Predicting the Hospitalized Duration of Treatment for Patients Contracted to Covid-19

Abbas Mahmoudabadi and Maedeh Pourmirza Dogahi
Industrial Engineering & Information Technology Department
MehrAstan University, Guilan, Iran
mahmoudabadi@mehrastan.ac.ir, m.pormirza1364@yahoo.com

Abstract

While the outbreak of Covid-19 pandemic is a severe global health problem, health authorities are focusing on managing patients' treatment and healthcare facilities. The aim of this study is to investigate the relationship between personal characteristics, symptoms, and underlying diseases nominated as predictors, and the hospitalized duration of treatment for the patients contracted to Covid-19 as a dependent variable. To this purpose, patients have been divided into three categories of recovered, transferred, and death followed by developing separate regression models with the above-mentioned variables. Data have been collected for 1015 patients who were positively contracted to Covid-19 and received clinical treatment or intensive care. The results revealed that age causes significant effects on the hospitalized duration of treatment among the aforementioned contributing factors, so the health service managers can estimate the patient hospitalized duration based on the age even if the symptoms and underlying diseases may cause significant effects on the infection rate.

Keywords
Covid-19, Regression Analysis, Hospitalized Duration, Health-care Management

1. Introduction

Covid-19, a viral infection and a severe respiratory disease with rapid human-to-human transmission, causes a significant risk to patients with weak immune systems (Di Mascio et al. 2020; Lai et al. 2020). Patients contracted to Covid-19 may experience a variety of symptoms based on their immune system because the virus hits the respiratory system (Almqvist, et al. 2020). In this case, the elderly people or those who are suffering from underlying diseases may suffer more severely than the others (Law et al. 2020), so they need to receive specific health care at the intensive-care unit (ICU) (Sun et al. 2014). This situation makes difficulties for health authorities to manage health care operations including hospital facilities, nursing operations, drug distribution, and ultimately manage admiring patients or transferring them to the other clinical service centers. Therefore, studying the specification of treatment should be received by the above patients is the concept behind this research work, where statistical methods are utilized to investigate the relationship between personal characteristics and treatments received by the patients.

Pattern recognition is one of the most applied models among statistical methods in many scientific fields in which the foremost goal is to predict a variable based on predictors known as the dependent variables. Regression analysis is a statistical model for predicting a variable based on one or more other variables where linear regression predicts a linear combination of independent variables to estimate the dependent variable (Hosmer et al., 2011). In the multiple linear regression, the coefficients of the linear model are simultaneously estimated based on the observations in which the sum of square errors, between the model outputs and observations, is minimized (Alexopoulos, 2010). The mathematical relation between dependent and independent variables is formulated by equation (1) where X_1, X_2, …,X_k are predictors or independent variables, and Y represents the dependent variable. The most practical method to determine the coefficients (β_1, β_2, …, β_k) is least-square errors in which the sum of square errors is minimized through determining the estimating coefficients (Montgomery et al. 2012).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \quad (1)$$

In the last decades, many studies have been conducted utilizing forecasting methods where data have been analyzed by data mining techniques and multiple linear regressions to improve the accuracy of estimations. For example, a

study on predicting the cost of public healthcare in Tsuyama Hospital, Japan, has been conducted to develop a linear regression and concluded that the forecasting models are capable of more accurately predicting health care costs (Panay et al. 2019). Studies on estimating patients' mortality ratios or other medical measurements have been also observed in the literature in which regression analysis is widely utilized over populations under the study (Hogan et al. 2015). Many studies have been conducted to investigate on diseases like the prognosis and likelihood of heart disease with various symptoms, because heart diseases kill one person every 40 seconds, according to the American Heart Association (Polaraju& Prasad 2017). Comparing situations is another research field where in terms of long-term healthcare, the study of the healthcare system forecast and its impact on health costs through linear regression in Colombia showed that long-term treatments are costly for insurers and patients (Riascos & Serna 2017).

As the recent studies, the Johns Hopkins University conducted a study to predict the prevalence of Covid-19 based on regression analysis and determined the most contributing factors in disease outbreak in short term (Pandey et al. 2020) to manage health care operations during the pandemic of Covid-19. At the beginning of the outbreak, the Iranian big cities received many travelers from other countries mainly from UAE, Chian, Oman, Iraq, so the case-studies on Covid-19 are required to focus more on managing the healthcare system because there is no evidence of stopping the virus outbreak.

Following the above, developing a model to estimate the ICU occupation condition based on patients' symptoms, personal characteristics and underlying diseases would support medical authorities to manage healthcare operations and their nursing capabilities. Therefore, the multivariate regression method is utilized to predict the number of days that patients may have to stay in the hospital if they are suspicious to be contracted to Covid-19.

