RESEARCH ARTICLE

Using machine learning models to investigate the relationship between corporeal workload and clinical and epidemiological features of patients infected with COVID-19 in Egypt

Ahmed Ali Torad[1], Fayiz El-Shamy[2], Ahmed Mahmoud Kadry[3], Zeinab Sami Ahmed[4]

Abstract

Objective: To explore if clinical and epidemiological features of patients positive for coronavirus disease-2019 are affected by somatic work stress.

Method: The retrospective study was conducted at Kafrelsheik University Hospital, Egypt, and comprised data of patients admitted between April 1, to June 6, 2020, with confirmed coronavirus disease-2019 infection. Health records of healthy subjects who had come to the hospital as part of their routine check-up were also included for comparison, and the researchers were blinded during the gathering and analysis phase. Demographic features, vital signs, infection severity, somatic workload of the patients' jobs at admission, and detailed discharge profile was noted. The relationship between clinical features and somatic work stress was evaluated. Data was analysed using SPSS 26.

Results: Of the 1072 cases, 602(56.2%) were men and 470(43.8%) were women. The overall median age was 43 years (interquartile range: 29 years). The healthy group had 500 random subjects. There were significant differences in all vital signs between the patients and healthy controls ($p<0.05$). Among the patients, infection severity was higher in men, but it was not significant (p>0.05). The overall mortality was 69(6.4%); 46(4.3%) men and 23(2.2%) women. There was no significant association between gender and outcome ($p>0.05$). There were 816(76.11%) patients with low-intensity physical workload pre-infection, 136(12.68%) moderate and 120(11.19%) high. Infection severity was significantly high in the low-intensity group ($p<0.05$). However, the fate of the patients was not significantly associated with their pre-infection work profile ($p>0.05$).

Conclusion: Coronavirus disease-2019 significantly affected patients' vital signs, and infection severity was significantly associated with physical work stress. However, mortality and pre-infection somatic workload were not associated.

Keywords: Sulphur dioxide, COVID-19, Epidemiologists, Healthy volunteers, Polymerase, Demography.

DOI: 10.47391/JPMA.EGY-S4-48

Introduction

The coronavirus disease-2019 (COVID-) was initially described by the World Health Organisation (WHO) as an unidentified outbreak of viral pneumonia in Wuhan, China, in late December 2019.[1,2] The virus subsequently spread rapidly across the country and then to several countries, which caused a huge number of infections and massive social phobia.[3]

Like the rest of the world, COVID-19 represented for Africa one of the greatest health challenges. With more than 58 million children suffering from malnutrition, more young people were likely to be at risk in Africa than elsewhere in the world.[2]

With an incubation period of 1-14 days, mostly 3-7 days, the basic signs of the infection are fever, dry cough and myalgia.[4] Patients appear normal during the incubation phase, but can pass the infection to others even without symptoms.[5] The main transmission method for COVID-19 was through huge particles when people tossed or sneezed, and this could be confused with a common influenza pestilence.[6] Contact with the infected products and devices further spreads the infection.[2] Th resulting pneumonia was additionally irresistible.[7]

More data was found to be necessary about the characteristics of patients in Egypt with COVID-19, especially related to patients who were not seriously ill but whose condition deteriorated rather quickly.[8]

A sharp decrease in COVID-19 cases occurred by rapidly increasing testing, isolating the infected patients, and precise contact tracing.[7]

Documentation of patients at serious risk of being COVID-

[1]Department of Basic Science, Kafrelsheikh University, Egypt.
[2]Department of Physical Therapy for Women's Health, Kafrelsheikh University, Egypt. [3]Physical therapy department for integumentary disorders, Kafrelsheikh University, Egypt. [4]Department of Physical Therapy for Cardiovascular and Respiratory Disorder and Geriatrics, Cairo University, Egypt.
Corrospondience: Ahmed Ali Mohammed Torad
email: ahmed.ali.torad@gmail.com

19-positive (COVID-19+) may help optimise resource use and care.

After treatment isolation, numerous patients developed anxiety and sleep disruption. Anxiety as a mental burden can trigger physiological events that may enhance or decrease immunity.[9] There is growing evidence, for example, that lifestyle factors, adapting to everyday stress, and dietary behaviour are significant cofactors in the immune system.[10]

The current study was planned to explore if clinical and epidemiological features of COVID-19+ patients are affected by somatic work stress.

