Advances in Informatics, Management and Technology in Healthcare
J. Mantas et al. (Eds.)

doi:10.3233/SHTI220713

Using Logistic Regression to Predict Long COVID Conditions in Chronic Patients

Adnan KULENOVIC[a,1] and
Azra LAGUMDZIJA-KULENOVIC[a]
[a] *Absolute Information Age, Inc. Toronto, Canada*

Abstract. Chronic diseases pose significant challenges to patients and healthcare systems, and the *COVID-19* pandemic has further deteriorated that situation. This paper presents a method for predicting selected *long COVID* conditions in chronic and multimorbidity patients. It produces a logistic regression model for each *long COVID* condition by examining electronic medical records (*EMR*s) of *COVID-19* patients and taking their chronic conditions as predictors. The models were developed and tested using the *Jumpstart EMR* database, provided in the *COVID-19 Research Environment* of *Hopkins University*, containing about 250,000 *EMR*s of the outpatient and ambulatory *COVID-19* patients across the US. They are illustrated by predictions of 20 prevalent acute and chronic *long-COVID* conditions in patients diagnosed with frequent *pre-COVID* chronic diseases. These models can aid in investigating *long COVID* impacts on various chronic patients, finding their underlying pathophysiology, and establishing guidelines for their treatment and prevention.

Keywords. *Long COVID*, chronic diseases, *EMR*, logistic regression, prediction.

1. Introduction

The objective of this study was to develop a set of models for the prediction of long-term *post-COVID* conditions in patients suffering from specific chronic diseases. The *long COVID,* also referred to as post-acute sequelae *SARS-CoV-2* infection *(PASC),* can be defined as the persistence of one or many conditions for more than three weeks after the acute *COVID-19* infection phase [1]. These conditions include persistent symptoms, organ dysfunction, and multisystem inflammatory syndrome. Potential contributors to *long COVID* complications, according to studies [2, 3], include injuries to one or several organs in the acute phase of *COVID-19*, persistence of *SARS-CoV-2* virus in certain tissues, re-activation of neurotrophic pathogens, *SARS-CoV-2* interactions with host-microbiome and virome, clotting and coagulation issues, and dysfunctional vagus nerve signalling. Study [4] further elaborated on *post-COVID* manifestations in each body system and described their pathophysiology and risk factors. Undoubtedly, these complications can seriously affect chronic patients whose overall medical conditions have been already compromised by synergic effects of dysfunction of multiple body systems and polypharmacy.

The next sections explain the models and the results of a sample study.

[1] Corresponding Author, Adnan Kulenovic, Absolute Information Age, Inc. Toronto, Canada; E-mail: adnan.kulenovic@abs-info-age.com.

2. Method

The method inputs are *EMRs of COVID-19* patients and a set of chronic *pre-COVID* conditions *(C),* and a set of *long COVID* conditions *(LC)*. Specific data used from the *EMR* database are the patients' age and the *ICD-10* codes and dates of their conditions. The result is a set of prediction models, one for each *long COVID* condition $lc_i \in LC$, $i=1..N$ and the set of chronic disease predictors $c_j \in C$, $j=1..M$. Each model is composed of a logistic regression formula and related statistical information.

The method steps are:

(1) Extract *EMRs* of *COVID-19* patients of a certain age having valid *ICD-10* condition diagnostic codes and dates. The presented study included patients older than 40, male and female. In addition, it included diagnoses recorded between *January 01, 2019*, and *December 1, 2021*, and *COVID-19* diagnoses between *December 1, 2019*, and *June 01, 2021*, which allowed at least six months to encounter *long COVID* conditions after the *COVID-19* diagnoses.

(2) Map the *EMR ICD-10* diagnostic codes into generic conditions, acute and chronic, according to their *ICD-10* and *Chronic Code Indicator (CCI)* classifications. Most frequent diagnostic codes are mapped to separate generic conditions; for example, *ICD-10* class *I00-I99 Diseases of Circulatory System* was split into *Hypertension (I10-I16)*, *Chronic Heart Diseases (I05-I09, I17-I52), Acute Heart Diseases (I40, I28-I33…)*, and *Other Circulatory Diseases.*

(3) Select patients' *long-COVID* conditions diagnosed after their *COVID-19* infection. According to the adopted definition, a *long COVID* condition should last three weeks or more. Also, *pre-COVID* acute and chronic conditions that lasted three months or more were excluded.

(4) Define the logistic regression variables. Assign a predictor binomial variable $cv_j \in CV$ for each *pre-COVID* condition $c_j \in C$, with a value of 1 for patients who had that condition before their *COVID-19* diagnoses, 0 otherwise. Similarly, define predicted binomial variables $lcv_i \in LCV$ that indicate the *odds* for developing each *long COVID* condition $lc_i \in LC$ for predictors $c_j \in C$.