2. Committed Variables and Data Collection

To develop the regression model, the hospitalized duration of treatment represents the dependent variable and symptoms, general conditions such as age, gender, and eventually underlying diseases which patient may be suffered from are predictors, so the hospitalization duration treatment is estimated based on the symptoms, general conditions, and underlying disease. The regression models are developed for three separate groups of recovered, transferred, and death, which all indicate the type of clearance in hospitalizing terminology. The patient symptoms naturally require being under control and are regularly measured and recorded (Farzandipour et al. 2010) while there are many underlying diseases (Emami et al., 2020) committed to the patients. In addition to underlying diseases that weaken the human immune system and make the patient be more committed to viruses, elderliness is also one of the risk factors for increasing the death of Covid-19, according to the previous studies (Kowalski et al., 2020).

Therefore, the "Age" is under the current investigation as one of the most important contributing factors. Data including age, gender, symptoms, and underlying disease for 1015 patients have been collected from February 18 to August 20, 2020, in the northern Iranian province of Guilan for six months. Age is measured by year, but symptoms and underlying disease are indicated by a binary digit of (0 and 1) where 1 indicates that the patient is suffering from underlying disease, otherwise 0. The similar values are also assigned for symptoms. Data is composed of 603 recovered, 251 transferred or cleared according to personal satisfaction and 161 deaths depicted in figure 1.

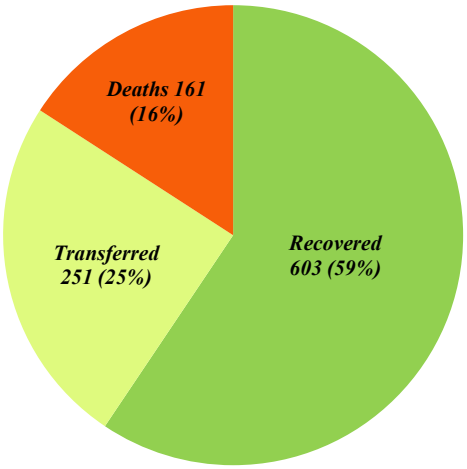

Figure 1: Proportions of clearance type

Demographic analysis has been carried out to know more on data and the results tabulated in Table 1. Age is also grouped into five groups with the different average times on the hospitalized duration of treatment and the number of patients in each group in the last column. All categories have been indicated by gender shown in the middle of the table. The aggregate numbers of females and males are 427 and 588.

Table 1.Demographic analysis of patients contracted to Covid-19

Variable	Group	Average HD*	Recovered		Dead		Transferred		Total
			Female	Male	Female	Male	Female	Male	
Age	01-20	9.52	5	13	0	1	0	2	21
	21-40	5.40	44	68	5	5	15	25	162
	41-60	5.82	89	120	15	21	24	35	304
	61-80	4.99	81	124	30	43	50	62	390
	81-100	4.91	30	29	21	20	18	20	138
Symptoms	At least one	5.18	91	114	46	47	31	44	373
	No Symptoms	5.50	158	240	25	43	76	100	642
Underlying Disease	At least one	5.51	170	248	33	53	81	102	687
	No Disease	5.12	79	106	38	37	26	42	328
Total		5.38	249	354	71	90	107	144	1015

*** HD= Hospitalized Duration: The number of days in which the patient should be under treatment in hospital**

3. Modeling Procedures and Analysis

In order to utilize the analytical procedures, three main regression models can be developed for the patients resulted in recovery, transferred, death, and total. The committed variables of "Age", patient symptom shown as "Symptom", and finally suffering from underlying diseases, indicated by "Disease" are considered in the regression modeling process. For the last two variables, if at least one symptom or disease appears, it is set to one otherwise 0.

The attractive significance level of regression modeling sets to 95%, so if the P-value is less than 0.05, means that regression modeling is significant (Pourhossein & Mahmoudabadi, 2019), otherwise, a linear regression model cannot interpret the relation between hospitalized duration and age, symptoms or underlying disease. Since the modeling procedure applies in four categories of recovered, transferred, dead, and total patients, they are developed separately. For each category, two regression models have been developed. At the initial step, all predictors have been entered in the modeling procedure whereas at the final step the most reliable predictor is selected and the model re-developed based on that. The number of patients, who had a chance to be recovered from Covid-19, was 603 out of 1015 of the sample investigated in this study.

