Impact of Urbanization and Labor Vulnerability Over the COVID-19 Dynamic Across Countries: A Survival Analysis

Augusto Delgado
Waseda University

This study analyzes the evolution of daily recorded cases of COVID-19 among all countries. In more detail, using survival models, we study the average duration of the increase of cases of COVID-19 before reaching its maximum and reduce the number of infected daily. Additionally, we incorporate dummies to control the levels of urbanization and labor vulnerability as the main control variables. Among the main results, we observed that countries with 50% of the population living in urban areas tend to have more extended periods of infection before reaching their maximum levels and entering a plateau. Besides, the vulnerability variable ends up not being significant by itself. Still, in interaction with the urbanization variable, we observe that countries with high urbanization levels and high levels of vulnerable employment reach the maximum infection level earlier than other countries.

Keywords: COVID-19, urbanization, labor vulnerability

INTRODUCTION

The COVID-19 pandemic has undoubtedly affected the economy of all countries in different degrees. These countries have also shown different reactions to the pandemic's evolution, depending on specific characteristics. Regarding the Coronavirus evolution, by December 31, 2019, a new type of flu was reported in Wuhan Province in China. With the sustained increase in this new virus's infections, by January 4, 2020, pneumonia clusters were reported, but no deaths were recorded in Wuhan Province, China. The WHO response to these events established a comprehensive set of guidelines to advise all countries on detecting, testing, and managing possible infection cases. On January 13, the first case of infection outside of China was recorded in Thailand. The next day it was established that contagion is possible from human to human. On March 11, 2020, after observing several infections in Asian countries, the WHO determined that the new coronavirus could be characterized as a pandemic. Since then, governments have been making efforts to test enough of their populations to detect the pandemic's severity within their borders.

Given this new global context, academia has responded with studies related to the analysis of the COVID-19 pandemic. This research aims to complement the existing literature on the subject. In this sense, this research's main objective is to use the survival models to analyze the evolution of the cases of daily COVID-19 infections and find the duration of the increases of everyday infections until reaching the maximum point before descending steadily. Also, the probability of occurrence of this maximum point of daily infections is analyzed. Following the main objective, the null hypothesis is that countries with specific characteristics, such as high levels of urbanization or high levels of vulnerable employment, have different probabilities of the maximum number of daily infections to occur at a certain period.

Many studies on COVID-19 use micro-level survival analysis methods to determine the disease's duration in those infected. For example, Wang et al. (2020) analyzed 538 COVID-19 patients in China's Sichuan province and found that patients living in areas with more significant medical resources spent less time in the hospital. With a similar analysis for India's case, Kundu et al. (2020) study the case of 26 815 patients of COVID-19 using Kaplan-Meier survival estimates, and Cox's model finds that western and central India have lower survival rates compared to other areas of India. For his part, Sousa et al. (2020) analyze the risk factors associated with COVID-19 in the state of northern Brazil for a total of 2070 people, finding that the most vulnerable people at risk of death from COVID-19 are those patients with advanced age and cardiovascular disease present at the time of infection. For Mexico's case, Salinas-Escudero et al. (2020) analyze the instances of COVID-19 for 16 752 confirmed cases of COVID-19 using Kaplan-Meier curves and Cox's proportional hazard model. They find that the risk of death is exceptionally high for men, elderly individuals, those with chronic kidney disease, and those hospitalized in public health services. Finally, Jin et al. (2020) analyzed data for 43 hospitalized, 37 patients who died from COVID-19, and 1019 patients who survived China's disease. Also, data were entered for the 2013 SARS in Beijing for 524 patients and 139 deaths. Among the main findings, men are more likely to die from COVID-19 than women; specifically, men are 2.4 times more likely to die. These results are independent of the age of the patients.

Meilijson and Alon (2020) modify the SIR epidemiological equations model to analyze the cases of COVID-19 by using the random-time transformation RTT of Bassan et al. (1997). They find that the current epidemic situation will remain strong until December, and for Germany and Italy nearly over by September. Using a machine learning approach, Nemati et al. (2020) perform a spread analysis of the COVID-19 using patient discharge time as a variable of interest and survival analysis to predict them. The authors compare the predictive ability of a set of models. Among the results, it is observed that the Gradient Boosting survival model performs better than other models for predicting the survival of COVID-19 patients. Similarly, Friedman et al. (2020) identify 383 prediction models on COVID-19; of these models, only seven met the criteria of analysis and performance to predict COVID-19 cases.

From a different perspective, Zhao and Dilip (2020) collect information on COVID-19 deaths and democracy indices for 2018 from various sources such as Johns Hopkins University for COVID-19 and the Economist Intelligence Unit. Through a survival analysis, they find that less democratic countries have performed better in controlling the virus and the number of deaths caused by the pandemic. These results are essential because they confirm that specific socio-economic characteristics affect the pandemic's development in each of the different economies.

This paper is divided as follows: section 2 for Data and Methodology; section 3 for presenting the main results and the interpretation of them; finally, section 4 for conclusions.

