Order of Occurrence of COVID-19 Symptoms

Janusz Wojtusiak, PhD[1], **Wejdan Bagais**[*,1], **Jee Vang, PhD**[1], **Amira Roess**[2], **Farrokh Alemi**[1]

[1]Department of Health Administration and Policy, College of Health and Human Services, George Mason University

[2]Department of Global and Community Health, College of Health and Human Services, George Mason University

Abstract

Background: COVID-19 symptoms change after onset—some show early, others later.

Objectives: This paper examines if the order of occurrence of symptoms can improve diagnosis of COVID-19 before test results are available.

Methods: 483 individuals who completed a COVID-19 test were recruited through listservs. Participants then completed an online survey regarding their symptoms and test results. The order of symptoms was set according to: (a) whether the participant had a "history of the symptom" due to a prior condition, and (b) whether the symptom "occurred first," or prior to, other symptoms of COVID-19. Two LASSO regression models were developed. The first model, referred to as "time-invariant," used demographics and symptoms, but not the order of symptom occurrence. The second model, referred to as "time-sensitive," used the same data set, but included the order of symptom occurrence.

Results: The average cross-validated Area under the Receiver Operating Curve (AROC) for the time-invariant model was 0.784. The time-sensitive model had an AROC of 0.799. The difference between the two accuracy levels was statistically significant (alpha<.05).

Conclusion: The order of symptom occurrence made a statistically significant, but small, improvement in the accuracy of the diagnosis of COVID-19.

Keywords

COVID-19 diagnosis; order of symptoms; symptom screening; predictive model; LASSO regression

Introduction

COVID-19 is a viral disease that is now endemic in the United States. No clear guidelines are available for the diagnosis of COVID-19 based on symptoms. Clinicians need to triage patients to testing sites or COVID-19 reception areas at outpatient clinics (e.g., wait in car). Symptom screening can prevent subjecting patients—who do not have COVID-19—to iatrogenic exposure.

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

Rapid antigen home tests are freely available by mail in the United States, but less than 25% of patients with COVID-like symptoms test at home during their illness [1]. Thus, the majority of patients who call their clinicians have not done a rapid test; the clinician has to rely on reported symptoms to make triage decisions. Even patients who have access to testing need to consult a clinician about interpreting their results; manufacturers require these tests to be evaluated in a clinical context. This is due to varied sensitivity among rapid tests: "In people with signs and symptoms of COVID-19, sensitivities are highest in the first week of illness when viral loads are higher" [2]. The average sensitivity of rapid antigen tests is 68% [3], which is too low to be clinically useful by itself. Sensitivity can be improved by combining antigen tests with symptom screening [4]. Thus, symptom screening is essential to the use of in-home rapid antigen tests, either in interpreting the results or improving the sensitivity of these tests.

There is confusion about what symptoms should be screened to identify COVID-19, and how to differentiate COVID-19 from other diseases. Early on, fever and difficulty breathing were identified as signature symptoms of COVID-19; however, these are also present in a range of other respiratory illnesses [5]. COVID-19 is now considered a systemic disease with multiple manifestations, including many non-respiratory symptoms [6,7]. COVID-19 symptoms differ by age [8], by SARS-CoV-2 variant [9], and by underlying condition [10]. Diagnosis based on symptoms is challenging, in part, because symptoms can also vary by stage of infection [11]. Some patients are asymptomatic, while others develop mild symptoms early in the course of infection [12]. Studies have shown that patients' symptoms, such as cough and fever, are not specific to COVID-19, and no clear clinical guidelines exist on how to diagnose COVID-19 from observed symptoms [2].

Some investigators have suggested that the order of symptoms (e.g., whether cough occurs before or after fever) may be informative. Larsen and colleagues conducted a simulation study that suggested symptom order matters in the diagnosis of COVID-19 [13]. Their study identified a common sequence: fever first, then cough, followed by diarrhea and gastrointestinal symptoms. The sequence was compared to the progression of other respiratory diseases, such as influenza, severe acute respiratory syndrome, and middle eastern respiratory syndrome. This simulation suggested that the order of symptoms can be useful in distinguishing respiratory illnesses from each other.