Materials and Methods

The retrospective study was conducted at Kafrelsheik University Hospital, Egypt, and comprised data of patients admitted between April 1, to June 6, 2020, with confirmed COVID-19 infection. The health facility was government-designated to accept COVID-19+ patients. After approval from the institutional ethics committees of Kafrelsheik University Hospital and Faculty of physical therapy-Kafrelsheik University (approval number: P.T/ BAS /1/2021/5, data was retrieved from the electronic health record (ERH) database. Data was related to patients infected with COVID-19, confirmed by reverse transcription polymerase chain reaction (RT-PCR).

Health records of healthy subjects who had come to the hospital as part of their routine check-up were also included for comparison, and the researchers were blinded during the gathering and analysis phase.

Demographic features, like age, gender, body mass index (BMI) and self-reported smoking status, were noted. Also noted were vital signs, such as systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate (HR), respiratory rate (RR), temperature (T), oxygen saturation (SpO2), and associated diseases present at the time of admission, like cancer, cardiac diseases, paralysis, chronic lung disease, renal insufficiency, liver deficit, pancreatic problems, morbid obesity, and respiratory circulation problem.

Infection severity was categorised as mild, moderate and severe, with mild = no care inside a critical care facility, moderate = intensive care unit (ICU) treatment, but no assistance required for mechanical breathing assistive devices, and severe = needed support for mechanical respirator.[8]

Medical records with negative COVID-19 PCR, comorbidities present on admission, or/and missed data were excluded.

Somatic workload of the patients' jobs at admission was classified into three categories as per the International standard classification of occupation-three categories of physical activity (ISCO-3CPA).[11,12] The categories were low, moderate and high, with low = low-intensity physical workload (LIPW), moderate = moderate-intensity physical workload (MIPW), and high = high-intensity physical workload (HIPW). Detailed discharge profile was noted. The outcome of each patient was recorded as either dead or alive. Python 3.9 was used for modelling analysis. Data was then further prepped. There were 114 missing values across the variables, and k-Nearest Neighbours (KNN) imputer was used which involved calculating the mean of the five values that were immediately adjacent to the missing value on both sides of the data.[13] After that, data was checked for outliers and winsorised to lessen the impact of extreme results[14]. Winsorisation included replacing extreme values with values that fell inside a particular percentile of the original distribution. All research variables were rescaled to recode values above the 95th percentile at the 95th percentile value.[14]

Patients were divided into three categories based on infection severity. The dataset was used to train classifiers on the model after the features were sorted. In order to classify data, models including Logistic Regression, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier, KNN Classifier, Multi-layer Perceptron (MLP) Classifier, Support Vector Classifier (SVC) model, Gaussian Naive Bayes model, and Bagging Classifier were utilised.[15]

To some extent, overfitting and selection bias were avoided by training each model using a 10-fold cross-validation procedure[15]. Each of the models was run 10,000 times after being randomly divided into the training set (90%) and the test set (10%). Area under the curve (AUC) of the receiver characteristic operator (ROC) and mean and median "Accuracy" were used to evaluate the classification models.

Data was analysed using SPSS 26. Continuous data was expressed mean ± standard deviation (SD), and categorical and nominal data as frequencies and percentages. The connection between categorical and nominal data was found using Pearson's chi-square test. Statistically significant variations in the continuous variables between the two groups were found using t-test. Age was utilised as a co-variate in a 3x2 mixed model analysis of covariance (MANCOVA). Prior to the primary analysis, Shapiro-Wilks test for normality was used with histograms to check all vital sign variables for normality assumptions between the groups. Exponential, power, arcsine, and logarithmic transformations were used for non-normally distributed data. The histograms and Shapiro-Wilks tests did not show any differences. Where necessary, a pair-wise comparison using Tukey's post-hoc analysis was performed with 0.05 as the threshold. All analyses underwent a Bonferroni correction.[16,17] The level of statistical significance was set as $p<0.05$.

The code for the current project is presented in the following GitHub link:

Dr AhmedTorad/COVID_19_and_Corporeal_workload (github.com)

Results

The modelling results showed that the Random Forest Classifier was the best one to predict the severity of the case based on vital signs with feature importance ranked as follows: SpO2: 0.23, RR: 0.13, age: 0.129, T: 0.127, SBP: 0.12, DBP: 0.117, HR:0.112, and occupation physical workload: 0.035. The mean accuracy was 0.7477 with **95%** confidence interval (CI) 0.737-0.759. The quartile accuracies were: Minimum: 0.69, Q1: 0.736, Q2: 0.747, Q3: 0.769, maximum: 0.791, interquartile range (IQR): 0.032, and AUC: 0.766. The sensitivity was 0.648, specificity of 0.517, negative log losses (the error) 0.62.

