



# HHS Public Access

Author manuscript

*Qual Manag Health Care*. Author manuscript; available in PMC 2024 January 01.

Published in final edited form as:

*Qual Manag Health Care*. 2023 ; 32(Suppl 1): S21–S28. doi:10.1097/QMH.0000000000000399.

## The Role of Symptom Clusters in Triage of COVID-19 Patients

Janusz Wojtusiak, PhD<sup>1</sup>, Wejdan Bagais, MS<sup>\*,1</sup>, Jee Vang, PhD<sup>1</sup>, Elina Guralnik, MPH<sup>1</sup>,  
Amira Roess, PhD<sup>2</sup>, Farrokh Alemi, PhD<sup>1</sup>

<sup>1</sup>Department of Health Administration and Policy, College of Health and Human Services, George Mason University

<sup>2</sup>Department of Global and Community Health, College of Health and Human Services, George Mason University

### Abstract

**Background and Objective:** COVID-19 manifests with a broad range of symptoms. This study investigates whether COVID-19 can be differentiated from other causes of respiratory, gastrointestinal, or neurological symptoms.

**Methods:** We surveyed symptoms of 483 subjects who had completed COVID-19 laboratory tests in the last 30 days. The survey collected data on demographic characteristics, self-reported symptoms for different types of infections within 14 days of onset of illness, and self-reported COVID-19 test results. Robust LASSO regression was used to create three nested models. In all three models, the response variable was the COVID-19 test results. In the first model, referred to as “main effect model,” the independent variables were demographic characteristics, history of chronic symptoms, and current symptoms. The second model, referred to as “hierarchical clustering model,” added clusters of variables to the list of independent variables. These clusters were established through hierarchical clustering. The third model, referred to as “interaction terms model,” also added clusters of variables to the list of independent variables; this time clusters were established through pairwise, and triple-way interaction terms. Models were constructed on a randomly selected 80% of the data and accuracy was cross validated on remaining 20% of the data. The process was bootstrapped 30 times. Accuracy of the three models was measured using average of cross-validated Area under the Receiver Operating Curves (AROC).

**Results:** In 30 bootstrap samples, the main effect model had an AROC of 0.78. The hierarchical clustering model had an AROC of 0.80. The interaction terms model had an AROC of 0.81. Both the hierarchical cluster model and the interaction model were significantly different from main effect model ( $\alpha = 0.04$ ). Patients with different races/ethnicities, genders, and ages presented with different symptom clusters.

**Conclusions:** Using clusters of symptoms, it is possible to more accurately diagnose COVID-19 among symptomatic patients.

### Keywords

COVID-19 diagnosis; clusters of symptoms; hierarchical clustering; LASSO regression

---

\*Corresponding author can be reached at wbagais@gmu.edu.

## Introduction

Diagnosis of COVID-19 in the community continues to be fraught with difficulties. COVID-19 is a systemic disease that presents with a variety of symptoms [1,2]. New variants of SARS-CoV-2 may present with different symptoms [3]. Patient age groups have been shown to manifest different COVID-19 symptoms [4,5]. Early and late progression of COVID-19 have different manifestations [6,7]. The order of occurrence of symptoms may matter [8,9]. Furthermore, simple rules of predictions based on symptoms most frequently associated with the disease are poor predictors of COVID-19 [10]. Further, algorithms that assume symptoms are independent of each other have proven to have low accuracy [8]. One way to resolve these challenges in diagnosing COVID-19 is to examine clusters of symptoms.

Symptom clusters occur in many diseases, from chronic kidney disease [11] to cancer [12], to COVID-19 [13]. Symptoms are often correlated with each other [14,15]. For example, nausea-vomiting, anxiety-depression, and dyspnea-cough clusters are common in cancer patients [16]. Symptom clusters can differentiate response to treatment better than clinician's diagnosis [17]. For example, symptom clusters affect quality of life [18] and patients' ability to function [19]. It is evident from the published literature that symptom clusters matter. The current study examines if symptom clusters can improve accuracy of triage decisions for patients suspected of COVID-19.

This study compares two approaches for constructing symptom clusters: (1) hierarchical clustering and (2) use of interaction terms. Historically, scientists have used hierarchical clustering [20,21] to combine variables and create clusters. The second method, i.e., relying on interaction terms to define clusters [22,23], of clusters to ability to predict the outcome. Thus, in the second method clusters not only indicate subgroups of patients with similar etiology but also with similar outcome and clinical response [24].

## Methods

### Source of Data:

Data were obtained from 483 patients who had taken a laboratory COVID-19 test, recruited through online listservs (with permission of listserv administrator). Participants learned about this study via emails to neighborhood listservs in the Washington DC area with the permission of the moderators. Study team members also shared the link to the survey with their personal networks. Individuals were directed to access the survey following a link at the end of the invitation message/emails. The first thing in the link was the consent page and then participant was directed to the survey.

