Mevcut sayıya ait içindekiler listesine DergiPark üzerinden ulaşılabilir

Selçuk Üniversitesi Fen Fakültesi Fen Dergisi

Dergi web sayfası: dergipark.org.tr/tr/pub/sufefd

Research Article

Analysis of Symptoms and Demographic Characteristics in Diagnosis of COVID-19 by Logistic Regression Model

Caner Tanış[a,1*]

[a] *Department of Statistics, Çankırı Karatekin University, Çankırı, 18100, Turkey, ror.org/011y7xt38*

ARTICLE INFO

Article History
Received 1 August 2023
Revised 17 September 2023
Accepted 24 October 2023

Keywords
Binary logistic regression
COVID-19
Infectious disease
Symptoms
Odds ratio

ABSTRACT

The new coronavirus COVID-19 is an infectious disease that started spreading globally in December 2019. Some symptoms are known to give clues as to whether the COVID-19 virus is infected. Therefore, the main purpose of this paper was to determine specific symptoms related to COVID-19 for the rapid diagnosis of COVID-19 cases. The data set consists of 25985 individuals including PCR results, 2 demographic properties (age, gender), and 5 symptoms such as headache, shortness of breath, sore throat, fever, and cough is considered in this study. We analyzed the relationship between these covariates and PCR results by binary logistic regression model. A total of 16405 (63.1%) individuals having to positive PCR results were included in this study. The research population was divided into two age groups (<60 and ≥60). The findings regarding the symptoms observed in COVID-19 patients can be listed as follows: Headache (25.8%), shortness of breath (2.2%), sore throat (11.2%), fever (16.3%), and cough (26.2%). The findings of binary logistic regression analysis show that any individual in the elder group has more probability of a positive PCR result approximately 1.6 times (odds ratio [OR]: 1.681), 95% confidence interval [CI]: 1.535-1.840). Also, an individual with symptoms of headache is approximately %7 more likely to have a positive PCR result than a nonexistent one (OR: 1.068, CI: 1.006-1.135).

Araştırma Makalesi

COVID-19 Tanısında Semptomların ve Demografik Özelliklerin Lojistik Regresyon Modeli ile Analizi

MAKALE BİLGİSİ

Makale Geçmişi
Geliş 1 Ağustos 2023
Revizyon 17 Eylül 2023
Kabul 24 Ekim 2023

Anahtar Kelimeler
İkili lojistik regresyon
COVID-19
Bulaşıcı hastalık
Semptomlar
Odds oranı

ÖZ

Yeni koronavirüs COVID-19, Aralık 2019'da küresel olarak yayılmaya başlayan bulaşıcı bir hastalıktır. Bazı semptomların COVID-19 virüsünün enfekte olup olmadığına dair ipuçları verdiği bilinmektedir. Bu nedenle, bu makalenin temel amacı COVID-19 vakalarının hızlı teşhisi için COVID-19 ile ilgili spesifik semptomları belirlemektir. PCR sonuçları, 2 demografik özellik (yaş, cinsiyet) ve baş ağrısı, nefes darlığı, boğaz ağrısı, ateş ve öksürük gibi 5 semptomu içeren 25985 kişiden oluşan veri seti bu çalışmada dikkate alınmıştır. Bu ortak değişkenler ile PCR sonuçları arasındaki ilişki ikili lojistik regresyon modeli ile analiz edilmiştir. PCR sonucu pozitif olan toplam 16405 (%63,1) birey bu çalışmaya dahil edilmiştir. Araştırma popülasyonu iki yaş grubuna ayrılmıştır (<60 ve ≥60). COVID-19 hastalarında gözlenen semptomlara ilişkin bulgular şu şekilde sıralanabilir: Baş ağrısı (%25,8), nefes darlığı (%2,2), boğaz ağrısı (%11,2), ateş (%16,3) ve öksürük (%26,2). İkili lojistik regresyon analizi bulgularına göre, yaşlı gruptaki herhangi bir bireyin pozitif PCR sonucu alma olasılığı yaklaşık 1,6 kat daha fazladır (odds oranı [OR]: 1,681), %95 güven aralığı [CI]: 1.535-1.840). Ayrıca, baş ağrısı semptomları olan bir bireyin pozitif PCR sonucuna sahip olma olasılığı olmayanlara göre yaklaşık %7 daha fazladır (OR: 1.068, CI: 1.006-1.135).

