriptive statistics for all variables*

	CFR	65 ≥ y/o	deaths	female_ smokers	male_ smokers	diabetes_ prevalence	cardiovascular _deathrate	total_ cases	GDP	population _density	population
N	2833	3014	3337	2379	2379	3337	3014	3336	3337	3337	3337
Missing	504	323	0	958	958	0	323	1	0	0	o
Mean	2.93	6.38	732	5.44	30.8	8.49	270	40731	19926	60.4	7.50e+6
Median	1.90	3.44	62	1.80	25.4	7.20	273	2533	8827	49.1	4937796
SD	2.35	3.93	1386	7.36	13.7	4.86	75.1	87083	21898	57.2	9.79e+6
Min.	0.13	3.00	0	1.5	18.1	2.42	126	1	753	15.3	53192
Max.	23.1	13.9	6415	23	52.3	17.7	382	371356	67335	213	4813867

Appendix B. *Correlation Matrix for Model 2*

		male_ smokers	female_ smokers	diabetes_ prevalence	cardiovascular_ deathrate
male_ smokers	Pearson's r	-			
	p-value	-			
female_ smokers	Pearson's r	**-0.225**	-		
	p-value	**< .001**	-		
diabetes_ prevalence	Pearson's r	**-0.106**	**-0.406**	-	
	p-value	**< .001**	**< .001**	-	
cardiovascular_ deathrate	Pearson's r	**0.708**	**-0.811**	**-0.095**	-
	p-value	**< .001**	**< .001**	**< .001**	-