Data were collected between November 2020 and January 2021, prior to widespread vaccination and emergence of Delta variant of COVID-19. Participants were eligible if they were adults, 18 years or older, and had a COVID-19 test within 30 days of participation in the study. Participants with inconclusive or pending test results were excluded from the analyses, resulting in 461 patients in the final analysis set. The majority of patients (90%) had at least one symptom before taking the test.

**Dependent Variable:**

Study participants self-reported the result of their COVID-19 laboratory test. The models developed here predict whether the patient has COVID-19 or not. For patients who do not have COVID-19, the model does not say which gastrointestinal, inflammatory, neurological, or other diagnosis is consistent with the patients' symptoms.

**Independent Variables:**

Study participants self-reported their current symptoms, history of chronic symptoms, and demographic characteristics (age, gender, race, and ethnicity). The 29 examined symptoms from different types of infections (and their prevalence in the study population) were:

---

1. General Symptoms	4. Inflammatory Symptoms
1.1. Fever or feeling feverish, (62%)	4.1. Joint/other unexplained pain (myalgia/arthralgia), (18%)
1.2. Muscle aches/myalgia, (35%)	4.2. Red/purple rash or lesions on toes, (5%)
1.3. Pinkeye/Conjunctivitis, (5%)	4.3. Unexplained rashes, (5%)
1.4. Fatigue (more than usual), (38%)	4.4. Excessive sweating, (7%)
1.5. Chills, (20%)	4.5. Bluish lips or face, (1%)
2. Neurological Symptoms	5. Respiratory Symptoms
2.1. Headaches, (45%)	5.1. Cough, (57%)
2.2. Loss of balance, (9%)	5.2. Sore throat, (38%)
2.3. New confusion, (6%)	5.3. Difficulty breathing (Dyspnea), (21%)
2.4. Unusual shivering or shaking, (11%)	5.4. Shortness of breath (Hypoxia), (20%)
2.5. Loss of smell, (8%)	5.5. Runny nose (Rhinorrhea/nasal symptoms), (58%)
2.6. Loss of taste, (20%)	5.6. Chest pain (chest tightness), (19%)
2.7. Numbness, (7%)	5.7. Wheezing, (23%)
2.8. Slurred speech, (3%)	
3. Gastrointestinal Symptoms	
3.1. Diarrhea, (28%)	
3.2. Stomach/abdominal pain, (14%)	
3.3. Change in or loss of appetite, (23%)	
3.4. Nausea or vomiting, (16%)	

---

The history of chronic symptoms included 4 variables measuring history of symptoms due to chronic conditions: (1) history of respiratory, (present in 21% of patients), (2) history of inflammatory, (present in 6% of patients), (3) history of neurological, (present in 9% of patients), and (4) history of gastrointestinal symptoms, (present in 11% of patients).

**Model Construction.**

We examined the relationship between symptoms and COVID-19 test results using the Least Absolute Shrinkage and Selection Operator (LASSO) regression models. The response variable was COVID-19 test results. In regressions, robust predictors were identified by randomly selecting 80% sample of training data and repeating LASSO regressions 23 times. Variables that remained in 95% of the regressions were kept as robust variables. In addition, in all LASSO regressions, we focused on large effects. The variable was ignored if the absolute value of regression coefficient did not exceed 0.01.

It is often useful to compare the performance of the model to accuracy of other models or clinical expectations. We have reported the accuracy of naïve models in a separate publication [25]. Naïve Bayes model was not accurate enough to be used in a clinical setting. In contrast, the LASSO regression model provides a high enough cross-validated accuracy that can be used in clinical triage decisions. The main rationale for selecting LASSO

regression is its ability to perform variable selection, and straight-forward interpretation of coefficients. In contrast, methods such as ridge regression or elastic net, perform “smoothing” of coefficients to remove extreme values, thus not allowing for the direct interpretation of the coefficients. In a separate paper in this supplement, we provide details of how the cluster of symptoms can also be used to increase accuracy of home tests [26].

### Different Clustering Approaches:

We examined three different methods of constructing symptom clusters:

1. **Main Effect Model:** COVID-19 test results were LASSO regressed on main effects (with no clusters or interaction terms) of symptoms, demographics (age, race, and ethnicity), and history of symptoms due to a chronic condition. Robust predictors were then selected as those that are most predictive of COVID-19 diagnosis. The main effect model assumes that no symptom cluster is used in predicting COVID-19 (see Supplement 2 for complete analysis).
2. **Hierarchical Clustering Model:** We combined predictors using hierarchical clustering and then built models for predicting COVID-19 using the identified clusters. Agglomerative clustering was applied to create a hierarchy of symptom clusters. Norm-1 similarity was used in which distance between two symptoms was defined as the number of people in the data that do not share the same symptom (other similarity measures, including Jaccard, led to comparable results). Hierarchical clustering was done separately for positive and negative COVID-19 cases. An optimal cut-off point in the hierarchical dendrogram was identified that maximized the accuracy of the models in the training data (see Supplement 3 for complete analysis).
3. **Interaction-Term Model:** Symptom clusters were identified through interaction terms in LASSO regression. As before, the response variable was COVID-19 test result. The independent variables were single, pair, and triplets of demographics, symptoms, and history of chronic symptoms.