* Sorumlu Yazar
E-posta adresi: caner.tanis@gmail.com (C. Tanış)
[1] ORCID: 0000-0003-0090-1661
Doi: 10.35238/sufefd.1335965
E-ISSN: 2458-9411

Atıf / Cite as

Tanış, Caner. “Analysis of Symptoms and Demographic Characteristics in Diagnosis of COVID-19 by Logistic Regression Model”. *Selçuk Üniversitesi Fen Fakültesi Fen Dergisi* 50 (1) 2024, 1-5, 10.35238/sufefd.1335965

Makale Bilgisi Article Information

Makale Türü **Article Type**
Araştırma Makalesi Research Article
Geliş Tarihi **Date Received**
1 Ağustos 2023 1 August 2023
Revizyon Tarihi **Date Revised**
17 Eylül 2023 17 September 2023
Kabul Tarihi **Date Accepted**
24 Ekim 2023 24 October 2023
Yayım Tarihi **Date Published**
24 Nisan 2024 24 April 2024
Değerlendirme **Review Process**
İki Dış Hakem, Çift Taraflı Körleme Two External Reviewers, Double-Blind Peer Review
Etik Beyan **Ethical Statement**
Bu çalışmanın hazırlanma sürecinde bilimsel ve etik ilkelere uyulduğu ve yararlanılan tüm çalışmaların kaynakçada belirtildiği beyan olunur (C. Tanış). It is declared that scientific and ethical principles have been followed while carrying out and writing this study and that all the sources used have been properly cited (C. Tanış).
İntihal Kontrolü **Plagiarism Check**
Bu makale, iTenticate yazılımı ile taranmış ve intihal tespit edilmemiştir. This article has been scanned with iTenticate software and no plagiarism detected.
Çıkar Çatışması **Conflict of Interest**
Yazar, bu makalede bildirilen çalışmayı etkiliyor gibi görünebilecek bilinen hiçbir rakip mali çıkarları veya kişisel ilişkileri olmadığını beyan eder. The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Finansman **Funding**
Yazar, anonimleştirilmiş Kovid-19 hasta kayıtlarına kamu erişimi sağladığı için İsrail Sağlık Bakanlığı'na ve Kovid-19 hastalarının verilerinin elde edilmesindeki yardımlarından dolayı Prof. Dr. David Gurwitz'e müteşekkirdir. The author is thankful to the Israel Ministry of Health for providing public access to anonymized Covid-19 patient records and to Prof. Dr. David Gurwitz for his help in obtaining data of Covid-19 patients.

1. Introduction

The new coronavirus (COVID-19) appeared in Wuhan, China in December 2019 and rapidly exceeded its borders and affected the whole world. The World Health Organization (WHO) officially reported that COVID-19 is a global pandemic on March 11, 2020 (WHO, 2020). As of February 21, 2023, it was globally announced that there have been 757,264,511 confirmed COVID-19 cases with 6,850,594 deaths by the WHO (WHO, 2023). The rapid spread of COVID-19 has devastated the health system even in developed countries. Rapid investigation and identification of epidemiologic features of new infectious diseases are critical for reducing transmission and effective treatment management (Ki, 2020). Early diagnosis of the disease is very important to combat this epidemic affecting the world. In this regard, some clinical symptoms have helped in the early diagnosis of COVID-19. The main symptoms of COVID- 19, such as cough, fever, and dyspnea are important in the early diagnosis of COVID-19 disease (Alimohamadi et al., 2020). Identifying the clinical symptoms of COVID-19 in children is crucial in terms of accurately predicting the course of the disease (Tang et al., 2021). Many scientific studies have been conducted to specify the role of these symptoms in the diagnosis of COVID-19. (Ki, 2020) examined the epidemiologic properties of early COVID-19 cases in Korea. Choi and Ki (2020) designed a mathematical model to determine the effects of social distancing on the spread of COVID-19 in Korea. Yang et al. (2020) provided a systematic analysis of the prevalence of some symptoms in COVID-19 patients. Zhou et al. (2020) estimate the fatality of COVID-19 patients via a logistic regression model. Guan et al. (2020) examined the clinical characteristics of 1,099 COVID-19 patients in China. In another Chinese research, Xu et al. (2020) provided clinical findings of characteristics in COVID-19 patients. Feng et al. (2020) studied the early prediction of illness progression in COVID-19 patients with some clinical characteristics. Yupari-Azabache et al. (2021) analyzed the mortality risk factors in COVID-19 patients by using the logistic regression model. Hills and Eraso (2021) discussed the reasons for non-adherence the social distancing during the COVID-19 pandemic using a logistic regression analysis. Sonoda et al. (2021) focused on the determination of common symptoms in COVID-19 patients from Japan. Xiong et al. (2020)studied on the estimation of the fatality rates of COVID-19 patients from California using a logistic regression. In this regard, they considered some factors such as gender, race, and age to estimate the fatality rates of COVID-19 patients. Fleitas et al. (2020) analyzed the role of clinical symptoms in COVID-19 patients from Argentina via a logistic regression model. They examined the roles of 23 different symptoms in discrimination of being with COVID-19 positive. Liu et al. (2022) analyzed the clinical properties and risk factors in COVID-19 patients from Wuhan, China by using the logistic regression model.