(5) Form a logistic regression model for each *long COVID* condition. Use a statistical package to form a logistic regression model from extracted *EMRs*, one for each *long COVID* condition lc_i. Each model includes a binomial regression formula to calculate the odds of a *long COVID* condition lc_i over predictors $lcv_i \in LCV$. Apart from logistic regression coefficients, each model includes, for each pair lc_i and c_j, the odds ratios OR_{ij}, 95% confidence intervals CI_{ij}, and p-values P_{ij}. When $OR_{ij} > 1$, condition c_j is associated with a higher odds of condition lc_i. In addition, when $P_{ij} < 0.05$, the association is significant (the null hypothesis is true in less than 5% of cases). More about the interpretation of *OR*, *CI* and *p-values* can be found, for example, in [5, 6].

3. Results and Discussion

The models were tested on the *Healthjump* database in the *COVID-19 Research Platform of Hopkins University*. The platform provides anonymized *EMR* data on outpatient and ambulatory patients collected from practices across the *US.* The environment includes

various database management and statistical analysis tools, such as *Snowflake/ SQL* and *STATA*. The authors were authorized to use this platform and data for this research.

Table 1 summarizes the statistical data and prediction indicators for the most frequent acute and chronic *long COVID* conditions lc_i *(N=20)* in patients suffering from major *pre-COVID* chronic conditions c_j *(M=12).* 37,515 *EMRs* met the constraints from Steps 1-3 above, meaning that they had at least one of the selected chronic *pre-COVID* conditions and one or more *long COVID* ones. The *pre-COVID* chronic conditions c_j are listed in the first column, and their patient counts are next. The remaining columns show the *long COVID* conditions lc_i and their case counts. Each cell displays the odds ratio OR_{ij} and p-value P_{ij} for the association of lc_i and c_j. The cells are shaded when $OR_{ij}>1$ and $P_{ij}<0.05$, meaning that *pre-COVID* condition c_j is a significant predictor of a *long COVID* condition lc_i. In most cases, P_{ij} was found to have a value of 0.000, indicating a strong association between the predictor and predicted diseases. Confidence intervals CI_{ij} are not presented, as they correlate to P_{ij} [6]. Associations marked with ***x*** indicate preexisting chronic conditions, disregarded according to Step 3.

		Acute *Long COVID* Conditions										**Chronic *Long COVID* Conditions**									
Chronic *Pre COVID* Conditions		Dorsopathies	Soft Tissue Disorders	Joint Disorders	Digestive System Symptoms	Cough	Upper Respiratory Infections	Abnormal Breathing	Infectious Diseases	Malaise and Fatigue	Anemia	Hyperlipidemias	Anxiety Disorders	Obesity	Vitamin D Deficiency	Gastro Esophageal Diseases	Heart Diseases	Nervous System Diseases	Hypertension	Kidney Diseases	Diabetes Mellitus
	Case #	5077	4512	3737	3686	3629	3628	3392	2826	2682	1818	4778	4155	4152	3934	3922	3683	3353	3301	2282	1945
Hypertension	23237	1.2	0.9	1	0.9	1	1	1	0.9	1	1.3	1.2	0.9	1.3	0.9	1.1	1.7	1.2	x	1.8	1.4
Hyperlipidemias	21034	1.2	1.1	1.1	1.1	1.2	1	1.2	1	1.2	1.2	x	1	1	1.3	1.1	1.2	1	0.9	1.2	1
Diabetes Mellitus	12073	0.9	1.1	1	1.2	0.9	0.8	1	1.3	0.8	1.2	1.5	0.8	1.3	0.9	0.9	1.2	1.3	2	1.5	x
Gastro Esophageal Diseases	9795	1.4	1.2	1.1	1.5	1.3	1.2	1.2	1.2	1.1	1.3	0.9	1.2	0.9	1	x	1.1	1	0.9	1	0.9
Obesity	9452	1.2	1.2	1.2	1.1	1.2	1.2	1.4	1.2	1	1.1	1	1.1	x	1.3	1	0.9	1	1	0.9	1.6
Anxiety Disorders	8944	1.4	1.2	1.2	1.3	1.2	1.5	1.1	1	1.3	1.1	0.9	x	1	1.2	1.3	0.8	1	0.8	0.7	0.7
Upper Respiratory Diseases	8921	1.2	1.3	1.1	1.2	1.6	2.2	1	1.2	1.2	1	0.9	1	1	1	1.2	0.8	0.9	0.7	0.7	0.8
Athropathies	8523	1.3	1.4	1.6	1	0.9	1	1	1.1	1.2	1.1	0.8	0.8	0.9	0.9	1	1.2	1.1	1	1.2	0.9
Depressive Disorders	8289	1.2	1.2	1.1	1	1	1.1	1	1.2	1.2	1.1	0.7	1.7	1	0.9	1.1	0.9	1.1	0.8	0.9	0.9
Vitamin D Deficiency	7906	1.2	1.2	1.2	1.3	1.1	1.1	1.2	1.1	1.5	1.7	1	1	1	x	1.1	0.9	0.9	0.7	1	0.9
Heart Diseases	7639	1	0.9	0.9	1.2	1	0.8	1.6	1	1.3	1.5	0.9	0.8	0.8	0.9	1.1	x	1.3	1.9	1.7	1.2
Thyroid Disorders	7240	1	1.2	0.9	1.1	1	1.1	1.1	0.9	1.1	1.3	0.9	1	0.9	1.2	1	1.2	1.1	0.9	1	0.8