For all models, robust LASSO regression was used to select the most predictive predictors. The regularization strength ( $\lambda$ ) was optimized to maximize cross-validated accuracy within the training set (see Supplement 4 for complete analysis).

### Accuracy:

The proposed regression models were developed in 80% of data and cross-validated in 20% of data set aside for validation. The process was repeated 30 times for randomly selected subsets. The average Area under the Receiver Operating Curve (AROC) was reported as the measure of cross-validated accuracy of predictions.

### Online Supplements:

Python source code and complete analysis are available in Jupyter Notebook and HTML format. Supplement 1 includes common functions and Supplements 2–4 construction and testing of specific models.

**IRB Approval:**

This study was approved by George Mason University IRB number 1668273-8.

**Results**

Table 1 describes the demographic distribution of the study participants. Most participants were white, Non-Hispanic/Latino, and between 25 to 44 years old. A large proportion of the study participants (42%) were healthcare or essential workers.

The hierarchical clusters of positive COVID-19 cases are shown in the dendrogram in Figure 1. Similarly, the hierarchical clusters of negative COVID-19 cases are shown in Figure 2. Comparison of the figures show clear differences in clusters between positive and negative cases. The optimal number of clusters when using hierarchical clusters was 15, most of which were single-predictor clusters.

Table 2 provides coefficients from LASSO logistic regression models. Empty cells correspond to the variables dropped by LASSO or not present in the first place. The main effect model had average AROC of 0.78 (variance 0.00199), the lowest accuracy level. In this model all predictors (except age 30 years or older) increased probability of COVID-19. By themselves, headache (1.06), chest pain (0.69), cough (0.43) and loss of taste (0.38) were the strongest predictors of COVID-19 diagnosis.

Table 2 shows the average AROC, sensitivity, and specificity for the three models. When hierarchical clustering was used as predictors of test results, the average AROC was 0.79 (variance 0.00295). At the most optimal cutoff point, the hierarchical cluster technique identified 10 clusters, 8 of which were individual predictors. The 2 combinations of symptoms were: (1) “cough, fever, runny nose, and headaches,” with regression coefficient of 1.57; and (2) “chest pain, difficulty breathing, wheezing, chills, and shortness of breath,” with regression coefficient of 2.22.

The Interaction-Term cluster model had average AROC of 0.81 (variance 0.00279) the highest AROC value (see Table 2). Surprisingly, no main effects, were present in this model, showing that symptom clusters were more predictive of COVID-19 than any single symptom. The strongest predictor positively associated with COVID-19 diagnosis was “fatigue & chest pain among white individuals,” with regression coefficient of 1.56. The strongest predictor that was negatively associated with predicting COVID-19 was “runny nose, observed among non-Hispanic or Latino individuals,” with regression coefficient of -0.91.

The impact of symptom clusters on diagnosis of COVID-19 depended on the age, gender, and race of the patient, as patients in different demographic subgroups presented with different symptom clusters. Table 2 shows symptom clusters in distinct demographic subgroups. Race was a factor in 7 clusters. Ethnicity was a factor in 6 clusters. Age appeared in 2 clusters. Likewise, female gender appeared in 2 clusters.

In 30 bootstrap samples, the Interaction-Term cluster model was significantly more accurate than the main-effect model (paired t-statistic 5.43, 29 degrees of freedom,  $p = 0.00$ ). The

Hierarchical Clustering model was not more accurate than the main-effect model (paired t-statistic = 1.50, degrees of freedom = 29,  $p = 0.07$ ).

Figure 3 visualizes the LASSO regression with clusters using network modeling procedures [27]. All clusters that predict COVID-19 are shown, yellow nodes show clusters that increase the probability and blue rectangles show clusters that reduce the probability of COVID-19. The impact of listed single symptoms, single history of chronic symptom, or single demographic characteristics are mediated through these clusters. It is important to note that 10 single symptoms were not present in any of the clusters and are not listed in Figure 3. Those were slurred speech, diarrhea, bluish lips, confusion, unexplained rash, shivering, abdominal pain, excessive sweating, loss of balance, and red rash. In addition, history of neurological, gastrointestinal, and immune function symptoms did not show in the final regression equation and are not listed in Figure 3.

## Discussion

The Interaction-Term model has a cross-validated AROC of 0.81, which is a relatively high level of accuracy and can be a clinically meaningful screening method for COVID-19. This level of accuracy is comparable to in-home, rapid, antigen test results [28]. Currently, the Centers for Disease Control (CDC) website provides a list of symptoms and no guidance on how to use these symptoms in diagnosis of COVID-19 [29]. Now that COVID-19 is endemic, clinicians need advice on how to diagnose COVID-19 from its symptoms. To date, this has not been possible because symptom screening has not been accurate enough to be used in clinics or triage decisions. This study showed that symptom clusters can improve the accuracy of diagnosis of COVID-19 to levels that are more acceptable in clinical management of patients.