2. Materials and Methods

The research population consisted of 25685 individuals with 16405 (63.1%) COVID-19 positive and 9580 (36.9%) COVID-19 negative. The research data is open-access data released to the public by the Israel Ministry of Health (Health, 2021). This anonymized dataset was constructed on July 29, 2020. This data is constantly updated. The last update date is March 7, 2021 (Health, 2021) for the used data set in this paper. These data were collected December 11-31 2020 in Israel. The Binary logistic regression model was designed to determine which symptoms or demographic variables affect the being COVID-19 positive. In this regard, the dependent variable is considered PCR result (negative = 0, positive = 1) and gender, age, cough, fever, shortness of breath, and sore throat are chosen as independent covariates in the binary logistic regression model. A two-sided α of less than 0.05 was considered statistically significant. The demographic findings were expressed as frequency (%).

3. Results and Discussion

The epidemiological characteristics of a total of 25685 individuals in this research are presented in Table 1. It is seen that the < 60 group (n = 23347(89.8%)) and the ≥ 60 group (n = 2638(10.2%)) are presented.

Table 1. The descriptive statistics of covariates in regression models.

Covariate	Category	Frequency (%)
PCR result (Response)	Positive	16405 (63.1)
	Negative	9580 (36.9)
Gender	Male	11889 (45.8)
	Female	14096 (54.2)
Age	<60	23347 (89.8)
≥ 60 2638 (10.2)		
Headache	Absent	19398 (74.7)
	Present	6587 (25.3)
Shortness of breath	Absent	25449 (97.9)
	Present	536 (2.1)
Sore throat	Absent	23146 (89.1)
	Present	2839 (10.9)
Fever	Absent	21698 (83.5)
	Present	4287 (16.5)
Cough	Absent	19321 (74.4)
	Present	6664 (25.6)

Table 2 provides the results of the univariate logistic regression model. Age, headache, shortness of breath, sore throat, and cough were included in the multivariate analysis, as they were found to be statistically significant (p < 0.05) in Table 2.

Table 2. The results of the univariate logistic regression model.

	$\hat{\beta}$	SE $\left(\hat{\beta}\right)$	p-value	OR	95 % OR
Gender	0.021	0.026	0.411	1.021	0.971-1.074
Age	0.518	0.046	<0.001*	1.678	1.533-1.836
Headache	0.068	0.03	0.022*	1.07	1.01-1.135
Shortness of breath	0.225	0.094	0.016*	1.252	1.042-1.504
Sore throat	0.085	0.042	0.041*	1.089	1.004-1.181
Fever	-0.046	0.035	0.182	0.955	0.893-1.022
Cough	0.085	0.03	0.004*	1.089	1.028-1.154

Table 3 provides the results of the multivariate logistic regression model. According to Table 2, it is seen that age and headaches are significant in the multivariate logistic regression model (p < 0.05). Those older than 60 years of age were 1.681 times more likely to be PCR positive than those younger than 60 years old (OR: 1.681, [1.535−1.840]). Those with headache symptoms are 1.068 times more likely to be PCR positive than those without headache symptoms (OR:

1.068[1.006−1.135]). In other words, older age and the presence of headache are statistically significant in determining the PCR test result and are variables that increase the probability of being COVID-19 positive.

Table 3. The results of the multivariate logistic regression model.

