SURVIVAL ANALYSIS OF COVID-19 INCIDENCE IN KOSOVO

Roberta Bajrami, Adelina Gashi, Medain Hashani

AAB College - Kosovo, Faculty of Economics

Corresponding author: Roberta Bajrami, roberta.bajrami@aab-edu.net

ABSTRACT: The human race has been at the edge of the COVID-19 pandemic since the start of 2020. While the disease is easily transmissible, a large proportion of the people affected are recovering. Most recovered patients do not suffer COVID-19 death, even though they have been observing for a long time. In the sense of survival analysis, they can be viewed as long term survivors (cured population). In this study, we present some statistical methods for estimating the cure fraction in Kosovo of COVID-19 patients. Proportional hazards Mixture cure model is used to estimate the fraction of cure and the effect of gender and age covariations on lifetime. For this analysis the data available on the https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.xlsx' website is used. The result revealed that the covariates, diabetes prevalence and hospital beds per thousand have highly statistically significant coefficients, while others, that is stringent index, total cases, gdp per capita (economic variable), respondent's age, handwashing facilities are not statistically significant, implying that these variables are not really contributing to the hazard ratio of covid-19 incidence.
KEYWORDS: GDP per capita, Covid 19, respondent's age, handwashing facilities

1. BACKGROUND TO THE STUDY

The outbreak of the coronavirus disease has had tremendous impact of individuals and the economies of nations. It has negatively affected extremely poor, as well as advanced economies. The impact has left no one in doubt of the fragility of the world economy, particularly in terms of the health facilities and manpower.

As of January 2020, a global health epidemic has been created by the outbreak of the new 2019 coronavirus disease (COVID-19). On 31 December 2019, COVID-19 was first identified as a pneumonia of unknown etiology in Wuhan, Hube Province, China (Adhikari, et al., 2020). By May 19,2020, the virus had spread all over the world. Nearly 50 million people were infected (4982937) and there were more than three million and twenty thousand deaths (324554), with more than 19 million people suffering from the outbreak (1958416). On 11 March 2020, the World Health Organization (WHO) announced the outbreak of the new coronavirus (COVID-19) a global pandemic (Cucinotta & Vanelli, 2020). On the 13th of March, the first cases of coronavirus disease were recorded in Kosovo. The number of new daily cases, however, did not rise until March 22, with the first coronavirus death confirmed. Older adults are at increased risk of becoming seriously ill from the outbreak, and one-third of all virus infections in Kosovo are made up of people aged 60 and over. Prishtina, the most populous district in the Kosovo, has been badly affected and accounts for more than 90% of all COVID-19 casualties in Kosovo. There were 46,909 confirmed coronavirus infections and 1203 deaths from the disease in Kosovo as of August 20, 2020. The construction of hospital centres and rapid recovery efforts ensure that the lion's share of victims from the outbreak were fully recovered.

Typical mathematical models in the study of epidemiological data are the SIR model (Susceptible, Infected and Recovered) and related compartment models (Pollicott, Wang & Weiss, 2012). Several COVID-19 patient data analyses were recorded using SIR models and other normal statistical techniques (Sreedevi & Sankaran, 2020). Our goal in this research is to analyze patient data from Kosovo COVID-19 using survival models.

We model time in the study of survival for an event of importance to occur. We may describe death in this sense as the occurrence of concern due to COVID-19. Time is regarded as life. For different factors, the event's final time of incidence is not available for certain people (patients). This state leads to the phenomenon of censorship known as the partial lifespan data is only valid for censored patients; it is longer than a given duration. Censorship capability helps longevity statistics to be analyzed independently from all other statistical fields. Date of disease conformation and date of death/recovery due to (due to the disease, records on age and gender are available for some COVID-19 patients in Kosovo. The recovered and hospitalized patients may be treated for lifetime censorship until the case is described as death due to COVID-19, as we do not have any details on them after the date given (as per records).

This study uses cox regression model to analyze data on COVID-19 patients in Kosovo.

2. METHODOLOGY

2.1. Model

Suppose we have n patients under study in general. The time to death of a COVID-19 patient from the date the disease is identified is described as T. T would then be counted as the number of days in the hospital. For patients whose death happens, the number of days hospitalized is counted as life time experienced. When the patient is healed, we just know that the incident is not actually happening to the patient, so the number of days treated with such patients is called a censored lifespan. For patients with existing status as 'hospitalized' after the specified date, there is still no details available. Those lives are also treated as censored lives, too. Survival analysis is a set of mathematical techniques for analyzing the frequency and timing of occurrences and is useful in studying many forms of social and natural science events.

2.2 Cox Regression Model

The aim of the model is to determine the influence of many variables on survival simultaneously. In other words, it helps one to analyze how such variables affect the probability of occurrence of a single case (e.g., illness, death) at a given point in time. This rate is frequently referred to as the hazard rate. In the survival-analysis literature, indicator factors (or variables) are typically called covariates.

The Cox model is represented by the *hazard function* denoted by $h(t)$. In short, the hazard function can be defined as the chance of dying at the moment of death. Estimated below as:

$$h(t) = h_0(t) \times exp(b_1x_1 + b_2x_2 + \ldots + b_px_p)$$

where,

- t represents the survival time
- $h(t)$ is the hazard function determined by a set of p covariates $(x_1, x_2, \ldots, x_p)$
- the coefficients $(b_1, b_2, \ldots, b_p)$ measure the impact (i.e., the effect size) of covariates.
- the term h_0 is called the baseline hazard.
- The Cox model can be written on the variables x_i as a multiple linear regression of the hazard logarithm, with the baseline hazard being an intercept term that differs with time.

The $exp(b_i)$ quantities are referred to as hazard ratios (HR). A b_i value greater than zero, or a hazard ratio greater than one, means that the incident hazard increases as the value of the i^{th} covariate increases and thus the length of life decreases.

In other words, a hazard ratio above 1 reveals a covariate that is positively correlated with the likelihood of an occurrence and is thus adversely associated with the period of survival.

In synopsis, the $HR = 1$ means no effect, while $HR < 1$ means reduction in the hazard and $HR > 1$ indicates an increase in hazard

3. RESULTS

The data used was obtained online from https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.xlsx. The complete COVID-19 dataset is a collection of the COVID-19 data maintained by *Our World in Data*. It is updated daily and includes data on confirmed cases, deaths, and testing, as well as other variables of potential interest. Some of the variables used affect the economy such as, GDP per capita and extreme poverty

The variables used are defined below:

X_1= Total cases
X_2= Stringent index
X_3= GDP per capita
X_4= Extreme poverty
X_5= Diabetes prevalence
X_6= Handwashing facilities
X_7= Hospital bed

Table 1: The variables used affect the economy

<table>
<tr><th>_t</th><th>Haz. Ratio</th><th>Std. Err.</th><th>z</th><th>P>|z|</th><th colspan="2">[95% Conf. Interval]</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Total cases</td><td>.9858058</td><td>.0216116</td><td>-0.65</td><td>0.514</td><td>.9443449</td><td>1.02908</td></tr>
<tr><td>Stringent index</td><td>.9272075</td><td>.1336446</td><td>-0.52</td><td>0.600</td><td>.6990167</td><td>1.22989</td></tr>
<tr><td>GDP per capita</td><td>.9277744</td><td>.050462</td><td>-1.38</td><td>0.168</td><td>.83396</td><td>1.032142</td></tr>
<tr><td>Extreme poverty</td><td>1.004179</td><td>.0059553</td><td>0.70</td><td>0.482</td><td>.9925743</td><td>1.015919</td></tr>
<tr><td>Diabetes prevalence</td><td>1.057945</td><td>.0278935</td><td>2.14</td><td>0.033</td><td>1.004663</td><td>1.114052</td></tr>
<tr><td>Handwashing facilities</td><td>1.070181</td><td>.0586589</td><td>1.24</td><td>0.216</td><td>.9611716</td><td>1.191553</td></tr>
<tr><td>Hospital beds</td><td>1.013832</td><td>.0007342</td><td>18.97</td><td>0.000</td><td>1.012394</td><td>1.015272</td></tr>
</table>

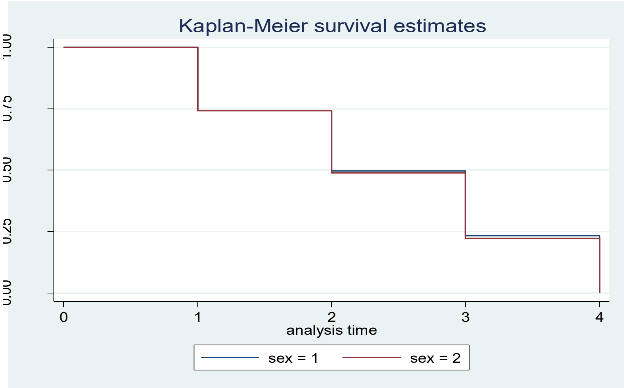

Fig. 1: Kapler-Meier survival estimate

The model obtained is given as

$$h_1(t) = h_0(t)\exp\,(0.9858X_1 + 0.9272X_2 + 0.9278X_3 + 1.0042X_4 + 1.0579X_5 + 1.0701X_6 + 1.0138X_7).$$

The model is stratified by total deaths and this indicates that hospital beds is statistically significant. The likelihood results showed that the p-value of the chi-square is 0.0000 and it indicates that we reject the null hypothesis that the $\hat{\beta}$ coefficients are statistically different from zero.

In Table 1, the column marked z displayed the performance records and the ratio of each regression coefficient to its standard error, a Wald statistic which, under the assumption that the corresponding β is 0, is asymptotically standard regular. The frequency of diabetes covariates and per thousand hospital beds have extremely statistically significant coefficients, while others, which is a strict graph, total incidents, GDP per population, age of respondent, handwashing facilities are not statistically important, indicating that these variables may not actually add to the covid-19 incidence risk ratio. The Kapler-Meier survival function on multiple predictors revealed the approximate survival function S(t) for the time-to-death Cox regression. A point-wise 95 percent confidence envelope around the survival feature reveals the broken lines.

The exponentiated coefficients in the second column of the first panel (and in the first column of the second panel) of the output are interpretable as multiplicative effects on the hazard. Thus, for example, holding the other covariates constant, diabetes prevalence increases the hazard by a factor of $e^{\beta_5} = e^{1.057945}$ on extreme poverty, that is, by 5.6 percent. Similarly, hospital beds per thousand increases the hazard by a factor of 1.013832, or $e^{1.01382}$ that is 1.3 percent.

4. CONCLUSIONS

The regression model is applied to actual survival data of patients within the framework of this research. In this study, only the prolonged resection aspect does not meet the assumption of proportional hazards, and diabetes prevalence and hospital beds per thousand are found to be essential risk factors affecting the failure.

In conclusion, in this work, the incidence of diabetes and hospital beds per thousand have the greatest probability of patient survival, while the other factors have little or no effect on patients' survival. It is for that reason that the advocacy on the prevalence of diabetes should be increased in terms of understanding and alignment, as well as creating more bed spaces to serve patients.

REFERENCES

[1] **Adhikari S. P., Meng S., Wu,Y. J., Mao Y. P., Ye R. X., Wang Q. Z., ... & Zhou H.** (2020). Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. *Infectious diseases of poverty*, *9*(1), 1-12.

[2] **Brinda E. M., Attermann J., Gerdtham U. G., & Enemark U.** (2016). Socio-economic inequalities in health and health service use among older adults in India: results from the WHO Study on Global AGEing and adult health survey. *Public Health*, *141*, 32-41.

[3] **Cucinotta D. & Vanelli M.** (2020). WHO declares COVID-19 a pandemic. *Acta bio-medica: Atenei Parmensis*, *91*(1), 157-160.

[4] **Győrffy B., Lanczky A., Eklund A. C., Denkert C., Budczies J., Li Q., & Szallasi Z.** (2010). An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast cancer research and treatment*, *123*(3), 725-731.

[5] **Hancock D. R. & Algozzine B.** (2017). *Doing case study research: A practical guide for beginning researchers*. Teachers College Press.

[6] **Oyeranti O. & Sokeye B.** The evolution and spread of covid-19 in Kosovo, CPEEL's Covid-19 Volume II Discussion Papers Series
https://cpeel.ui.edu.ng/sites/default/files/publication/evolution-and-spread-covid-19-nigeria.pdf

[7] **Pollicott M., Wang H. & Weiss H.** (2012). Extracting the time-dependent transmission rate from infection data via solution of an inverse ODE problem. *Journal of biological dynamiStringent index*, *6*(2), 509-523.

[8] **Sreedevi E. P. & Sankaran P. G.** (2020). Statistical methods for estimating cure fraction of COVID-19 patients in India. *medRxiv*.