DATA AND METHODOLOGY

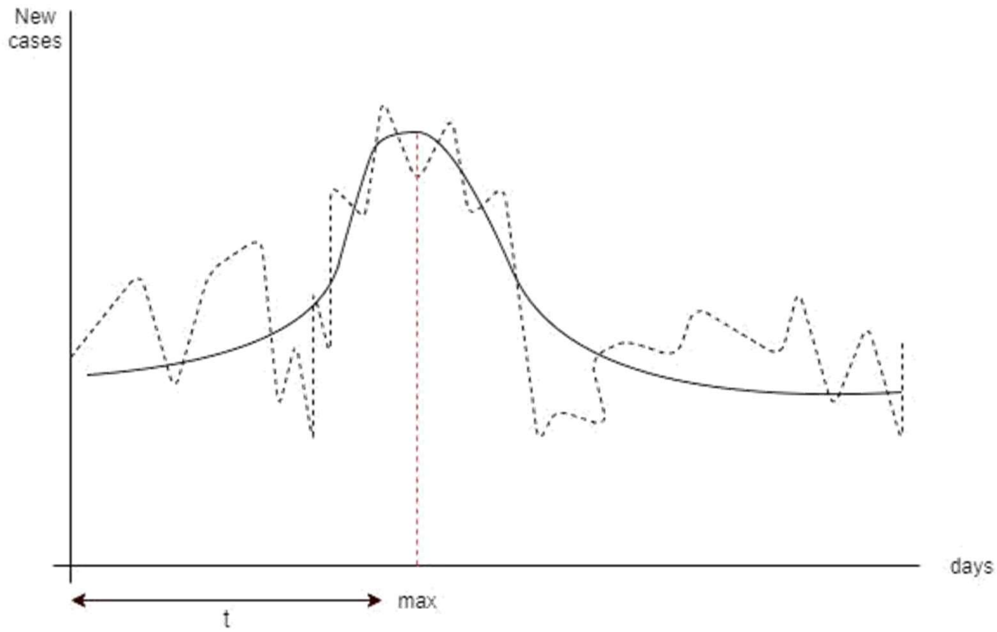
Data and Design

This section introduces a dataset used to identify the variable that will allow us to construct the time-to-event variable, which will work as an endogenous variable and control variables we will use.

First of all, we define the period of analysis used in the research. We collected information between January 1 to August 22 of 2020, for all variables related to COVID-19. For the rest of the variables (covariates), we use annual information. Finally, all data are set at country level for 172 countries in total.

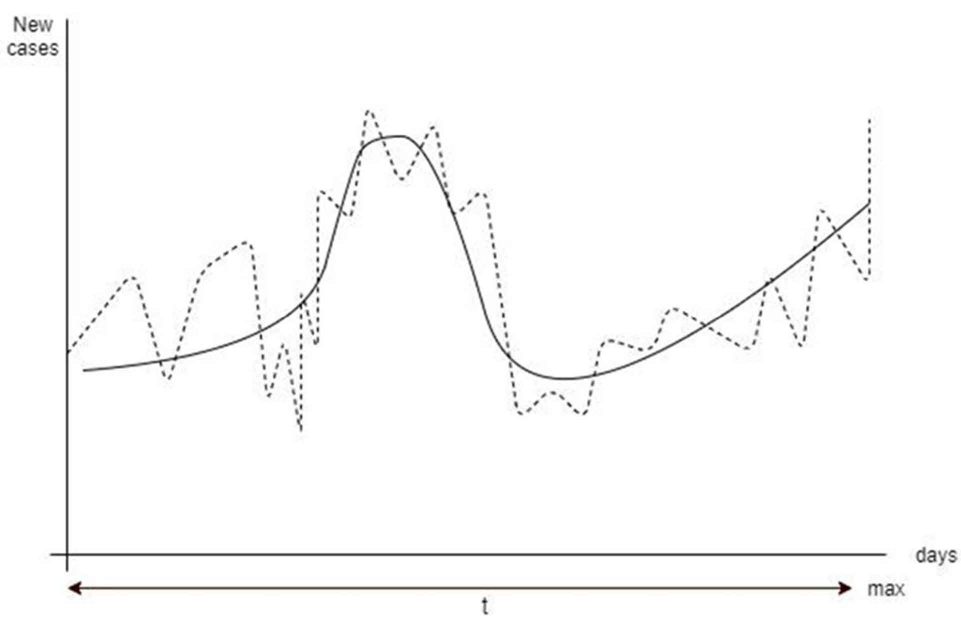
For the time-to-event variable (t), we use the number of days that we have information on the "new cases" of COVID-19 between January 1 to August 22 before the time the maximum daily number of "new cases" is detected. Therefore, if we find that the maximum number of "new cases" is located on February 20, then t is defined as 51. (See Figure 1)

FIGURE 1
NEW CASES EVOLUTION WITH MAXIMUM LEVEL



However, when the evolution of the "new cases" infections of COVID-19 does not show permanent reduction as in Figure 1, the number of days we use a variable t is defined as the number of days with data available. (See Figure 2)

FIGURE 2
NEW CASES EVOLUTION WITH MAXIMUM LEVEL



It is essential to mention that the event of interest determines the endogenous variable's value. In this case, the event we are analyzing is whether the daily number of new infections reached a maximum level and then observed sustained reductions. Then, for countries where the daily number of infected reaches a maximum and then falls steadily, such as in Figure 1, the variable t takes the value equivalent to the number of days where data is available until the day when the number of new infections reaches its maximum value. On the other hand, when there is no evidence of a global maximum, as in Figure 2, the variable t takes the value equivalent to the number of days where data are available until August 22. The data is collected from the European Centre for Disease Prevention and Control.