	$\hat{\beta}$	SE $(\hat{\beta})$	p-value	OR	95 % OR
Age					
<60 (Ref)					
>60	0.519	0.046	<0.001*	1.681	1.535-1.840
Headache					
Absent (Ref)					
Present	0.066	0.031	0.032*	1.068	1.006-1.135
Shortness of breath					
Absent (Ref)					
Present	0.166	0.095	0.080	1.180	0.980-1.421
Sore throat					
Absent (Ref)					
Present	0.062	0.043	0.153	1.064	0.977-1.157
Cough					
Absent (Ref)					
Present	0.039	0.031	0.207	1.039	0.979-1.104
Constant	0.453	0.017	<0.001*	1.572	

*: It means statistical significance (p<0.05).

Table 4. Results of goodness of fit test for the multivariate logistic regression model.

Tests of goodness of fit	Test	df	P
Omnibus Tests of Model Coefficients	151.038	5	<0.001
Hosmer and Lemeshow Test	4.085	3	0.252

From Table 4, the results of the Hosmer-Lemeshow test show that the logistic regression model fits the data (p=0.252>0.05). According to omnibus tests of model coefficients, the coefficients of covariates are statistically significant in the multivariate logistic regression model, and the model formed by the explanatory variables is statistically significant (p<0.05).

This study is designed to determine which symptoms or characteristic properties have an impact on positive PCR results. Considering the findings of previous studies for similar purposes, the results of our study are similar to the results of many previous studies. Yang et al. (2020) reported the most observed clinical symptoms can be listed as follows: fever (91.3%), cough (67.7%), fatigue (51%), and dyspnea (30.4%). It is observed that the most common two symptoms are fever and cough in (Alimohamadi et al., 2020; Guan et al., 2020; Liu et al., 2022). [16] found that headache (OR:1.71) significantly correlated with being COVID-19 positive in the investigation including 48748 individuals less than 56 years old. Thus, they determined that the headache is a factor that increases the probability of being COVID-19 positive. Fleitas et al. (2020) observed that although the most common symptoms are cough and fever, there is no correlation between these symptoms and being COVID-19 positive. It is reported that older age is significant for COVID-19 disease severity while fever and cough are not significant in determining COVID-19 disease severity(Fleitas et al., 2020; Yupari-Azabache et al., 2021). The findings of Sonoda et al. (2021) showed that headache (OR: 3.31) is significantly associated with being COVID-19 positive. They mentioned that the presence of fever was not utility to determine being COVID-19 positive. Guan et al. (2020) observed that some characteristics such as gender, fever, cough, sore throat, and shortness of breath are not significant to determine COVID-19 positive while age groups are crucial in the diagnosis of COVID-19. It is emphasized that age is an important factor in the risk of death due to COVID-19 (Fleitas et al., 2020; Xiong et al., 2020). Yupari-Azabache et al. (2021) mentioned that an older person is a risk factor mortality of hospitalized patients infected with COVID -19 (OR: 1.11). The findings of Feng et al. (2020), some characteristics such as gender, fever, cough, and shortness of breath are not statistically significant to estimate of illness progression in COVID-19 patients. According to our findings, the most common symptoms are cough and headache in COVID-19 patients. Headache (OR: 1.681) and age (OR: 1.068) were found statistically significant in the multivariate logistic regression model while gender, cough, fever, shortness of breath, and sore throat were found not significant in determining the PCR results.

4. Conclusions

This study includes many results that we think will contribute to the literature. First of all, headache and age were found to be significant in the multivariate logistic regression analysis, and it was concluded that these two variables were effective in determining the PCR test result. The presence of a headache increases the probability of being COVID-19 positive by 1.068 times, while someone ≥60 age is 1.681 times more likely to be COVID-19 positive than someone <60 age. As a result, we provide new important findings of some symptoms and characteristics in the literature. The conclusions of this research have supported previous findings. We consider that the results of this study will guide future studies. One of the important advantages of this study is that it was designed with a big sample size (n = 25985). The study has limitations. Firstly, the Covid-19 data sets are anonymous and open Access (Health, 2021). Secondly, the sample size is limited and an accurate assessment of the diagnosis of Covid-19 patients is not accurate.

Acknowledgements

The author is thankful to the Israel Ministry of Health for providing public access to anonymized Covid-19 patient records and to Prof. Dr. David Gurwitz for his help in obtaining data of Covid-19 patients.