Of the 1072 cases, 602(56.2%) men and 470(43.8%) were women. The overall median age was 43 years (IQR: 29 years). The healthy group had 500 random subjects. There were significant differences in all vital signs between the patients and healthy controls ($p<0.05$) (Table 1).

Among the patients, infection severity was higher in men, but it was not significant (p>0.05). The overall mortality was 69(6.4%); 46(4.3%) men and 23(2.2%) women. There was no significant association between gender and outcome ($p>0.05$) (Table 2).

There were 816(76.11%) pre-infection LIPW patients, 136(12.68%) MIPW and 120(11.19%) HIPW. Infection severity was significantly high in the LIPW group ($p<0.05$). However, the fate of the patients was not significantly associated with their pre-infection work profile ($p>0.05$) (Table 3).

Except HR ($p<0.05$), no significant association were found

Table-1: Comparison of vital signs between COVID-19 patients and healthy subjects.

Vital signs	Covid-19+ patients (Mean ± SD)	Normal (Mean ± SD)	*t-value*	*p-value*
SBP	118.4±13.3	112±8	9.952	<0.001
DBP	76.2±8.97	68.5±8.2	16.2818	<0.001
RR	22±3.4	12±4	51.2704	<0.001
T	37.4±0.59	36.8±0.8	16.6859	<0.001
HR	85.9±13.16	80±6	9.5705	<0.001
Sp02	90.4±7.19	97±2	20.1621	<0.001

COVID-19: Coronavirus disease-2019, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, RR: Respiratory rate, T: Temperature, HR: Heart rate, P: Probability value.

Table-2: Comparison of COVID-19 infection between male and female patients.

Variable		Male n (%)	Female n (%)	Total	Chi-square	*p-value*
Severity cases	Mild	460 (42.9%)	359(33.5%)	819(76.4%)	1.187	0.552of
	Moderate	107(10.0%)	90(8.4%)	197(18.4%)		
	Severe	35(3.3%)	21(2.0%)	56(5.2%)		
The fate of COVID-19 patients	Recovered	556(51.9%)	447(41.7%)	1003(93.6%)	3.309	0.069
	Expired	46(4.3%)	23(2.2%)	69(6.4%)		

COVID-19: Coronavirus disease-2019,

Table-3: The relationship of physical workload with severity of infection and patient outcome.

Variable		ISCO-3CPA					
		LIPW n (%)	MIPW n (%)	HIPW n (%)	Total	Chi-square	*p-value*
Severity cases	Mild	628(58.6%)	111(10.4%)	80(7.5%)	819(76.4%)	19.31	<0.001
	Moderate	149(13.9%)	13(1.2%)	35(3.3%)	197(18.4%)		
	Sever	39(3.6%)	12(1.1%)	5(0.5%)	56(5.22%)		
The fate of COVID-19+ patients	Recovered	762(71.1%)	127(11.9%)	114(10.6%)	1003(93.56%)	0.463	0.790
	Expired	54(5.0%)	9(0.8%)	6(0.6%)	69(6.44%)		

ISCO-3CPA: International standard classification of occupation-three categories of physical activity, LIPW: Low-intensity physical workload, MIPW: Moderate-intensity physical workload, HIPW: High-intensity physical workload, COVID-19: Coronavirus disease-2019, Sig: Significant.

Table-4: Effect of physical workload on vital signs of COVID-19 patients.

Vital signs			ISCO-3CPA		
	LIPW (Mean ± SD)	MIPW (Mean ± SD)	HIPW (Mean ± SD)	F	*p-value*
SBP	118.81±13.4	116.58±13.77	117.5±11.9	1.812	0.16
DBP	76.43±8.86	75.47±9.71	75.73±8.83	0.815	0.44
RR	22.09±3.26	21.82±3.63	21.84±3.95	0.559	0.57
T	37.39±0.61	37.42±0.5	37.41±0.52	0.244	0.78
HR	83.55±11.88	88.03±16.54	100.08±5.93	95.946	<0.001
SpO2	90.52±6.91	90.08±8.72	90.11±7.11	0.291	0.75

COVID-19: Coronavirus disease-2019, ISCO-3CPA: International standard classification of occupation-three categories of physical activity, LIPW: Low-intensity physical workload, MIPW: Moderate-intensity physical workload, HIPW: High-intensity physical workload, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, RR: respiratory rate, T: Temperature, HR: Heart rate, SpO2: Oxygen saturation.

between somatic workload and vital signs (p>0.05) (Table 4).