We acknowledge that the interaction-term model developed here (and the associated symptom clusters), while accurate, is difficult to use by unaided clinicians. The existence of smart phones, and access of a large portion of the population to the Internet, suggests that these models can be made part of an internet service. Patients can describe their symptoms to the service; it can score symptom clusters, and report if the patient is likely to have COVID-19. The patients can then report the service's presumed diagnosis to their clinician, who can use the service's findings as any laboratory test. It may be possible to overcome the difficulty of implementing the proposed screening of symptom clusters using an Internet service.

The Interaction-Term symptom cluster model was more accurate than the main-effect model, suggesting that clusters and not individual symptoms are helpful in diagnosis of COVID-19. The improvement in AROC was small but more pronounced when examining sensitivity/specificity: At constant specificity (0.93), cluster of symptoms increased sensitivity by 9 points (from 0.45 to 0.55) compared to the sensitivity of the model built on individual symptoms. In addition, the LASSO regression with interaction terms excluded any single symptom, suggesting that symptom clusters are more predictive than single symptoms.

There are many explanations why symptom clusters can make diagnosis of COVID-19 more accurate. Clusters of symptoms are combinations of symptoms that typically do not occur in many patients, but when they do, they are highly predictive and add more information to the diagnosis task. Similar findings have been reported in the literature, where symptom clusters, not single symptoms, determine the patient's response to depression treatment [30].

A review of statistically significant variables in the Interaction-Term model suggests that based on age, gender, race, and ethnicity different symptom clusters are useful in diagnosing COVID-19. For example, Cough among 18- to 29-year-old females was predictive of COVID-19 (coefficient of 0.36). Similarly, runny nose among non-Hispanic/Latinos was helpful in ruling out COVID-19 [also see 8]. These and other clusters show that symptoms of COVID-19 may vary based on demographic characteristics. There are many reasons for this, younger patients have fewer comorbidities and may not express the same symptoms as older patients for the same disease [31]. Men and women differ in expression of symptoms for same underlying disease [32,33]. Finally, racial differences in presentation of diseases have been known to occur. This study contributes to the growing literature that symptoms must be evaluated within age/gender/race groups.

## Limitations

This study had several limitations. First, the models that were constructed relied on a relatively small data sample, collected between October 2020 and January 2021, with only one variant of SARS-CoV-2 virus known to be prevalent at the time of data collection. Larger sample size may allow for discovery of more combinations of symptoms uniquely present in COVID-19 patients. Bigger data many include variables excluded by current model but reported in the literature [19].

The participants' age and gender distribution in our study sample does not correspond to the distribution of age and gender within the United States population. Our findings may not generalize to United States population.

COVID-19 test results were self-reported and established by a variety of diagnostic tests. This may have introduced bias in accurately assessing the true positive COVID-19 test results.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgment

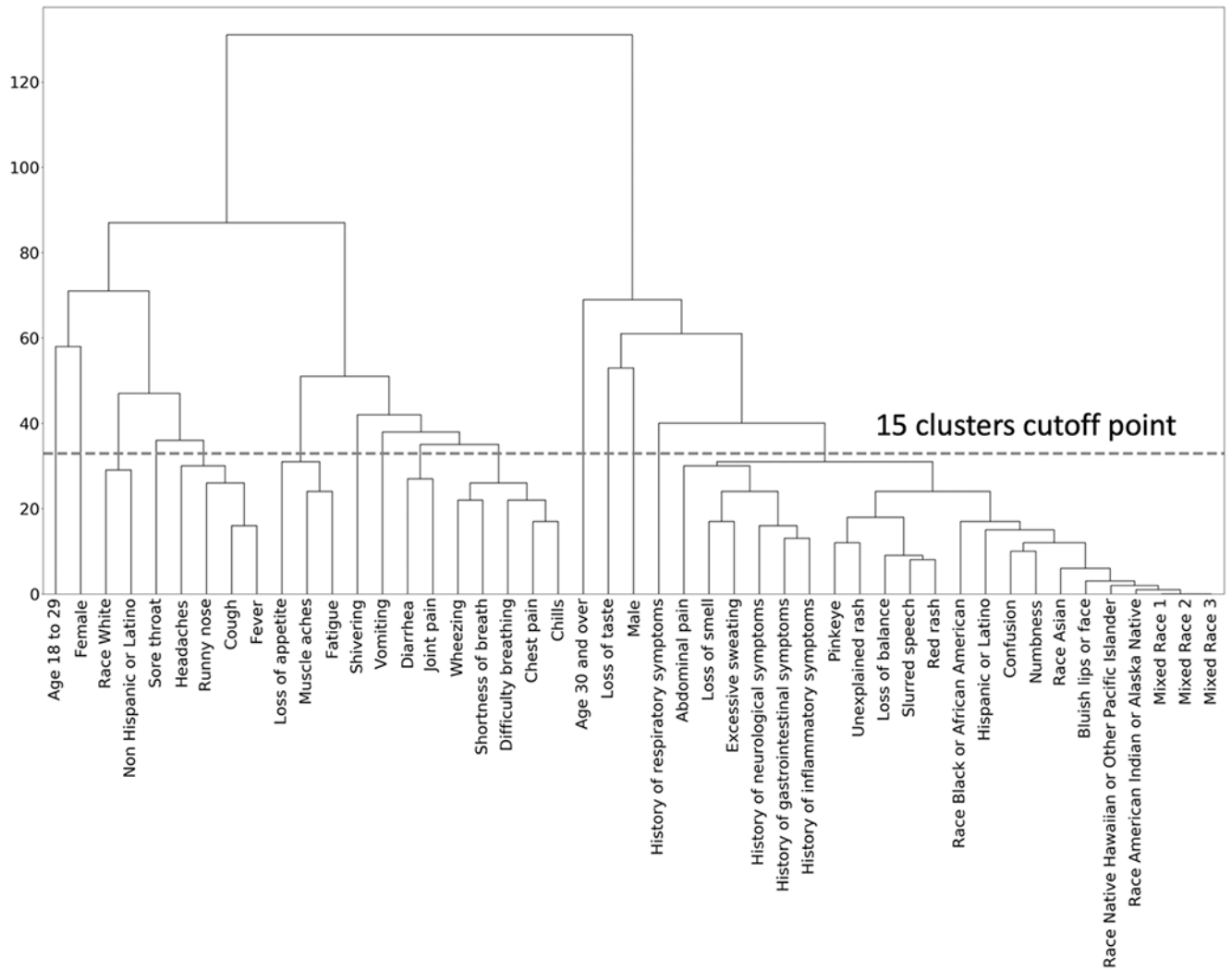
This project was funded by National Cancer Institute contract number 75N91020C00038 to Vibrent Health, Praduman Jain (Principal Investigator). All listed authors and acknowledged individuals were paid by the contract and had no conflicts of interest to declare.