Three variables are used as covariates: (1) the urban population (percentage of the population), where urban population refers to people living in urban areas as defined by national statistical offices. It is calculated using *World Bank* population estimates and urban ratios from the *United Nations World Urbanization Prospects*. The data source is the *World Development Indicators (World Bank)*. (urbanization), (2) the vulnerable employment, total (percentage of total employment) estimated by the *International Labor Organization (ILO)*, (vulnerability), (3) the percentage of the population older than 65, (age65) obtained from the *World Development Indicators (World Bank)*.

Finally, the categorical variable we use to agglomerate the estimation is based on the income groups classification defined as "Low-income countries" for countries with Gross National Income (GNI) per capita in current US$ lower than 1 036. "Middle-income countries" for those with GNI per capita in current US$ between 1 036 and 12 535. Finally, "High-income countries" for those with GNI per capita in current US$ higher than 12 535.

Methodology

In this section, we explain the Kaplan-Meier methods and the Cox regression, which will help us analyze the duration of the infection levels until they reach their maximum level.

Kaplan-Meier Method

The Kaplan-Meier method is a non-parametric methodology for estimating time-related events. In simple words, the method is used to analyze the 'death', or more specifically, the end of an event. For example, in labor economics, when analyzing when a person ceases to be unemployed, in health economics, when analyzing the time when a person dies from an illness. (Kaplan and Meier, 1958)

This methodology assumes a large reduction in the calculation volume because the survival time is calculated in each time period when the pre-established events occur and stop counting for future calculations. The Kaplan-Meier method has the following stages:

- List the time when the event (pre-established) occurs.
- Find for every participation time the number of subjects that continue to participate in the event.
- Establish the number of subjects who achieved the event (pre-established) within a time interval (nx).
- Calculus of the probability of occurrence of the event (pre-established). The calculus is based on the following equation:

$$qx = \frac{dx}{nx} \tag{1}$$

where x is the participation duration, nx is the time interval, and dx is the participation interval. Specifically, by using the COVID-19 data, the survival function $S(t)$ is expressed as the following rate:

$$S(t) = \frac{\text{Number of days until reach the max level of daily new cases of COVID-19}}{\text{Total number of days with daily new cases observed of COVID-19}}$$

This function denotes the probability of new cases duration until they reach the maximum level at the time t or longer before observing a sustained and permanent reduction of those new cases. This survival function is denoted by:

$$S(t) = P(T \geq t) = 1 - F(t), \tag{2}$$

where T represents the survival time or duration of the new cases of infection of COVID-19 before reaching its maximum level. F(t) is the distribution function of T, which measures the probability time of survival of daily new cases of COVID-19 duration up to time t.

Therefore, the estimator of the Kaplan-Meier survival model is given by as follows:

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right), \tag{3}$$

where t_i is the survival time of the daily new cases of COVID-19 at the point i; $d_i$is the number of days where the daily new cases of COVID-19 reach the maximum number beforet_i. The survival function is the product of the conditional probabilities. This method is usually used as a preliminary evaluation since it is a descriptive method for evaluating a single variable.

Cox Regression

The Cox regression can be used as a methodology to determine if a set of covariates or characteristics are affecting survival. Also, we can determine the magnitude and direction of those effects. In this sense, it is necessary to identify those covariates to calculate a predictive indicator of survival. The method to determine that indicator of survival is through the Cox proportional-hazard regression. This semi-parametric method enables us to determine the effect of each covariate over the hazard (Cox, 1992). The model is determined by:

$$\lambda_i(t) = e^{X'_i\beta} \lambda_0(t), \qquad i = 1,2,3,\ldots \tag{4}$$

where n is the number of individuals or countries of our data set, 172 for our case (Riffenburgh, 2012). $X_i = (X_{i1}; X_{i2}; X_{i3}; \ldots; X_{ik})'$ is the vector of covariates, $\beta = (\beta_1, \beta_2, \beta_3, \ldots, \beta_k)'$ is the vector of regression coefficients, $\lambda_i(t)$ is the hazard calculated for each individual i, and λ_0 (t) is the baseline hazard. This baseline hazard function corresponds in our case to the probability of reach the maximum number of daily new cases of COVID-19 when all the variables are 0. A more detailed explanation of the model can be found in Ham and Rea (1987), and Mittelbock (2004). More specifically, in our case, the model can be defined as:

$$\lambda_i(t) = \lambda_0(t)\exp\,[\beta_1\text{urbanization}+\beta_2\text{vulnerability}+\beta_3(\text{urbanization})x(\text{vulnerability}) + \beta_4\text{age65}+\varepsilon_i], \quad i = 1,2,3,\ldots,172 \tag{5}$$

where *urbanization* is 1 when the percentage of the population living in urban areas is equal or higher than 50%; 0, otherwise. *vulnerability* is defined as 1 when the share of vulnerable employment is equal to or higher than 50%; 0, otherwise. The *ILO* calculates these numbers. Finally, *age65* measures the percentage of the population is 65 years old or older.

RESULTS

This section analyzes the covariates' impact on the survival of the daily new cases of COVID-19 by using the econometric models developed in Section 2.

Figure 3 shows the Kaplan-Meier estimator and graphs the survival probability of the event we are analyzing. In particular, the point at which new daily COVID-19 infections reach their maximum level before steadily decreasing is explored here as an event. The graph shows that, on average, during the first 50 days, more than 80% of the countries have not yet reached their maximum daily transmission point. Likewise, 50% of the sample countries reach their maximum infection level and then steadily decline in approximately 160 days. Finally, more than 30% of the sample countries did not get their maximum point during the analysis period.

FIGURE 3
KAPLAN-MEIER. DAILY NEW CASES OF COVID-19

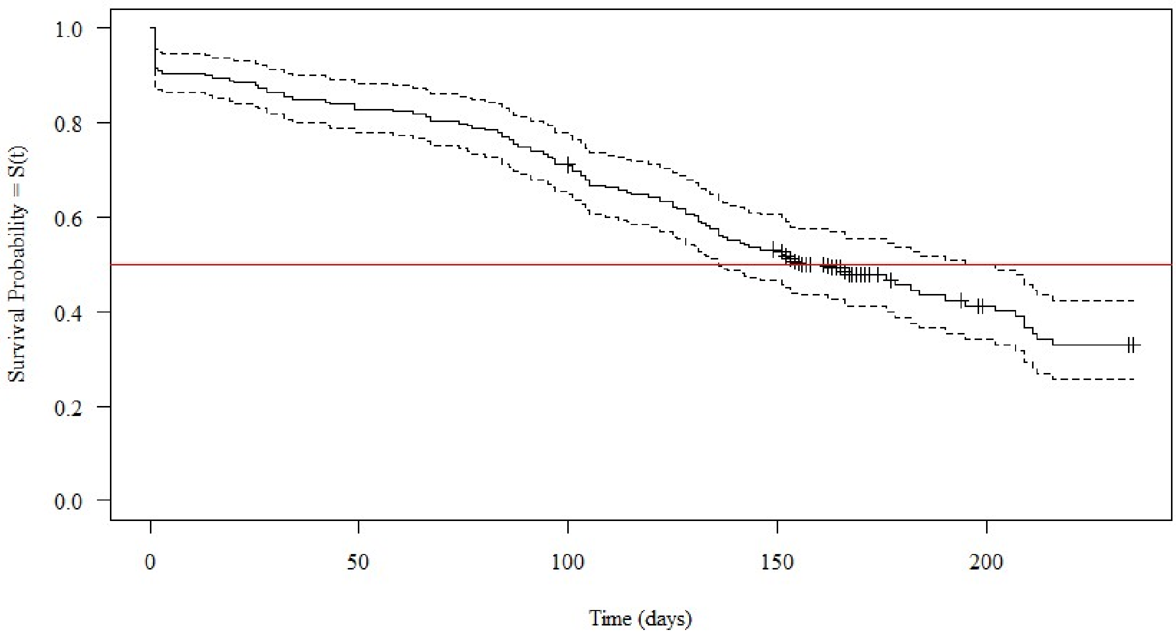


Figure 4 shows the estimation of the non-parametric model Kaplan Meier, but this time conditioned to urbanization's dummy variable. The figure shows that countries with high urbanization levels, countries where more than 50% of the population live in urban areas, have higher survival levels than countries with low urbanization levels (mostly rural countries). In other words, during the first 100 days, a primarily rural country has a probability of approximately 60% that it has not yet reached its peak of daily infections. On the other hand, a mostly urban country has a roughly 75% probability that it has not yet reached its peak of everyday infections. These results are consistent because, in more urbanized countries, infection rates are more persistent. After all, inhabitants are concentrated in more limited spaces, such as large cities. While in countries with a high percentage of the population living in rural areas, social distance is easier to maintain than in high concentration areas.

Similar to Figure 4, Figure 5 shows the survival probabilities using the vulnerability dummy variable. The figure shows that countries with levels of vulnerable employment greater than 50% have lower survival levels. A country with high vulnerable employment peaks in new daily COVID-19 infections more quickly than countries with non-vulnerable employment. Thus, during the first 150 days of analysis, countries with high levels of vulnerable employment have approximately a 40% probability of survival before reaching their peak of daily COVID-19 infections, while countries with non-vulnerable employment have almost 60% probability of survival before reaching their peak.

FIGURE 4
KAPLAN-MEIER. DAILY NEW CASES OF COVID-19 CONDITIONED TO THE LEVEL OF URBANIZATION

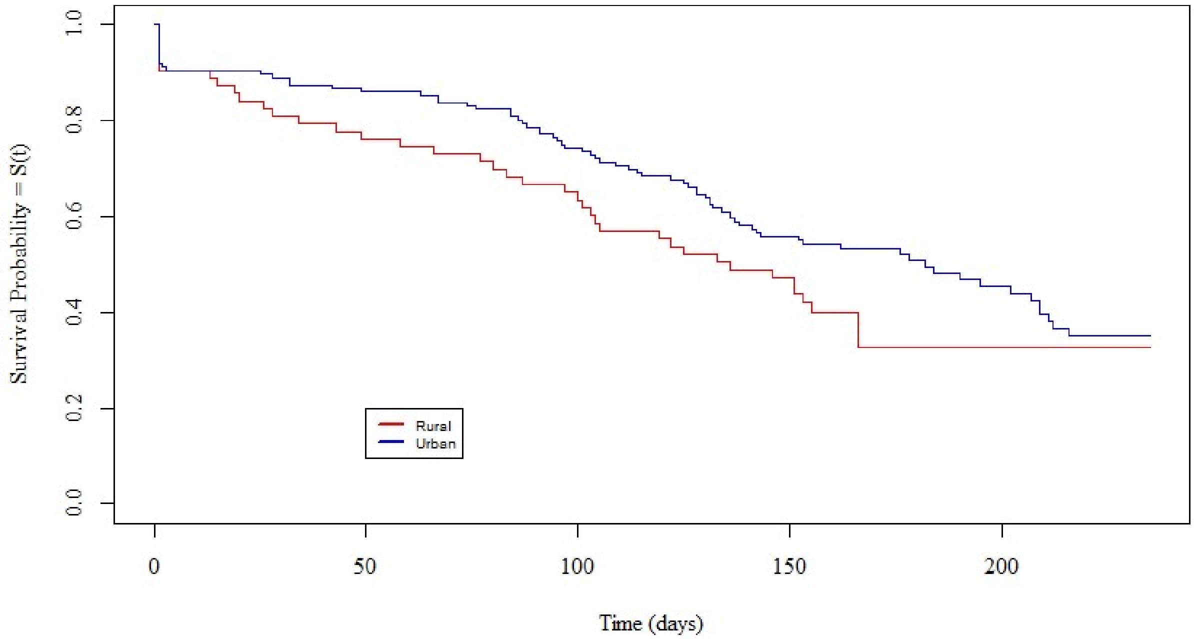


FIGURE 5
KAPLAN-MEIER. DAILY NEW CASES OF COVID-19 CONDITIONED TO THE LEVEL OF VULNERABILITY IN THE EMPLOYMENT

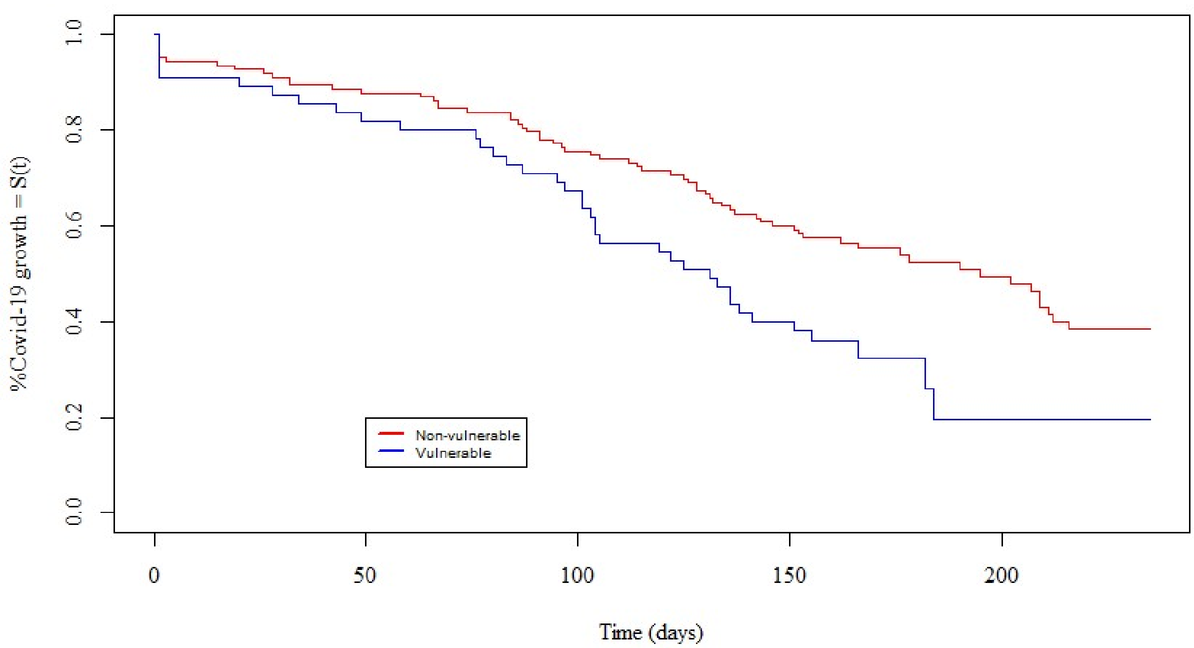


Table 1 shows the results of the estimation of the Cox model using as covariates the dummies of urbanization (*urbanization*), vulnerability (*vulnerability*), a dummy of interaction between urbanization and vulnerability (*urbanization*vulnerability*). Likewise, a continuous control variable is added on the size of the vulnerable age population (*age65*). The different structures and models are represented in each column.

TABLE 1
ESTIMATION RESULTS OF THE COX MODEL

	Dependent variable: Maximum level of daily new cases of COVID-19.						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
urbanization	0.693		0.88	0.684	0.812	0.818	0.78
	(0.199)		(0.269)	(0.362)	(0.37)	(0.37)	(0.249)
vulnerability		1.778	1.644	1.191	1.096	1.07	
		(0.208)	(0.266)	(0.389)	(0.39)	(0.392)	
urbanization*vulnerability				1.673	1.424	1.452	1.549
				(0.506)	(0.507)	(0.509)	(0.342)
age65					0.962	0.963	0.962
					(0.02)	(0.02)	(0.02)
Clusterized by Income Level	NO	NO	NO	NO	NO	YES	YES
Observations	199	178	178	173	177	172	172
Concordance	0.545	0.561	0.574	0.567	0.59	0.586	0.588
Max. Possible R2	0.996	0.996	0.996	0.996	0.996	0.996	0.996
Log Likelihood	551.23	479.78	479.67	467.28	476.69	464.86	464.87
Wald Test	3.39	7.69	7.94	139.93	11.6	38.45	47.05
LR Test	3.252	7.269	7.494	7.589	12.124	11.102	11.072
Score (Logrank) Test	3.425	7.893	8.151	8.1	12.044	11.004	10.928

Note: p<0.1; p<0.05; p<0.01.
The number in parenthesis below the coefficients are the standard errors.

Column 1 shows the results of the estimation when urbanization is used as an exogenous variable. The coefficient corresponding to urbanization with a value of 0.693, statistically significant at 10%, shows that countries with high levels of urbanization have 0.693 times the probability of reaching the maximum point of infection of new daily cases of COVID-19 compared to countries with low levels of urbanization. In simple terms, a country with an urban population greater than 50% is 30.7% less likely to reach the peak of daily COVID-19 infections simultaneously as a country with a mostly rural population. As explained above, this result is based on the idea that in highly urbanized countries, i.e. people are crowded together in limited spaces, as is the case of large cities. These results mean that the number of infections will continue to grow continuously and will take time to reach a maximum.

Similarly, column 2 shows the results when the vulnerability dummy is used as a covariate. The results show that countries with vulnerable employment greater than 50% have 1,778 times the probability of reaching their maximum daily COVID-19 infection point at a given point in time. This result is statistically significant at 1%. Additionally, this outcome can be explained by the fact that one of the characteristics of vulnerable employment is the volatility and unpredictability of jobs. For example, this fact forces people in labor vulnerability situations to remain in work to earn sufficient income even when this means not maintaining the social distance or quarantine established by governments.

Column 3 shows the results when both dummies are included in the estimate. It is observed that in this case, the urbanization variable loses significance, and only the vulnerability variable remains significant. Additionally, in column 4, we incorporate a variable of interaction between two dummy variables. This interaction variable helps us capture the duration and probability of reaching the maximum point of daily COVID-19 infections in countries with high urbanization and high levels of vulnerable employment. However, the variable does not appear to be statistically significant. In column 5, we include an additional covariate, age65, to control for the effect of a population susceptible to COVID-19. In this case, it is observed that countries with 1% more population over 65 years old reduce the probability of reaching the maximum level of daily infections by 3.8%. In short, countries with a larger population at risk take longer to reach the maximum point of everyday COVID-19 infections. This outcome may be because countries with high levels of vulnerable populations have greater precautionary behaviors to avoid or delay infection chances.

The control variables do not incorporate the GDP per capita or any similar variable because this could be correlated with urbanization and vulnerability, distorting the estimated coefficients. In this way, a way to control income levels is solved by clustering the data according to the sample's income levels. We defined "Low-income countries" for countries with Gross National Income per capita in current US$ lower than 1 036. "Middle-income countries" for those with GNI per capita in current US$ between 1 036 and 12 535. Finally, "High-income countries" for those with GNI per capita in current US$ higher than 12 535.

Column 6 shows the same results as the estimate in column 5, but the data is clustered by income level. Within the estimated coefficients, it is observed that the urbanization variable, the interaction variable between *urbanization* and *vulnerability*, and the variable that controls the size of the population over 65 years old are statistically significant. In the case of the *urbanization* variable, countries with high urbanization have 18.2% less probability of reaching the maximum level of daily infections at a given moment than a country with urbanization levels below 50%. Likewise, countries with high urbanization and high vulnerability have a 45.2% greater probability of reaching the maximum daily infection level. In other words, these countries achieve their maximum level of daily infections more quickly because the situation of labor vulnerability forces workers to leave in search of income even when this implies an increase in their risk of infection. Finally, the *age65* variable shows that countries with 1% more population over 65 years old reduce the probability of reaching the maximum point of daily infections by 3.7%.

In the end, in column 7, the results of Cox's model estimation are observed when the vulnerability variable is omitted. These results are also clustered by income level. To eliminate the vulnerability dummy variable, we use the contrast of likelihood ratios. With a p-value of 0.8623, it is not possible to reject the null hypothesis that the model estimated in column 6 and the model estimated in column 7 are statistically different. Thus, the vulnerability dummy variable can be excluded without losing the predictive power of Cox's model. The results of the estimation are similar to those observed in column 6.

Below the row where the number of observations is shown, the levels of "concordance" for each estimate are shown. In all cases, the agreement levels are greater than 55%; that is, in more than 55% of the cases, the model predicts well the occurrence of the maximum points of new daily COVID-19 infections. Additionally, Wald's and likelihood ratio tests show that the estimated coefficients are overall relevant in the estimated models. Finally, the score test (log-rank) shows that the survival probabilities among the different groups analyzed are different under the estimated coefficients' significance levels.

In the annex, Figures 6, 7, and 8 show that the coefficients estimated in the model in column 7 are linear since no structural changes are observed in the estimated errors. The figures show that the perpendicular line remains within the confidence interval for the estimated errors at the zero level. There is one exception in Figures 7 and 8, where there is a small portion where the line falls over the boundary or slightly outside the confidence interval. This fact remains a limitation of the model that can be resolved with new data or further analysis of structural change in the variables used in the estimation.

CONCLUSIONS

This study was motivated by the fact that the daily COVID-19 infections' behavior has been shown differently among countries. However, there was no clear evidence that certain countries' characteristics affect this variable's behavior until now. Therefore, this research seeks to contribute to the elucidation of this question. For this reason, we used a survival analysis of the new daily infections of COVID-19. Specifically, we use the Kaplan-Meier method and Cox regression. The first one to graph the survival levels before reaching the maximum point of daily COVID-19 infections; additionally, we based those survival levels on the dummy variables of urbanization (when more than 50% of the population lives in urban areas) and vulnerability (when more than 50% of the employment is in a situation of vulnerability). In the results, we observed that highly urbanized countries take longer to reach the peak of daily COVID-19 infections because it is more difficult to maintain social distance in urban areas. Many people live in agglomeration in reduced spaces, which causes everyday infections do not reduce quickly. Even more, they do not reach their peak early.

On the other hand, the vulnerability variable gives evidence that countries with high levels of vulnerable employment cause the peak of daily infections to be reached more quickly than in countries with non-vulnerable employment. This result is because people in vulnerable employment have undefined or permanent incomes. Many people search for work and income to solve this problem, even when this increases their risk of infection since there is a trade-o between hunger and COVID-19 infection in their decisions. This fact causes contagion to increase rapidly until it reaches the maximum point of daily contagion of COVID-19 sooner than countries with non-vulnerable employment.

On the other hand, when we use the Cox model, we observe that the covariates used are statistically significant after controlling the model by clustering the data by income levels. Specifically, we observe that urbanization's dummy variable causes countries with high levels of urbanization to be 22% less likely to reach the peak of daily infections at a given point in time than highly rural countries. The vulnerability dummy variable by itself does not prove to be significant. Still, when it interacts with urbanization, we observe that countries with high urbanization and high vulnerable employment are 54.9% more likely to reach their peak daily COVID-19 infections at a given point in time. In other words, countries with vulnerable employment and urbanization reach their peak of infection more quickly than their counterparts. Finally, the vulnerable population, a percentage of the population aged 65 or older, means that countries with 1% more population sensitive to the COVID-19 are 3.8% less likely to reach their peak. This result is possible because countries with many susceptible people have social behavior protocols that prevent infection levels from expanding rapidly.

Finally, we observe that all these results show that the countries' social characteristics affect the daily levels of COVID-19 infection. In that sense, they modify the probability of reaching the maximum point of everyday infections and then continuing with a permanent plateau. The survival of infections before reaching the maximum point is then affected by the conditions of urbanization and labor vulnerability of the sample countries.

These results can be used as inputs for developing public policies that help fight the spread of pandemic viruses such as COVID-19. First, the evidence that urbanized countries make infections spread by delaying the peak time would be relevant to establish behavioral protocols to avoid the sustained increase of daily infections. Second, vulnerable employment appears to be crucial as it increases contagion and prevents a peak before a sustained decrease. This fact generates the need to establish clear policies that allow workers to avoid searching for income and work even when the risk of contagion increases.

REFERENCES

Bassan, B., Marcus, R., Meilijson, I., & Talpaz, H. (1997). Parameter estimation in differential equations, using random time transformations. *Journal of the Italian Statistical Society*, *6*, 177-199.

Cox, D.R. (1992). Regression Models and Life-Tables. In *Springer Series in Statistics* (pp. 527-541). Springer New York.

Friedman, J., Liu, P., & Gakidou, E. (2020). Predictive performance of international COVID-19 mortality forecasting models. *medRxiv*, pp. 1-15.
Ham, J.C., & Rea, S.A. (1987). Unemployment Insurance and Male Unemployment Duration in Canada. *Journal of Labor Economics*, *5*, 325-353.
Jin, J-M., Bai, P., He, W., Wu, F., Liu, X-F., Han, D-M., . . . Yang, J-K. (2020). Gender Differences in Patients With COVID-19: Focus on Severity and Mortality. *Frontiers in Public Health*, *8*, 1-6.
Kaplan, E.L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, *53*, 457-481.
Kundu, S., Kirti, & Mandal, D. (2020). *A Study on Survival Scenario of COVID-19 patients in India: An Application of Survival Analysis on patient demographics* (pp. 1-15).
Meilijson, I., & Alon, N. (2020). *Random time transformation analysis of Covid19 2020* (pp. 1-15).
Mittelbock, M. (2004). Book Review: Survival Analysis: Techniques for Censored and Truncated Data, By John P. Klein and Melvin L. Moeschberger. *Biometrical Journal*, *46*, 379-379.
Nemati, M., Ansary, J., & Nemati, N. (2020). Machine-Learning Approaches in COVID-19 Survival Analysis and Discharge-Time Likelihood Prediction Using Clinical Data. *Patterns*, *1*, 1-8.
Riffenburgh, R. (2012). Survival, Logistic Regression, and Cox Regression. In *Statistics in Medicine*, Elsevier, pp. 491-508.
Salinas-Escudero, G., Carrillo-Vega, M.F., Granados-García, V., Martínez-Valverde, S., Toledano-Toledano, F., & Garduño-Espinosa, J. (2020). A survival analysis of COVID-19 in the Mexican population. *Research Square*, pp. 1-13.
Sousa, G.J.B., Garces, T. S., Cestari, V.R.F., Florencio, R.S., Moreira, T.M.M., & Pereira, M.L.D. (2020). Mortality and survival of COVID-19. *Epidemiology and Infection*, *148*, 1-6.
Wang, Z., Ji, J.S., Liu, Y., Liu, R., Zha, Y., Chang, X., . . . Wu, X. (2020). *Survival analysis of hospital length of stay of novel coronavirus (COVID-19) pneumonia patients in Sichuan, China* (pp. 1-16).
Zhao, Y., & Dilip, D. (2020). *Survival Analysis of COVID-19 on Democracy with Cox Proportional Hazards Model* (pp. 1-7).

APPENDIX

FIGURE 6
LINEARITY ESTIMATOR ANALYSIS FOR URBANIZATION ESTIMATOR

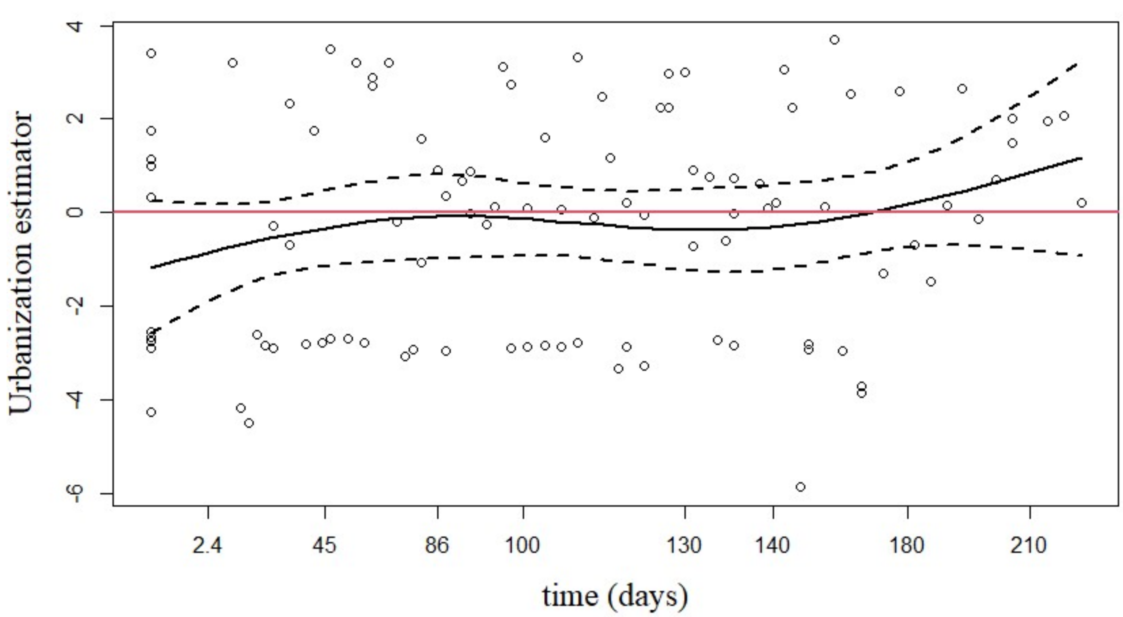


FIGURE 7
LINEARITY ESTIMATOR ANALYSIS FOR URBANIZATION*VULNERABILITY ESTIMATOR

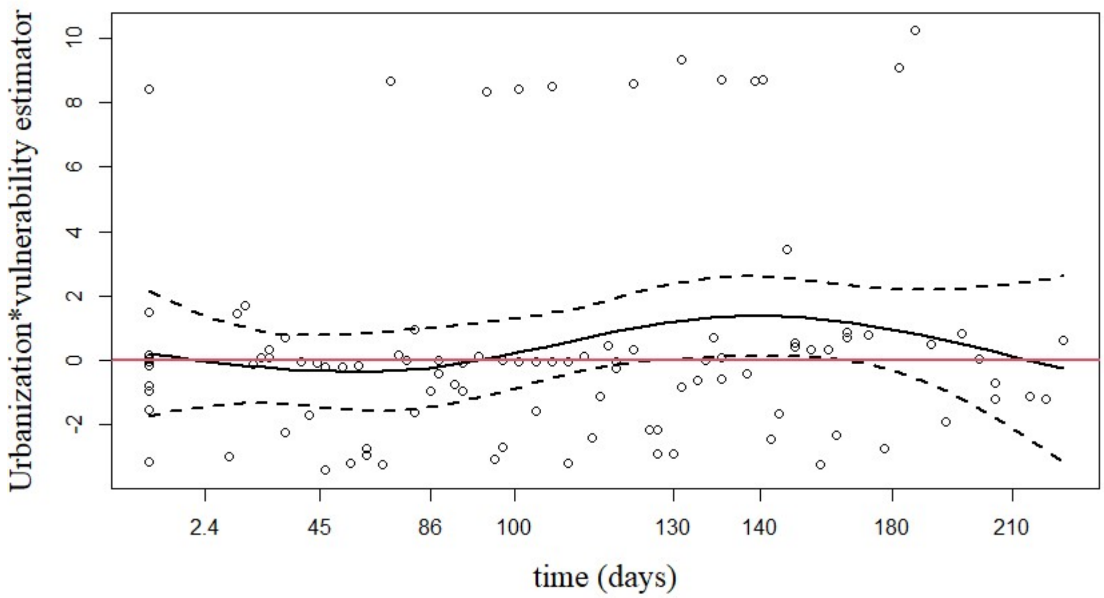


FIGURE 8
LINEARITY ESTIMATOR ANALYSIS FOR AGE65 ESTIMATOR