Discussion

To the best of our knowledge, the current study is the first to evaluate the association between somatic workload and COVID-19+ patients' outcomes. Most of the patients were admitted from different areas in Egypt. Among the COVID-19+ patients, a spike was seen in infected individuals aged 43 years, and the percentage of infection was much higher in males than in female patients. Men also had higher severity and mortality rate than women. According to the Egyptian census 2020, there are 13.6% Egyptians aged <5 years, 11.1% aged 5-10 years, and only 19% aged >45 years,[18] which means that Egypt has high youth percentage. Also, the current results were in agreement with Sijia et al., who stated that 47.7% of cases were aged 13-45 years.[19] An analysis of 425 COVID-19+ patients revealed that 56% were males.[20] An analysis of 140 cases revealed that 50.7% of the cases were men.[21] Male COVID-19+ prevalence was 2.4 times that of females, and although males and females had the same vulnerability, men were more likely to die.[22]

In the current study's general population group, there was much better values related to vital signs compared to the COVID-19+ patients. This was in line with earlier direct or indirect findings related to association of COVID-19 with vital signs, including BP, RR, HR, SpO2 etc.[23-28]

An interesting finding of the current study is the detection of an increased rate of COVID-19 patients and odds of mortality in LIPW than in MIPW and HIPW groups. These may be due to the similarity of MIPW and HIPW with moderate exercises. It has been revealed that physical activity may reduce the occurrence and severity of infection.[29] Also, the antibodies or white blood cells (WBCs) circulate more quickly.[30] A link has been suggested between moderate activity and immunity boost.[29]

The current study has its limitations. In line with every reverse analysis of real-world information, the disparity in local practice and the fullness of the data collection constrained the analysis and decreased the study's sample size. EHR was used to retrieve data from the central data store. The study was conducted during a pandemic when standards might have been compromised. Additional efforts were made to ensure that missing data was missing on a spontaneous basis and not due to systematic installation or patient disparities. The current analysis is applicable to Kafrelsheik University Hospital, Egypt, and generalisation should be done with caution. Also, the sample size was not calculated which could have affected the power of the study.

Conclusion

COVID-19 significantly affected patients' vital signs, and infection severity was significantly associated with physical work stress. However, mortality and pre-infection somatic workload were not significantly associated.

Disclaimer: None.

Conflict of Interest: None.

Source of Funding: None.

Reference

1. Yang QX, Zhao TH, Sun CZ, Wu LM, Dai Q, Wang SD, et al. New thinking in the treatment of 2019 novel coronavirus pneumonia. Complement Ther Clin Pract 2020;39:101131. doi: 10.1016/j.ctcp.2020.101131.
2. Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, et al. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. Sci China Life Sci 2020;63:457-60. doi: 10.1007/s11427-020-1637-5.
3. Kang L, Ma S, Chen M, Yang J, Wang Y, Li R, et al. Impact on mental health and perceptions of psychological care among medical and nursing staff in Wuhan during the 2019 novel coronavirus disease outbreak: A cross-sectional study. Brain Behav Immun 2020;87:11-7. doi: 10.1016/j.bbi.2020.03.028.
4. Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. J Med Virol