## References

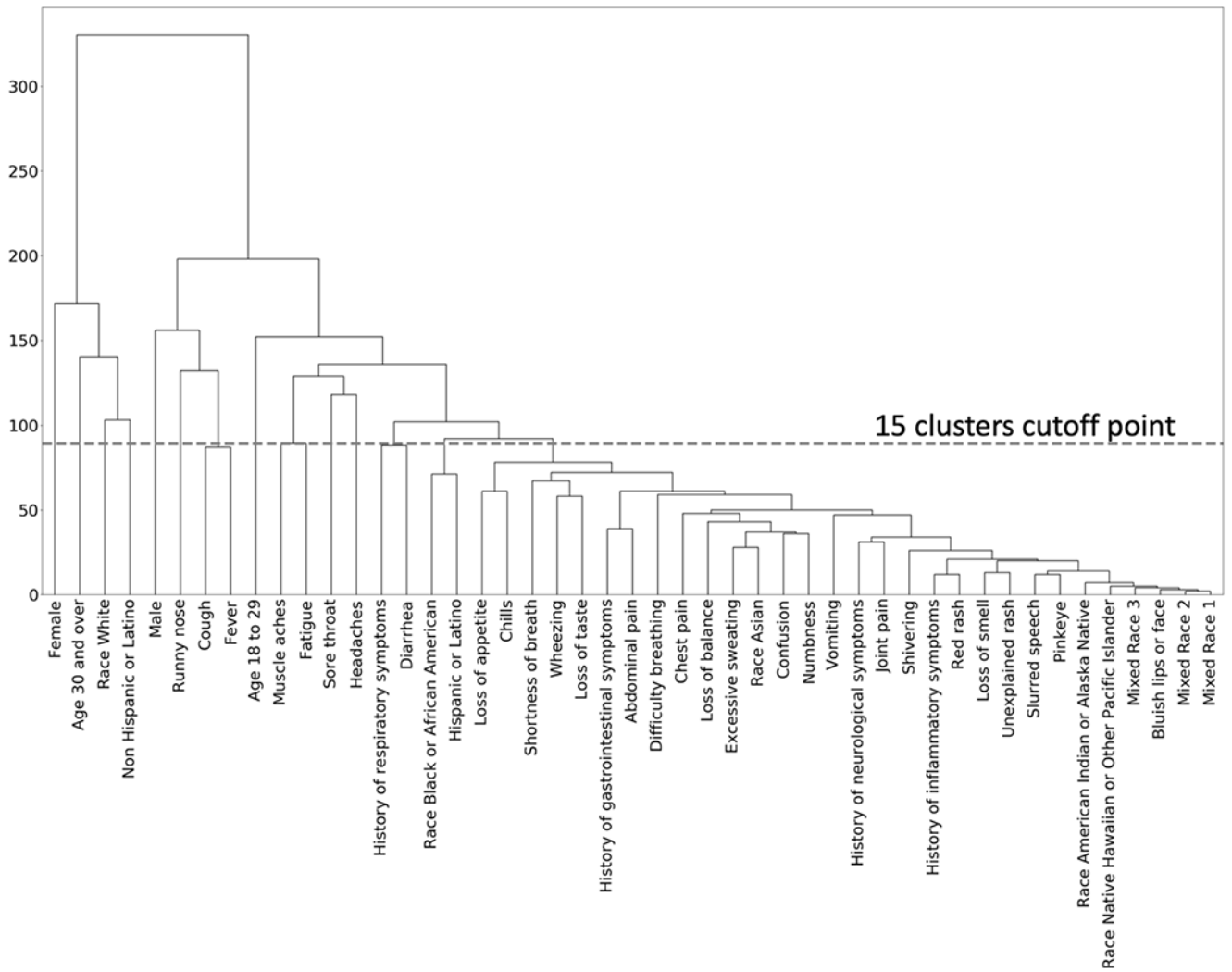
1. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, ... & Spector TD (2020). Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature medicine*, 26(7), 1037–1040.
2. Alemi F, Guralnik E, Vang J, Wojtusiak J, Wilson A, Roess A. Non-respiratory symptoms of COVID-19: A Meta Literature Review. 2021, available through first author at falemi@gmu.edu
3. Mahase E (2021). Covid-19: Sore throat, fatigue, and myalgia are more common with new UK variant, *BMJ*. 2021;372. doi:10.1136/bmj.n288
4. Trevisan C, Noale M, Prinelli F, et al. Age-Related Changes in Clinical Presentation of Covid-19: the EPICOVID19 Web-Based Survey. *Eur J Intern Med*. Apr 2021;86:41–47. doi:10.1016/j.ejim.2021.01.028 [PubMed: 33579579]
5. Jin S, An H, Zhou T, et al. Sex- and age-specific clinical and immunological features of coronavirus disease 2019. *PLoS Pathog*. Mar 2021;17(3):e1009420. doi:10.1371/journal.ppat.1009420 [PubMed: 33770147]
6. Chen J, Qi T, Liu L, et al. Clinical progression of patients with COVID-19 in Shanghai, China. *J Infect*. 2020;80(5):e1–e6. doi:10.1016/j.jinf.2020.03.004
7. Wagner T, Shweta F, Murugadoss K, et al. Augmented curation of clinical notes from a massive EHR system reveals symptoms of impending COVID-19 diagnosis. *eLife*. 2020;9 doi:10.7554/eLife.58227
8. Larsen JR, Martin MR, Martin JD, Kuhn P, & Hicks JB (2020). Modeling the onset of symptoms of COVID-19. *Frontiers in public health*, 8, 473. [PubMed: 32903584]
9. Wojtusiak J, Bagais WE, Roess A, Alemi F. Order of Occurrence of COVID-19 Symptoms. *Supplement Quality Management in Healthcare*, forthcoming.
10. Alemi F, Vang J, Guralnik E, Roess A. Modeling the Probability of COVID-19 Based on Symptom Screening and Prevalence of Influenza and Influenza-Like Illnesses. *Quality Management in Healthcare*, 2021; (in press)