The initial regression model has been developed based on three key predictors of Age, Symptoms, and Underlying disease, and results tabulated in Table 2. As shown, the initial model cannot interpret the relationship between hospitalized duration and the predictors where the significant F is greater than 0.05. It means the hospitalized duration for the recovered patients contracted to Covid-19 does not depend on Age, Symptoms, and underlying diseases. The number of patients, who transferred to another hospital or discharged, was 251 out of 1015 of the sample investigated in this study. Such as what has been done for recoveries, the initial regression modeling results are tabulated in the table, which reveals the hospitalized duration for the transferred or discharged patients contracted to Covid-19 does not depend on Age, Symptoms, and underlying diseases.

Table 2.The stats of prediction model for the patients resulted in Recovery and Transferred

Clearance	Model	Specification		Predictors	Coefficients	t-Stat	P-value
Recovered	Initial	Regression (SS*)	137.559	Intercept	6.2782	3.9469	0.0001
		Residual (SS*)	18487.008	Age	-0.0107	-0.8587	0.3908
N = 603		F-Stat	1.486	Symptom	-0.4761	-0.3722	0.7098
		Significant F	0.217	Disease	0.4418	0.3352	0.7376
Transferred	Initial	Regression (SS*)	31.913	Intercept	4.2681	2.4707	0.0142
		Residual (SS*)	2972.963	Age	-0.0141	-1.1300	0.2596
N = 251		F-Stat	0.884	Symptom	-0.6801	-0.5014	0.6166
		Significant F	0.450	Disease	-0.1259	-0.0900	0.9283

***SS = Sum of Square Errors**

The number of patients, who transferred to another hospital or discharged following their personal satisfaction, is 161 out of 1015 persons. The initial regression modeling results are tabulated in Table 3, where it can be concluded that the "Age" has a significant effect on the hospitalized duration of the dead patients who were contracted to Covid-19. A remarkable conclusion is that the existing relation is in a reverse direction, which means that the old patients had no chance to able to resist the virus and lost their lives in short times comparing to the other groups. In this case, the final regression model has been developed based on the "Age" with a significant F of 0.001, and the model formulated by equation (2).

$$\boldsymbol{Deaths}: Hospitalized - Duration(Day) = 15.047 - 0.1137 \times Age(Year) \quad (2)$$

The modeling procedure applied for the patients that resulted in death has been also repeated for all patients and results tabulated in the above-mentioned table. The same conclusion has been obtained where the initial modeling reveals that there is a significant relation between "Age" and "Hospitalized Duration" and equation (3) represents their reverse relationship.

$$\boldsymbol{Total}: Hospitalized - Duration(Day) = 6.708 - 0.022 \times Age(Year) \quad (3)$$

Table 3: The stats of prediction models for the patients resulted in dead

Clearance	Model	Specification		Predictors	Coefficients	t-Stat	P-value
Death	Initial	Regression (SS*)	562.250	Intercept	15.6372	5.0202	0.0000
		Residual (SS*)	8310.222	Age	-0.1132	-3.2033	0.0016
N = 161		F-Stat	3.541	Symptom	-0.5527	-0.2855	0.7756
		Significant F	0.016	Disease	-0.5677	-0.2973	0.7666
Death	Final	Regression (SS*)	557.227	Intercept	15.0472	6.1727	0.0000
		Residual (SS*)	8315.245	Age	-0.1137	-3.2642	0.0013
		F-Stat	10.655				
		Significant F	0.001				
Total	Initial	Regression (SS*)	189.835	Intercept	6.4007	5.6976	0.0000
		Residual (SS*)	32318.313	Age	-0.0214	-2.2029	0.0278
N = 1015		F-Stat	1.980	Symptom	0.0523	0.0599	0.9522
		Significant F	0.115	Disease	0.3529	0.3915	0.6955
Total	Final	Regression (SS*)	169.429	Intercept	6.7081	11.1540	0.0000
		Residual (SS*)	32338.719	Age	-0.0222	-2.3038	0.0214
		F-Stat	5.307				
		Significant F	0.021				

***SS = Sum of Square Errors**

4. Summary and Conclusion

Since the estimation of the hospitalized duration of patients' treatment is an important issue for health authorities, the statistical regression modeling procedure is utilized to predict it based on the personal characteristic of age, symptoms, and underlying diseases of the patients contracted to Covid-19. The research has been conducted in the Iranian northern province of Guilan, where the personal data of 1015 patients were available. The regression analysis revealed that the hospitalized duration of the patients does not depend on patients' symptoms and their underlying diseases but it depends on the age of all patients and they were ended to death.