References

Alimohamadi, Y., Sepandi, M., Taghdir, M. and Hosamirudsari, H., 2020, Determine the most common clinical symptoms in COVID-19 patients: a systematic review and meta-analysis, *Journal of preventive medicine and hygiene*, 61 (3), E304.

Choi, S. and Ki, M., 2020, Analyzing the effects of social distancing on the COVID-19 pandemic in Korea using mathematical modeling, *Epidemiology and health*, 42.

Feng, Z., Yu, Q., Yao, S., Luo, L., Zhou, W., Mao, X., Li, J., Duan, J., Yan, Z. and Yang, M., 2020, Early prediction of disease progression in COVID-19 pneumonia patients with chest CT and clinical characteristics, *Nature communications*, 11 (1), 4968.

Fleitas, P. E., Paz, J. A., Simoy, M. I., Vargas, C., Cimino, R. O., Krolewiecki, A. J. and Aparicio, J. P., 2020, Understanding

the value of clinical symptoms of COVID-19. A logistic regression model, *MedRxiv*, 2020.2010. 2007.20207019.

Guan, W.-j., Ni, Z.-y., Hu, Y., Liang, W.-h., Ou, C.-q., He, J.-x., Liu, L., Shan, H., Lei, C.-l. and Hui, D. S., 2020, Clinical characteristics of 2019 novel coronavirus infection in China, *MedRxiv*.

Health, I. M. o., 2021, https://data.gov.il/dataset/covid-19/, [March 7].

Hills, S. and Eraso, Y., 2021, Factors associated with non-adherence to social distancing rules during the COVID-19 pandemic: a logistic regression analysis, *BMC Public Health*, 21 (1), 1-25.

Ki, M., 2020, Epidemiologic characteristics of early cases with 2019 novel coronavirus (2019-nCoV) disease in Korea, *Epidemiology and health*, 42.

Liu, X.-q., Xue, S., Xu, J.-b., Ge, H., Mao, Q., Xu, X.-h. and Jiang, H.-d., 2022, Clinical characteristics and related risk factors of disease severity in 101 COVID-19 patients hospitalized in Wuhan, China, *Acta Pharmacologica Sinica*, 43 (1), 64-75.

Sonoda, S., Kuramochi, J., Matsuyama, Y., Miyazaki, Y. and Fujiwara, T., 2021, Validity of clinical symptoms score to discriminate patients with COVID-19 from common cold out-patients in general practitioner clinics in Japan, *Journal of clinical medicine*, 10 (4), 854.

Tang, Z., Li, M., Chen, W., Ran, X., Li, H. and Chen, Z., 2021, Clinical symptoms of COVID-19 pneumonia in children: A protocol for systematic review and meta-analysis, *Medicine*, 100 (1).

WHO, 2020, Director-General's Opening Remarks at the Media Briefing on Covid 19 - 11 March 2020.

WHO, 2023, https://covid19.who.int/, [February 22].

Xiong, D., Zhang, L., Watson, G. L., Sundin, P., Bufford, T., Zoller, J. A., Shamshoian, J., Suchard, M. A. and Ramirez, C. M., 2020, Pseudo-likelihood based logistic regression for estimating COVID-19 infection and case fatality rates by gender, race, and age in California, *Epidemics*, 33, 100418.

Xu, Y., Xu, Z., Liu, X., Cai, L., Zheng, H., Huang, Y., Zhou, L., Huang, L., Lin, Y. and Deng, L., 2020, Clinical findings in critically ill patients infected with SARS-CoV-2 in Guangdong Province, China: A multi-center, retrospective, observational study, *MedRxiv*, 2020.2003. 2003.20030668.

Yang, J., Zheng, Y., Gou, X., Pu, K., Chen, Z., Guo, Q., Ji, R., Wang, H., Wang, Y. and Zhou, Y., 2020, Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis, *International journal of infectious diseases*, 94, 91-95.

Yupari-Azabache, I., Bardales-Aguirre, L., Rodriguez-Azabache, J., Barros-Sevillano, J. S. and Rodríguez-Diaz, A., 2021, COVID-19 mortality risk factors in hospitalized patients: A logistic regression model, *Revista de la Facultad de Medicina Humana*, 21 (1), 19-27.

Zhou, F., Chen, T. and Lei, B., 2020, Do not forget interaction: Predicting fatality of COVID-19 patients using logistic regression, *arXiv preprint arXiv:2006.16942*.