11. Lockwood MB, Chung S, Puzantian H, Bronas UG, Ryan CJ, Park C, DeVon HA. Symptom Cluster Science in Chronic Kidney Disease: A Literature Review. *West J Nurs Res*. 2019 Jul;41(7):1056–1091. [PubMed: 30378466]
12. George MA, Lustberg MB, Orchard TS. Psychoneurological symptom cluster in breast cancer: the role of inflammation and diet. *Breast Cancer Res Treat*. 2020 Nov;184(1):1–9. [PubMed: 32803635]
13. Yifan T, Ying L, Chunhong G, Jing S, Rong W, Zhenyu L, Zejuan G, Peihung L. Symptom Cluster of ICU Nurses Treating COVID-19 Pneumonia Patients in Wuhan, China. *J Pain Symptom Manage*. 2020 Jul;60(1):e48–e53.
14. Miaskowski C, Dodd M, Lee K. Symptom clusters: the new frontier in symptom management research. *J Natl Cancer Inst Monogr*. 2004(32):17–21. [PubMed: 15263036]
15. Barsevick AM, Aktas A. Cancer symptom cluster research: new perspectives and tools. *Curr Opin Support Palliat Care*. 2013;7(1):36–37 [PubMed: 23314017]
16. Kirkova J, Aktas A, Walsh D, Davis MP. Cancer symptom clusters: clinical and research methodology. *J Palliat Med*. 2011 Oct;14(10):1149–66. [PubMed: 21861613]
17. Wise J Covid-19: Study reveals six clusters of symptoms that could be used as a clinical prediction tool. *BMJ*. 2020:m2911. doi:10.1136/bmj.m2911 [PubMed: 32690476]
18. Miaskowski C, Aouizerat BE, Dodd M, Cooper B. Conceptual issues in symptom clusters research and their implications for quality-of-life assessment in patients with cancer. *J Natl Cancer Inst Monogr*. 2007(37):39–46. [PubMed: 17951230]
19. Dodd MJ, Miaskowski C, Paul SM. Symptom clusters and their effect on the functional status of patients with cancer. *Oncol Nurs Forum*. 2001;28(3):465–470. [PubMed: 11338755]
20. Johnson SC (1967). Hierarchical clustering schemes. *Psychometrika*, 32(3), 241–254. [PubMed: 5234703]
21. Murtagh F, & Contreras P (2012). Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1), 86–97.