2020;92:401-2. doi: 10.1002/jmv.25678.
5. Thoracic Society. Expert consensus for bronchoscopy during the epidemic of 2019 novel coronavirus infection (Trial version). Zhonghua Jie He He Hu Xi Za Zhi 2020;43:199-202. doi: 10.3760/cma.j.issn.1001-0939.2020.03.012.
6. Khan S, Ali A, Siddique R, Nabi G. Novel coronavirus is putting the whole world on alert. J Hosp Infect 2020;104:252-3. doi: 10.1016/j.jhin.2020.01.019.
7. Wu P, Hao X, Lau EHY, Wong JY, Leung KSM, Wu JT, et al. Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020. Euro Surveill 2020;25:2000044. doi: 10.2807/1560-7917.ES.2020.25.3.2000044.
8. Sands KE, Wenzel RP, McLean LE, Korwek KM, Roach JD, Miller KM, et al. Patient characteristics and admitting vital signs associated with coronavirus disease 2019 (COVID-19)-related mortality among patients admitted with noncritical illness. Infect Control Hosp Epidemiol 2021;42:399-405. doi: 10.1017/ice.2020.461.
9. Rajeswari S, SanjeevaReddy N. Efficacy of Progressive Muscle Relaxation on Pregnancy Outcome among Anxious Indian Primi Mothers. Iran J Nurs Midwifery Res 2019;25:e23-30. doi: 10.4103/ijnmr.IJNMR_207_18.
10. LaPerriere A, Ironson G, Antoni MH, Schneiderman N, Klimas N, Fletcher MA. Exercise and psychoneuroimmunology. Med Sci Sports Exerc 1994;26:182-90. doi: 10.1249/00005768-199402000-00008.
11. Budlender D. Whither the International Standard Classification of Occupations (ISCO-88)? Geneva, Switzerland: International Labour Organization; 2003.
12. Brighenti-Zogg S, Mundwiler J, Schüpbach U, Dieterle T, Wolfer DP, Leuppi JD, et al. Physical Workload and Work Capacity across Occupational Groups. PLoS One 2016;11:e0154073. doi: 10.1371/journal.pone.0154073.
13. In: Berrar DP, Dubitzky W, Granzow M, eds. A practical approach to microarray data analysis. New York, USA: Kluwer academic publishers; 2003.
14. Tukey JW. The future of data analysis. Ann Math Statist 1962;33:1-67.
15. Berrar D. Cross-validation. Encyclopedia of bioinformatics and computational biology 2019;1:542–5. Doi: 10.1016/B978-0-12-809633-8.20349-X
16. Chernoff H. Large-sample theory: Parametric case. Ann Math Statist 1956;27:1-22.
17. Lehmann EL. Elements of large-sample theory. New York, NY: Springer-Verlag; 1999.
18. Lotfy N. Spatial analysis of completeness of death registration in Egypt. J Egypt Public Health Assoc 2020;95:12. doi: 10.1186/s42506-020-00040-3.
19. Tian S, Hu N, Lou J, Chen K, Kang X, Xiang Z, et al. Characteristics of COVID-19 infection in Beijing. J Infect 2020;80:401-6. doi: 10.1016/j.jinf.2020.02.018.
20. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020;382:1199-207. doi: 10.1056/NEJMoa2001316.
21. Zhang JJ, Dong X, Cao YY, Yuan YD, Yang YB, Yan YQ, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. Allergy 2020;75:1730-41. doi: 10.1111/all.14238.
22. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020;395:507-13. doi: 10.1016/S0140-6736(20)30211-7.
23. Yanai H. A Significance of High Prevalence of Diabetes and Hypertension in Severe COVID-19 Patients. J Clin Med Res 2020;12:389-92. doi: 10.14740/jocmr4218.
24. Park SY, Kim JH, Kim HJ, Seo B, Kwon OY, Chang HS, et al. High Prevalence of Asthma in Elderly Women: Findings From a Korean National Health Database and Adult Asthma Cohort. Allergy Asthma Immunol Res 2018;10:387-96. doi: 10.4168/aair.2018.10.4.387.
25. He J, Wu B, Chen Y, Tang J, Liu Q, Zhou S, et al. Characteristic Electrocardiographic Manifestations in Patients With COVID-19. Can J Cardiol 2020;36:966.e1-966.e4. doi: 10.1016/j.cjca.2020.03.028.
26. Capoferri G, Osthoff M, Egli A, Stoeckle M, Bassetti S. Relative bradycardia in patients with COVID-19. Clin Microbiol Infect 2021;27:295-6. doi: 10.1016/j.cmi.2020.08.013.
27. Dhont S, Derom E, Van Braeckel E, Depuydt P, Lambrecht BN. The pathophysiology of 'happy' hypoxemia in COVID-19. Respir Res 2020;21:198. doi: 10.1186/s12931-020-01462-5.
28. Pimentel MAF, Redfern OC, Hatch R, Young JD, Tarassenko L, Watkinson PJ. Trajectories of vital signs in patients with COVID-19. Resuscitation 2020;156:99-106. doi: 10.1016/j.resuscitation.2020.09.002.
29. Nash MS. Exercise and immunology. Med Sci Sports Exerc 1994;26:125-7. doi: 10.1249/00005768-199402000-00001.
30. Shephard RJ, Shek PN. Immunological hazards from nutritional imbalance in athletes. Exerc Immunol Rev 1998;4:22-48.