Of course, patients have radically different presentations based on the time elapsed from onset to hospital admission. Patients who were 7 days past onset of the disease were significantly more likely to present with fever, cough, shortness of breath, muscle ache, joint pain, fatigue, headache, confusion, or diarrhea [14]. Patients with less than 7 days had fewer occurrences of these symptoms. These data suggest that time from onset is a significant factor in presentation of the disease. In other diseases, incorporating the order of variable occurrence has been shown to improve accuracy of predictions [15]. This paper builds upon previous studies concerning symptom order and clarifies if this information can improve the diagnosis of COVID-19.

Material and Methods

Source of Data:

483 patients were surveyed between November 2020 and January 2021 (before the emergence of the Delta variant of COVID-19). Study participants were recruited through listservs with permission from moderators. Participants were eligible if they were adults 18 years or older and had a COVID-19 test 30 days prior to the survey. At the time of data collection, the national rate of COVID-19 was 5%. Participants with missing or inconclusive COVID-19 test results were removed, which left the final cohort with 461 patients.

Method of Identifying COVID-19:

The survey asked for the patient's test results. At the time, few patients had access to rapid home tests, so most patients had to get tested at area laboratories or hospitals.

Symptoms:

The survey captured participants' COVID-19 symptoms, exposures, general health status, and socio-demographic status. The 29 symptoms listed included general symptoms (fever or feeling feverish, muscle aches/myalgia, pinkeye/conjunctivitis, excessive fatigue, and chills), neurological symptoms (headaches, loss of balance, new confusion, unusual shivering or shaking, loss of smell, loss of taste, seizures), gastrointestinal symptoms (diarrhea, stomach/abdominal pain, change in or loss of appetite, and nausea or vomiting), inflammatory symptoms (joint/other unexplained pain or myalgia/arthralgia, red/purple rash or lesions on toes, unexplained rashes, excessive sweating) and respiratory symptoms (cough, sore throat, difficulty breathing or dyspnea, shortness of breath or hypoxia, runny nose or rhinorrhea/nasal symptoms, and chest pain or chest tightness). The demographic items included age, gender, and race.

Timing of Symptoms:

Two approaches for establishing symptom order were examined. In the first approach, we determined if the symptom was present at the onset of illness or occurred later. In particular, the following question was asked: "Which of the following symptoms occurred on the first day you felt ill?" This question was followed by a list of symptoms to be selected by the respondent.

The second approach clarified if the patient had a history of presenting with a particular symptom; some pre-existing chronic diseases have similar symptoms to COVID-19. The long-term presence of a symptom may alter its importance during diagnosis. For example, in patients with multiple sclerosis, loss of smell/taste may be common and may not be indicative of COVID-19. To assess the history of symptoms for different categories, the following questions were asked:

1. [The previous question asks about gastrointestinal symptoms]. Are these symptoms consistent with any chronic gastrointestinal conditions you have, such as Celiac, Crohn's, Diverticulitis, GERD, Irritable Bowel syndrome, food allergies, or other similar chronic conditions?

2. [The previous question asks about neurological symptoms]. Are these symptoms consistent with any chronic neurological conditions you have, such as migraines, stroke, or other similar chronic conditions?

3. [The previous question asks about immunological symptoms]. Are these symptoms consistent with any chronic immune system conditions you have, such as Graves' disease, Lupus, Lyme disease, Rheumatoid arthritis, or other similar chronic conditions?

4. [The previous question asks about respiratory symptoms]. Are these symptoms consistent with any chronic respiratory conditions you have, such as asthma, COPD, CHF, seasonal allergies, or other similar chronic conditions?

These two methods created three timeframes for any symptom (see figure 1), in which: the patient had a history of having the symptom; the patient had the symptom at the onset of the disease; or the patient had the symptom after the onset of the disease. As indicated in figure 1, the survey occurs after the presentation of symptoms, and patients are recalling their symptoms and when they occurred.

Model Construction:

Two models were constructed: time-sensitive and time-invariant. The time-invariant model included age, gender, race, and symptoms as independent variables—it did not include any information about the order of symptoms. The time-sensitive model included all the symptoms from the time-invariant model plus the same symptoms coded as 'first' or 'delayed' occurrence, and included variables that measured if the patient had a history of symptoms.

The inclusion of interactions led to a large number of variables in the training data set; Least Absolute Shrinkage and Selection Operator (LASSO) was used to limit the number of variables. To ensure robust models [16,17], LASSO regressions were repeated 23 times within 80% of random training data (see figure 2). The number 23 was dictated by our computational limits, and an odd number prevents the need for tie-breaking. Variables that were statistically significant in 95% of these regressions were considered robust predictors. The accuracy of the robust predictors was examined in the test data. Once the variables were identified, there were multiple models, each with different parameters and variables. To restrict the model to robust variables, a final LASSO regression was needed to specify the model parameters; only robust variables were included in this regression.

Measure of Accuracy:

To measure the accuracy of the models, data were randomly split into 80% training and 20% testing (see figure 2). Training data were used for model building, and test data were used to measure the accuracy of the models according to the Area under the Receiver Operating Curve (AROC). To get a more reliable measure of the model quality, the entire process was repeated 30 times, and the average AROC is reported.

Results

In total, 483 participants completed the survey; 22 respondents were awaiting their test results or had an inconclusive test result, leaving 461 patients for the analysis. Table 1 shows the demographic characteristics of the population recruited for our study.

Table 2 describes the patients' symptom history. Having a history of neurological conditions and their associated symptoms doubled the risk of testing positive for COVID-19. Similarly, a history of inflammatory and/or respiratory conditions increased the probability of a positive test. In contrast, having a history of symptoms found in chronic gastrointestinal diseases reduced the probability of testing positive for COVID-19.

The final time-sensitive model was based on 14 robust variables, constructed from the combination of age, gender, symptoms, and order of symptoms (first occurrence and/or symptom history). The time-invariant model had 11 robust variables, also constructed from the same combination of independent variables, but the order of symptoms was excluded.

The average cross-validated AROC for the time-invariant model was 0.784. The AROC for the time-sensitive model was 0.799. The difference between these two AROCs was statistically significant according to a t-test ($p < 0.05$). Table 3 provides a summary of the accuracy and numbers of predictors used in the models. Both models had a similar score for specificity, but the time-sensitive model scored higher for sensitivity. Figure 3 shows the ROC curve and highlights the best threshold for both models, identified by G-mean.

Table 4 provides the robust variables from the final LASSO regressions. For the time-invariant model, the symptoms that were predictive of COVID-19 included headaches, chills, coughs, difficulty breathing, chest pain, joint pain, loss of appetite, loss of smell, and loss of taste. Two of these symptoms (difficulty breathing and joint pain) dropped out of the LASSO equation when time-sensitive symptoms were included in the model. The time-sensitive symptoms in the final LASSO models included headaches at onset, chills at onset, cough at onset, runny nose at onset, chest pain delayed, headaches delayed, loss of appetite delayed, loss of smell delayed, loss of taste delayed, runny nose delayed, vomiting delayed, and wheezing delayed. Of particular interest was the fact that runny nose at onset increased the odds of testing positive for COVID-19 by 1.16 times (regression coefficient of 0.15), but delayed occurrence of runny nose decreased the odds of testing positive by 1.13 times (regression coefficient of −0.12). When the timing of runny nose was ignored, this symptom was not predictive of COVID-19.

Discussion

The inclusion of time-dependent predictors and the order of occurrence of symptoms significantly improved model accuracy, but the magnitude of improvement was small. Often, adding variables to a prediction model increases the AROC by a small amount, but this does not mean that the variables are weak predictors. A highly predictive variable may have a minor impact on AROC, especially when starting with a high baseline AROC value. For example, a variable that doubles the risk only adds 0.01 points to AROC [18]. This is also the case when highly predictive, but not frequently present, predictors are used.

In situations with many variables but few cases, LASSO regression is expected to find some combination of variables that have a large AROC. This is also the case in our study; the AROC for the time-invariant model is expected to be high. What is most important is not the small magnitude of improvement in the AROC, but that the improvement was statistically significant. This suggests that information about the order of symptoms can improve the accuracy of predictions, albeit in small ways. The results do not indicate that other symptoms may not be statistically significant predictors of COVID-19. Instead, LASSO regression finds an optimal combination of variables for the best prediction.

Table 4 showed that timed-symptoms had more information than untimed symptoms, and were thus more likely to be picked up by LASSO regression as predictive of COVID-19. When headaches, chills, cough, or runny nose occurred first, they were more explanatory than the same symptoms without chronology (runny nose is completely eliminated as a time-invariant symptom). Conversely, when chest pain, headaches, loss of appetite, and loss of taste occurred later in the illness, these symptoms were less informative than the same symptom without chronological information. Moreover, the coefficient for runny nose reversed its sign between first and delayed occurrences of the symptom, explaining why the symptom was eliminated as a time-invariant predictor. The preceding information confirms that the timing of symptoms matters in COVID-19 diagnosis.

It is important to point out that all the symptom histories were eliminated from the final time-sensitive model by LASSO, suggesting that what matters in predicting COVID diagnosis is not the history of the symptoms, but which symptoms occurred first. Of course, history of chronic symptoms indicates the presence of comorbidities that cause the symptoms. These comorbidities may contribute to the severity of COVID-19 [19], but not necessarily to its symptoms.

Conclusion

The order of symptoms is important in predicting COVID-19 test results. The model that uses the order of symptoms is significantly more accurate than the time-insensitive model, and can be used as a screening tool to help triage patients.

Limitations

This study had two main limitations. First, the number of variables was relatively large, given the small sample size available. Our analysis indicates that a larger sample size may have allowed for the identification of more robust time-variant predictors, and in turn, would have achieved higher accuracy. A much larger sample size would also allow for explicitly modeling sequences of symptom occurrence.

COVID-19 test results were self-reported and established by a variety of diagnostic tests available at that time (before the widespread availability of in-home tests). This may have introduced bias in the data and consequently impacted how accurately models assessed the number of true positive and negative COVID-19 test results.

Further validation regarding the importance of symptom order is needed to establish the models' external validity, specifically for independently-collected sample data.

Acknowledgment

References

1. Rader B, Gertz A, Iuliano AD, et al. Use of At-Home COVID-19 Tests — United States, August 23, 2021–March 12, 2022. MMWR Morb Mortal Wkly Rep 2022;71:489–494. DOI: 10.15585/mmwr.mm7113e1 [PubMed: 35358168]

2. Dinnes J, Deeks JJ, Berhane S, Taylor M, Adriano A, Davenport C, Dittrich S, Emperador D, Takwoingi Y, Cunningham J, Beese S, Domen J, Dretzke J, Ferrante di Ruffano L, Harris IM, Price MJ, Taylor-Phillips S, Hooft L, Leeflang MM, McInnes MD, Spijker R, Van den Bruel A; Cochrane COVID-19 Diagnostic Test Accuracy Group. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. Cochrane Database Syst Rev. 2021 Mar 24;3(3):CD013705. doi: 10.1002/14651858.CD013705.pub2. [PubMed: 33760236]

3. Khandker SS, Nik Hashim NHH, Deris ZZ, Shueb RH, Islam MA. Diagnostic Accuracy of Rapid Antigen Test Kits for Detecting SARS-CoV-2: A Systematic Review and Meta-Analysis of 17,171 Suspected COVID-19 Patients. J Clin Med. 2021 Aug 8;10(16):3493. doi: 10.3390/jcm10163493. [PubMed: 34441789]

4. Alemi F, et al. Accuracy of Rapid Antigen Testing. Supplement to Healthcare Quality Management, in review for 2022.

5. Alemi F, Vang J, Wojtusiak J, Guralnik E, Wilson A, Peterson R, Roess A, Jain P. Differential Diagnosis of COVID-19 and Influenza, (in review, 2022).

6. Alemi F, Guralnik E, Vang J, Wojtusiak J, Wilson A, Peterson R, Roess A. Guidelines for Triage of COVID-19 Patients Presenting with Non-respiratory Symptoms. Supplement to Healthcare Quality Management, in review for 2022.

7. AlSamman M, Caggiula A, Ganguli S, Misak M, Pourmand A. Non-respiratory presentations of COVID-19, a clinical review. Am J Emerg Med. 2020 Nov;38(11):2444–2454. [PubMed: 33039218]

8. Cheng WA, Turner L, Marentes Ruiz CJ, Tanaka ML, Congrave-Wilson Z, Lee Y, Jumarang J, Perez S, Peralta A, Pannaraj PS. Clinical manifestations of COVID-19 differ by age and obesity status. Influenza Other Respir Viruses. 2021 Oct 19.

9. Hagen A How Dangerous Is the Delta Variant (B.1.617.2)? American Society for Microbiology. 2021. [(accessed on 9 October 2021)]. Available online: https://asm.org/Articles/2021/July/How-Dangerous-is-the-Delta-Variant-B-1-617-2.

10. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, Ji R, Wang H, Wang Y, Zhou Y. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. Int J Infect Dis. 2020 May;94:91–95. [PubMed: 32173574]

11. Baj J, Karakuła-Juchnowicz H, Teresi ski G, Buszewicz G, Ciesielka M, Sitarz E, Forma A, Karakuła K, Flieger W, Portincasa P, Maciejewski R. COVID-19: Specific and Non-Specific Clinical Manifestations and Symptoms: The Current State of Knowledge. J Clin Med. 2020 Jun 5;9(6):1753. [PubMed: 32516940]

12. Kronbichler A, Kresse D, Yoon S, Lee KH, Effenberger M, Shin JI. Asymptomatic patients as a source of COVID-19 infections: A systematic review and meta-analysis. Int J Infect Dis. 2020 Sep;98:180–186. [PubMed: 32562846]

13. 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]

14. Williams S, Sheard N, Stuart B, Phan HTT, Borca F, Wilkinson TMA, Burke H, Freeman A; REACT COVID Investigators. Comparisons of early and late presentation to hospital in

COVID-19 patients. Respirology. 2021 Feb;26(2):204–205. doi: 10.1111/resp.13985. Epub 2020 Dec 6. [PubMed: 33283433]

15. Alemi F Worry less about the algorithm, more about the sequence of events. Math Biosci Eng. 2020 Sep 27;17(6):6557–6572 [PubMed: 33378866]

16. Fan J, Lv J. A Selective Overview of Variable Selection in High Dimensional Feature Space. Stat Sin. 2010 Jan;20(1):101–148. [PubMed: 21572976]

17. Koch B, Vock DM, Wolfson J. Covariate selection with group lasso and doubly robust estimation of causal effects. Biometrics. 2018 Mar;74(1):8–17. doi: 10.1111/biom.12736. Epub 2017 Jun 21. [PubMed: 28636276]

18. Martens FK, Tonk EC, Kers JG, Janssens AC. Small improvement in the area under the receiver operating characteristic curve indicated small changes in predicted risks. J Clin Epidemiol. 2016 Nov;79:159–164. [PubMed: 27430730]

19. Kompaniyets L, Pennington AF, Goodman AB, Rosenblum HG, Belay B, Ko JY, Chevinsky JR, Schieber LZ, Summers AD, Lavery AM, Preston LE, Danielson ML, Cui Z, Namulanda G, Yusuf H, Mac Kenzie WR, Wong KK, Baggs J, Boehmer TK, Gundlapalli AV. Underlying Medical Conditions and Severe Illness Among 540,667 Adults Hospitalized With COVID-19, March 2020-March 2021. Prev Chronic Dis. 2021 Jul 1;18:E66. doi: 10.5888/pcd18.210123. [PubMed: 34197283]
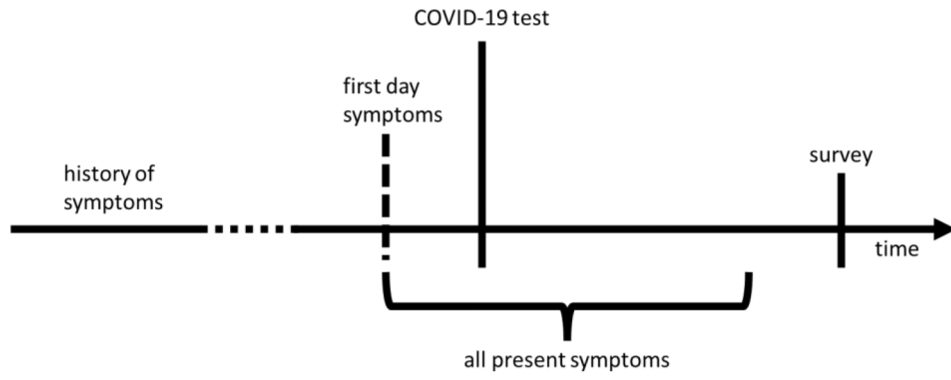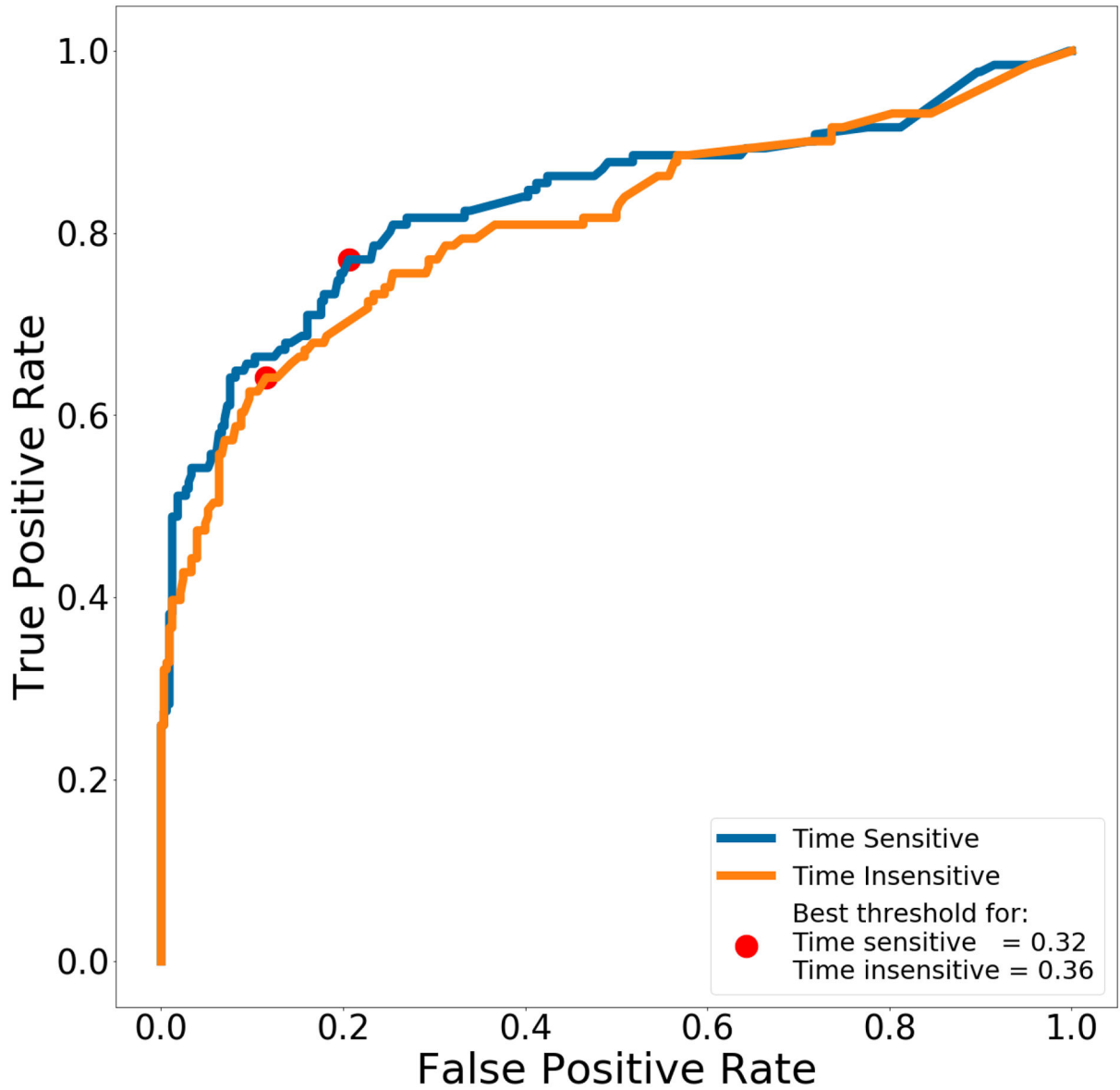
Figure 1:
Typical sequence of symptoms, COVID-19 testing and survey.
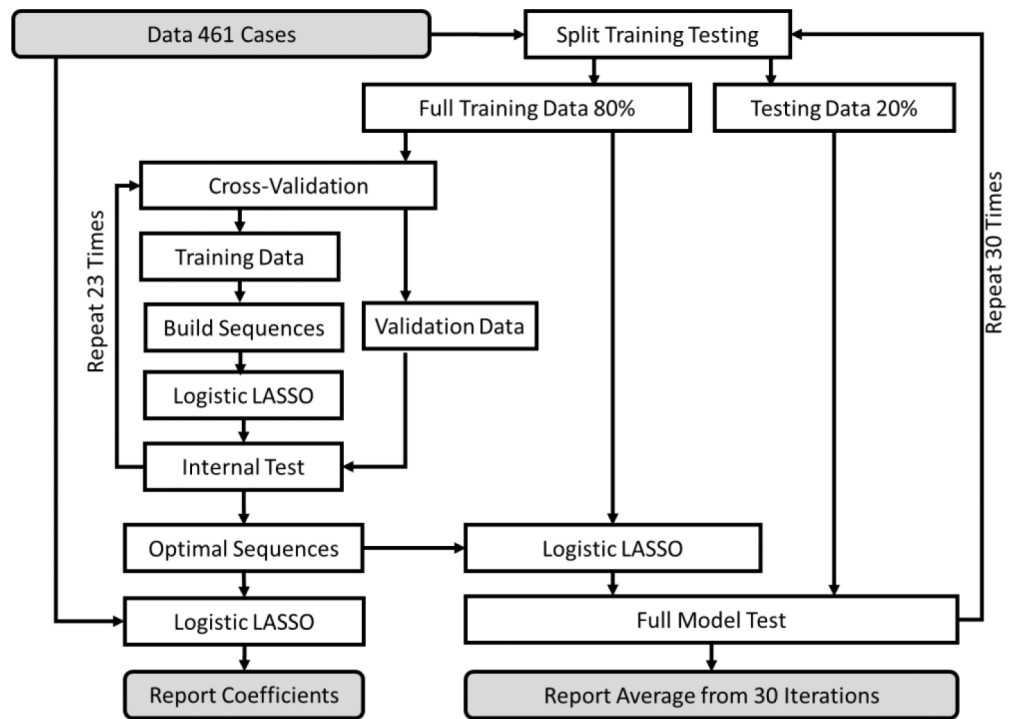
Figure 2:
Experimental design: Model Construction, Validation and Testing process

ROC curve

Figure 3:

ROC curve for time sensitive and time insensitive models

Table 1:

Characteristics of study sample

Variable	Values	Number of Cases (%)
COVID-19 Test Results	Negative	330 (68.32%)
	Positive	131 (27.12%)
	Results Pending	15 (3.11%)
	Inconclusive	7 (1.45%)
Age	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%)
Gender	Female	279 (57.76%)
	Male	203 (42.03%)
Ethnicity	Hispanic Latino	60 (12.42%)
	Non-Hispanic/Latino	401 (83.02%)
	Unknown	22 (4.55%)
Race	Other	18 (3.75%)
	Asian	25 (5.18%)
	Black or African American	60 (12.42%)
	White	380 (78.67%)

Table 2:

Distribution of History of Symptoms

Cases with	Positive COVID-19	Negative COVID-19	Likelihood Ratio	Total
History of Respiratory Symptoms	32 (24%)	66 (20%)	1.22	98 (21%)
History of Gastrointestinal Symptoms	13 (10%)	37 (11%)	0.89	50 (11%)
History of Neurological Symptoms	19 (15%)	23 (7%)	2.08	42 (9%)
History of Inflammatory Symptoms	12 (9%)	17 (5%)	1.78	29 (6%)
Number of Cases	131 (100%)	330 (100%)		461 (100%)

Table 3:

Model summary for predicting COVID-19 Test Results

	Included Time Sensitive Information	Excluded Time Sensitive Information
Average Cross-validated Area under the Receiver Operating Curve	0.799	0.784
Average Specificity	0.823	0.822
Average Sensitivity	0.685	0.642
Number of robust predictors in final model	14	11
Number of Possible Predictors	72	39

Table 4:

Robust Predictors of COVID-19 Test Results

Timing	Included Time Sensitive Information		Excluded Time Sensitive Information	
	Predictors	**Coefficients**	**Predictors**	**Coefficients**
Time-Invariant	Intercept	−2.42	Intercept	−2.21
	Age 30+	−0.22	Age 30+	−0.26
	Race White	0.26	Race White	0.11
			Headaches	1.06
			Chills	0.37
			Coughs	0.43
			Difficulty breathing	0.2
			Chest pain	0.69
			Joint pain	0.35
			Loss of appetite	0.27
			Loss of smell	0.17
			Loss of taste	0.38
Occurred at Onset of Illness	Headaches at onset	1.4		
	Chills at onset	0.4		
	Cough at onset	0.86		
	Runny nose at onset	0.15		
Occurred Later in Illness	Chest pain delayed	0.35		
	Headaches delayed	0.53		
	Loss of appetite delayed	1.15		
	Loss of smell delayed	0.49		
	Loss of taste delayed	0.32		
	Runny nose delayed	−0.12		
	Vomiting delayed	0.46		
	Wheezing delayed	0.23		