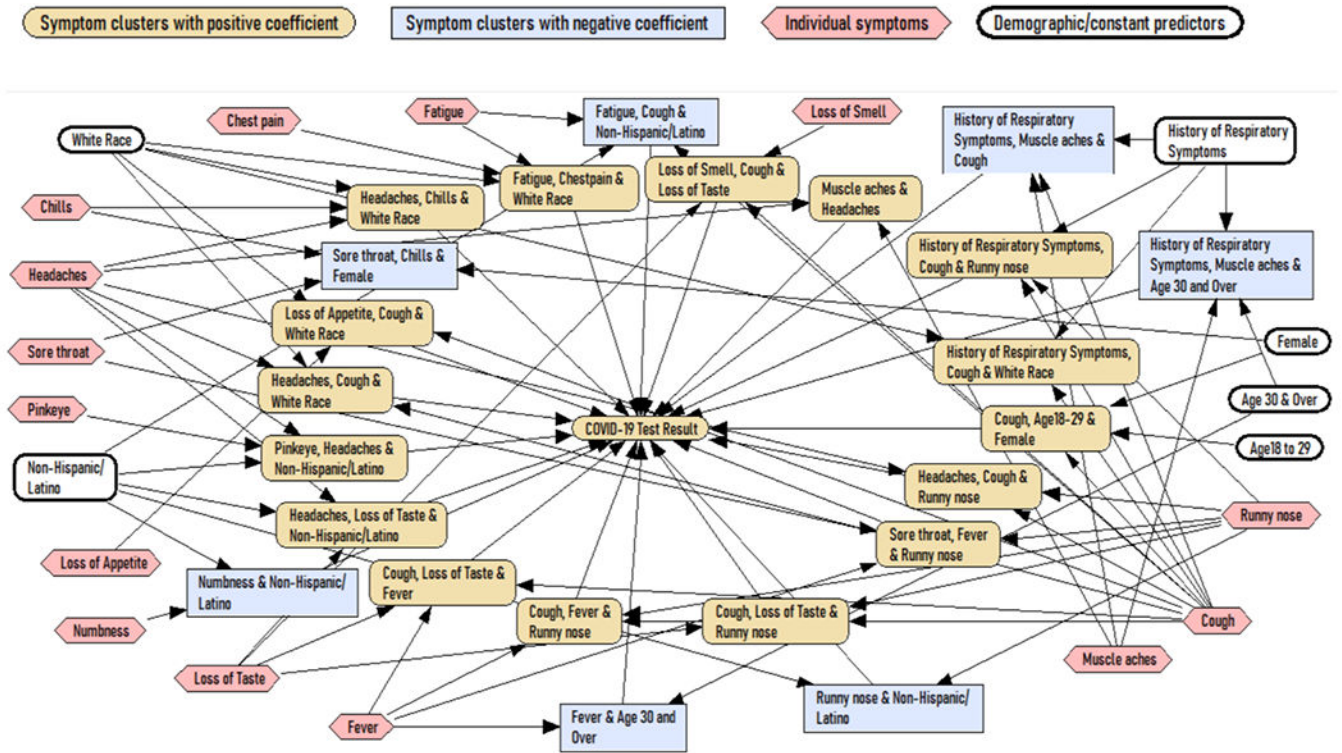
22. Michalski RS (1978). Pattern recognition as knowledge-guided computer induction (Tech. Rep. No. 927). Urbana-Champaign: University of Illinois, Department of Computer Science.
23. Wnek J (2012) Constructive Induction. In: Seel NM (eds) Encyclopedia of the Sciences of Learning. Springer, Boston, MA. 10.1007/978-1-4419-1428-6\_1063
24. Miaskowski C, Barsevick A, Berger A, et al. Advancing Symptom Science Through Symptom Cluster Research: Expert Panel Proceedings and Recommendations. *J Natl Cancer Inst.* 2017;109(4).
25. Alemi F, Vang J, Guralnik E, Roess A. Modeling the Probability of COVID-19 Based on Symptom Screening and Prevalence of Influenza and Influenza-Like Illnesses. *Qual Manag Health Care.* 2022 Apr-Jun 01;31(2):85–91. doi: 10.1097/QMH.0000000000000339. [PubMed: 35195616]
26. Alemi et al. Combined Symptom Screening and At-Home Tests for COVID-19. Supplement on COVID-19, *Qual Manag Health Care.* 2022
27. Alemi F Constructing Causal Networks Through Regressions: A Tutorial. *Qual Manag Health Care.* 2020 Oct/Dec;29(4):270–278. doi: 10.1097/QMH.0000000000000272. [PubMed: 32991546]
28. Alemi F, Vang J, Bagais WH, Guralnik E, Wojtusiak J, Moeller G, Schilling J, Peterson R, Roess A, Jain P. Combined Symptom Screening and Home Tests for COVID-19. *COVID-19 Supplement to Healthcare Quality Management Journal*, in review.
29. Similarities and differences between flu and COVID-19. Centers for Disease Control and Prevention. <https://www.cdc.gov/flu/symptoms/flu-vs-covid19.htm>. Accessed Nov. 19, 2020.
30. Bondar J, Caye A, Chekroud AM, Kieling C. Symptom clusters in adolescent depression and differential response to treatment: a secondary analysis of the Treatment for Adolescents with Depression Study randomised trial. *Lancet Psychiatry.* 2020 Apr;7(4):337–343. [PubMed: 32199509]
31. Mueller AL, McNamara MS, Sinclair DA. Why does COVID-19 disproportionately affect older people? *Aging (Albany NY).* 2020 May 29;12(10):9959–9981. doi: 10.18632/aging.103344. Epub 2020 May 29. [PubMed: 32470948]
32. Cavanagh A, Wilson CJ, Kavanagh DJ, Caputi P. Differences in the Expression of Symptoms in Men Versus Women with Depression: A Systematic Review and Meta-analysis. *Harv Rev Psychiatry.* 2017 Jan/Feb;25(1):29–38. doi: 10.1097/HRP.0000000000000128. [PubMed: 28059934]
33. Regitz-Zagrosek V. Sex and gender differences in health. *Science & Society Series on Sex and Science. EMBO Rep.* 2012 Jun 29;13(7):596–603. doi: 10.1038/embor.2012.87. [PubMed: 22699937]