In terms of application, the results support health authorities to manage patients' admission procedures where they are supposed to consider that the patient's age has reverse and significant effects on hospitalized duration. Researchers interested in working in this field are recommended to focus more on specific personal characteristics such as lifestyle, food, place of birth, and the other factors contributing to the immunity of the human system.

References

Alexopoulos, E. C. (2010). Introduction to multivariate regression analysis.Hippokratia, 14 (Suppl 1), 23.)

Almqvist, J., Granberg, T., Tzortzakakis, A., Klironomos, S., Kollia, E., Öhberg, C., & Ineichen, B. V. (2020). Neurological manifestations of coronavirus infections–a systematic review.Annals of clinical and translational neurology.

Di Mascio, D., Khalil, A., Saccone, G., Rizzo, G., Buca, D., Liberati, M., ...&D'Antonio, F. (2020). Outcome of Coronavirus spectrum infections (SARS, MERS, COVID 1-19) during pregnancy: a systematic review and meta-analysis. American journal of obstetrics & gynecology MFM, 100107.

Emami, A., Javanmardi, F., Pirbonyeh, N., &Akbari, A. (2020). Prevalence of underlying diseases in hospitalized patients with COVID-19: a systematic review and meta-analysis. Archives of academic emergency medicine, 8(1).

Farzandipour, M., Sheikhtaheri, A., &Sadoughi, F. (2010). Effective factors on accuracy of principal diagnosis coding based on International Classification of Diseases, the 10th revision (ICD-10). International Journal of Information Management, 30(1), 78-84.

Hogan, H., Zipfel, R., Neuburger, J., Hutchings, A., Darzi, A., & Black, N. (2015).Avoidability of hospital deaths and association with hospital-wide mortality ratios: retrospective case record review and regression analysis. bmj, 351, h3239.

HosmerJr, D. W., Lemeshow, S., & May, S. (2011). Applied survival analysis: regression modeling of time-to-event data (Vol. 618).John Wiley & Sons.

Kowalski, L. P., Sanabria, A., Ridge, J. A., Ng, W. T., de Bree, R., Rinaldo, A., ...&Paleri, V. (2020). COVID-19 pandemic: effects and evidence-based recommendations for otolaryngology and head and neck surgery practice. Head & neck, 42(6), 1259-1267.

Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., &Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. International journal of antimicrobial agents, 105924.

Law, S., Leung, A. W., &Xu, C. (2020). Severe acute respiratory syndrome (SARS) and coronavirus disease-2019 (COVID-19): From causes to preventions in Hong Kong. International Journal of Infectious Diseases.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012).Introduction to linear regression analysis (Vol. 821).John Wiley & Sons.

Panay, B., Baloian, N., Pino, J. A., Peñafiel, S., Sanson, H., &Bersano, N. (2019).Predicting Health Care Costs Using Evidence Regression.In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 31, No. 1, p. 74).

Pandey, G., Chaudhary, P., Gupta, R., & Pal, S. (2020). SEIR and Regression Model based COVID-19 outbreak predictions in India. arXiv preprint arXiv:2004.00958.

Polaraju, K. and D. D. Prasad (2017). "Prediction of heart disease using multiple linear regression model." International Journal of Engineering Development and Research Development 5(4): 1419-1425.

Pourhossein, F. and Mahmoudabadi, A., Original Paper Smart Urban Street Advertising Pattern Using Internet of Things Based on Environmental and Traffic Conditions 2(2): 44-60.

Riascos, A. and N. Serna (2017).Predicting Annual Length-Of-Stay and its Impact on Health. Medical Informatics and Healthcare.

Sun, L., DePuy, G. W., & Evans, G. W. (2014). Multi-objective optimization models for patient allocation during a pandemic influenza outbreak. Computers & Operations Research, 51, 350-359.

Biographical Statements:

Dr. Abbas Mahmoudabadi, is director of the master program in Industrial Engineering at MehrAstan University, Guilan, Iran. He received his Ph.D. degree in Hazmat transport optimization in January 2014, followed by receiving Thesis Dissertation Award from IEOM Society in 2015. He has published more than 85 papers in industrial engineering, transportation, traffic safety, and e-commerce. In addition to effective cooperation with national and international agencies, he teaches the education courses on the above fields as well as more than 26 years of executive experiences on traffic, road safety planning, and public transportation in developing countries.

Ms. Maedeh Pourmirza Dogahi, is student in Information Technology and E-commerce at Mehrastan University, Guilan, Iran and is currently investigating the contributing factors of Covid-19 patients on hospital treatment. She is working as an IT expert in department of education, medical and research center of a hospital located in the northern Iranian province of Guilan to manage patients' registering data for Health Information System.