Table 1. Shaded cells indicate that a *pre-COVID* condition is a significant predictor of a *long COVID* condition as its logistic regression coefficient has *OR >1* and *p-value < 0.05.*

There are several observations of these results to be discussed here. First, the acute *long COVID Soft Tissue Disorders* were common sequelae of the majority of examined chronic *pre-COVID* conditions (10 of 12), along with *Dorsopathias* (9), *Malaise and Fatigue* (9) and *Anemia* (9). On the other hand, strong predictors of acute *long COVID* conditions are *Gastro-Esophageal Diseases* (all 10), *Hyperlipidemia* (8), *Vitamin D Deficiency* (8), *Anxiety Disorders* (8), *Obesity* (7), and *Upper Respiratory Diseases* (7).

Second, regarding the chronic *long COVID* conditions, preexisting *Hypertension* has a strong association with *long COVID Heart Diseases (OR=1.7), Kidney Diseases (1.8),*

Hyperlipidemia (1.2) and *Obesity (1.3).* There are high odds that the *Diabetes Mellitus* patients could develop *post-COVID Hypertension (OR=2), Hyperlipidemia (1.5), Kidney Diseases (1.5), Obesity (1.3)* and *Nervous System Diseases (1.3).* These results are consistent with the study [7] on *long COVID* complications in diabetic patients. *Heart Diseases* can lead to a higher incidence of *long COVID Hypertension (1.9)*, *Kidney Diseases (1.7)* and *Nervous System Diseases (1.3).* These results suggest that patients suffering from frequent multimorbidities, such as *Hypertension*, *Diabetes Mellitus* and *Heart Diseases,* could be at increased risk for severe *long COVID* complications.

The above results align with the findings of some meta-studies on *long COVID*. For example, a meta-study [8] examined the results of 25 observational studies. The most frequent *long COVID* sequelae were chest pain, fatigue, dyspnea, and cough, which correspond to acute *Cough*, *Abnormal Breathing* and *Malaise and Fatigue* reported in Table 1. However, according to that table, this list should be further extended with *Dosropathias, Soft Tissue Disorders* and the above chronic *long COVID* conditions frequently reported in the examined *EMRs*.

4. Conclusions

To the best of our knowledge, the originality of this method is in finding direct associations between the preexisting chronic diseases and their *long COVID* manifestations by using the corresponding prediction models based on logistic regression.

The presented findings can be a background for various studies investigating the underlying pathophysiology of *long COVID* conditions in chronic and multimorbidity patients. Furthermore, the logistic regression formulas can be directly used to predict the odds of specific *long COVID* conditions in these patients and thus help doctors apply proper on-time treatments and preventive measures to mitigate their possible consequences.

References

[1] Amenta EM, Spallone A, Rodriguez-Barradas MC, El Sahly HM, Atmar RL, Kulkarni PA. Postacute COVID-19: An Overview and Approach to Classification. Open Forum Infect Dis. 2020 October 21;7(12).

[2] Proal AD, VanElzakker MB. Long COVID or Post-acute Sequelae of COVID-19 (PASC): An Overview of Biological Factors That May Contribute to Persistent Symptoms. Front Microbiol. 2021 June 23; 12.

[3] Raveendran AV, Jayadevan R, Sashidharan S. Long COVID: An overview. Diabetes Metab Syndr. 2021 May-Jun;15(3):869-875.

[4] Pavli A, Theodoridou M, Maltezou HC. Post-COVID Syndrome: Incidence, Clinical Spectrum, and Challenges for Primary Healthcare Professionals. Arch Med Res. 2021 Aug;52(6):575-581.

[5] Szumilas M. Explaining odds ratios. Journal of the Canadian Academy of Child and Adolescent Psychiatry, 2010;19(3), 227–229.

[6] du Prel JB, Hommel G, Röhrig B, Blettner M. Confidence interval or p-value?: part 4 of a series on evaluation of scientific publications. Dtsch Arztebl Int. 2009 May;106(19):335-9.

[7] Hayden MR. An Immediate and Long-Term Complication of COVID-19 May Be Type 2 Diabetes Mellitus: The Central Role of β-Cell Dysfunction, Apoptosis and Exploration of Possible Mechanisms. Cells. 2020 Nov 13;9(11):2475.

[8] Cabrera Martimbianco AL, Pacheco RL, Bagattini ÂM, Riera R. Frequency, signs and symptoms, and criteria adopted for long COVID-19: A systematic review. Int J Clin Pract. 2021 Oct;75(10).