**Figure 1:**  
Hierarchical Clustering among positive COVID-19 Patients



**Figure 2:**  
Hierarchical Clusters for Negative COVID-19 Patients



**Figure 3.** Bayesian network model for predicting COVID-19

**Table 1:**

## Characteristics of study sample

Variable	Number of Cases (%)
<b>COVID-19 Test Results</b>	
Negative	330 (68.32%)
Positive	131 (27.12%)
Results Pending	15 (3.11%)
Inconclusive	7 (1.45%)
<b>Age</b>	
18-24	84 (17.39%)
25-34	210 (43.48%)
35-44	156 (32.30%)
45-54	20 (4.43%)
55-84	13 (2.69%)
<b>Gender</b>	
Female	279 (57.76%)
Male	203 (42.03%)
<b>Ethnicity</b>	
Hispanic Latino	60 (12.42%)
Non-Hispanic/Latino	401 (83.02%)
Unknown	22 (4.55%)
<b>Race</b>	
Other	18 (3.75%)
Asian	25 (5.18%)
Black or African American	60 (12.42%)
White	380 (78.67%)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2:**

Robust LASSO Logistic Regressions

Model	Main Effects	Interaction-Term Clusters	Hierarchical Clustering
Average Area under the Receiver Operating Curve	0.78	0.81	0.79
Variance of Area under the Receiver Operating Curve	0.0019	0.0027	0.0029
Average Sensitivity	0.64	0.67	0.64
Average Specificity	0.82	0.85	0.86
Total number of possible predictors	47	17,343	23
Number of robust predictors	11	23	10
	Coefficients of LASSO Regression Models		
Predictors	Main Effect	Interaction-Term Clusters	Hierarchical Clustering
Intercept	-2.21	-1.77	-2.03
Cough	0.43		
Chills	0.37		
Headaches	1.06		0.61
Joint pain	0.35		
Chest pain	0.69		
Runny nose			-0.55
Loss of smell	0.17		
Loss of taste	0.38		0.77
Loss of appetite	0.27		
Difficulty breathing	0.20		
Muscle aches			0.22
History of respiratory symptoms			0.24
Age 30+	-0.26		-0.55
White	0.11		0.38
Female			0.33
Cough, fever, runny nose & headaches			1.57
Chest pain, difficulty breathing, wheezing, chills & shortness of breath			2.22
Fever & age 30+		-0.44	
Muscle aches & headaches		0.69	
Numbness & non-Hispanic/ Latino		-0.57	
Runny nose & non-Hispanic/ Latino		-0.91	
Cough, fever & runny nose		0.54	
Cough, loss of taste & fever		0.32	
Headaches, cough & runny nose		0.59	
Headaches, cough & White		0.64	
Headaches, chills & White		0.63	

Model	Main Effects	Interaction-Term Clusters	Hierarchical Clustering
Sore throat, fever & runny nose		0.47	
Fatigue, chest pain & White		1.56	
Cough, loss of taste & runny nose		0.06	
Sore throat, chills & female		-0.21	
Loss of appetite, cough & White		0.47	
Loss of smell, cough & loss of taste		0.91	
Cough, age 18 to 29 & female		0.36	
Fatigue, cough & non-Hispanic/Latino		-0.26	
Pinkeye, headaches & non-Hispanic/Latino		0.71	
History of respiratory symptoms, cough & runny nose		0.20	
History of respiratory symptoms, cough & White		0.60	
History of respiratory symptoms, muscle aches & cough		-0.63	
Headaches, loss of taste & non-Hispanic/Latino		0.39	
History of respiratory symptoms, muscle aches & age 30+		-0.86	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript