COVID

MDPI

Article

Symptoms Predicting SARS-CoV-2 Test Results in Resident Physicians and Fellows in New York City

Tania P. Chen [1,2], Meizhen Yao [2], Vishal Midya [2], Betty Kolod [2], Rabeea F. Khan [2], Adeyemi Oduwole [2], Bernard Camins [3], I. Michael Leitman [4,5], Ismail Nabeel [2], Kristin Oliver [2] and Damaskini Valvi [2,*]

1 Stanford Center for Clinical Research, Stanford University, Palo Alto, CA 94304, USA
2 Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
3 Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
4 Department of Surgery, Department of Medical Education, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
5 Department of Graduate Medical Education, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
* Correspondence: dania.valvi@mssm.edu

Abstract: Accurate prediction of SARS-CoV-2 infection based on symptoms can be a cost-efficient tool for remote screening in healthcare settings with limited SARS-CoV-2 testing capacity. We used a machine learning approach to determine self-reported symptoms that best predict a positive SARS-CoV-2 test result in physician trainees from a large healthcare system in New York. We used survey data on symptoms history and SARS-CoV-2 testing results collected retrospectively from 328 physician trainees in the Mount Sinai Health System, over the period 1 February 2020 to 31 July 2020. Prospective data on symptoms reported prior to SARS-CoV-2 test results were available from the employee health service COVID-19 registry for 186 trainees and analyzed to confirm absence of recall bias. We estimated the associations between symptoms and IgG antibody and/or reverse transcriptase polymerase chain reaction test results using Bayesian generalized linear mixed effect regression models adjusted for confounders. We identified symptoms predicting a positive SARS-CoV-2 test result using extreme gradient boosting (XGBoost). Cough, chills, fever, fatigue, myalgia, headache, shortness of breath, diarrhea, nausea/vomiting, loss of smell, loss of taste, malaise and runny nose were associated with a positive SARS-CoV-2 test result. Loss of taste, myalgia, loss of smell, cough and fever were identified as key predictors for a positive SARS-CoV-2 test result in the XGBoost model. Inclusion of sociodemographic and occupational risk factors in the model improved prediction only slightly (from AUC = 0.822 to AUC = 0.838). Loss of taste, myalgia, loss of smell, cough and fever are key predictors for symptom-based screening of SARS-CoV-2 infection in healthcare settings with remote screening and/or limited testing capacity.

Keywords: SARS-CoV-2; COVID-19; screening; physician trainees; medical residents; healthcare workers

Citation: Chen, T.P.; Yao, M.; Midya, V.; Kolod, B.; Khan, R.F.; Oduwole, A.; Camins, B.; Leitman, I.M.; Nabeel, I.; Oliver, K.; et al. Symptoms Predicting SARS-CoV-2 Test Results in Resident Physicians and Fellows in New York City. *COVID* **2023**, *3*, 671–681. https://doi.org/10.3390/covid3050049

Academic Editor: Somenath Chakraborty

Received: 24 February 2023
Revised: 15 April 2023
Accepted: 16 April 2023
Published: 25 April 2023

1. Introduction

The Coronavirus Disease 2019 (COVID-19) was first confirmed in the US by the Centers for Disease Control and Prevention (CDC) on 20 January 2020 [1,2]. New York City was one of the first epicenters in the United States [3], with the first case reported in New York State on 1 March 2020 [4]. Healthcare workers (HCWs) were at high risk for SARS-CoV-2 infection during the earliest surge of the pandemic due to direct exposure to COVID-19 patients, shortages of personal protective equipment and uncertainty about infection control protocols and containment strategies [5–7]. According to the CDC, 49,370 (16%) out of 315,531 COVID-19 cases reported in the US between 12 February and 19 April 2020 were HCWs [8].

Accurate prediction of SARS-CoV-2 infection based on symptoms can be a cost-efficient tool for remote screening in healthcare settings with limited SARS-CoV-2 testing capacity. Several studies have been undertaken to identify the combination of symptoms most predictive of COVID-19 infection, to guide precautionary self-isolation measures and to control transmission of SARS-CoV-2 [9–11]. Population-based studies have identified loss of taste or smell and fever to be strongly associated with SARS-CoV-2 infection among other reported symptoms [9,11,12]. However, a meta-analysis of 28 studies in 119,883 HCWs who tested positive for SARS-CoV-2 infection found fever being the most frequently reported symptom (27.5%), followed by cough (26.1%) and fatigue (23.4%), and substantial heterogeneity across studies conducted in China, the USA, the Netherlands, Germany and Spain [13]. Previous studies have included physicians, nurses, laboratory technicians and dentists, among other HCWs [8,13,14], but none has focused on physician trainees who are a relatively younger and healthier subgroup among the HCW population. COVID-19 infection in HCWs leads to shortages in personnel due to sick leaves and isolation during the quarantine period and recovery [15], which can hamper the quality of healthcare provided [11,15]. The early detection of symptoms and rapid testing are a critical screening strategy to control COVID-19 transmission [16]. Further research can contribute to optimizing symptom-based screening among HCW subgroups for the timely diagnosis of SARS-CoV-2 infection and the implementation of containment strategies to prevent further transmission among HCWs and the immediate community. This knowledge can enable low-resource healthcare systems to effectively initiate containment strategy protocols and the reduction of COVID-19 burden at a larger scale [11,15].

We therefore used a machine learning approach to investigate symptoms of SARS-CoV-2 infection that best predict IgG antibody and/or reverse transcriptase polymerase chain reaction test results in physician trainees from the larger healthcare system in New York City. We further examined whether prediction is more accurate when combining information about reported symptoms with other risk factors for SARS-CoV-2 infection previously identified in physician trainees, including sociodemographic and occupational risk factors [5,14]. This study advances existing knowledge about symptom-based screening of COVID-19 infection, as a useful, cost-efficient, remote screening tool in healthcare settings with limited SARS-CoV-2 testing capacity [9,17–19].

2. Materials and Methods

2.1. Study Design and Population

We conducted a retrospective cohort study of 328 physician trainees (residents and fellows) of Mount Sinai Health System (MSHS) that comprises eight hospitals in New York City and Long Island, NY. All active residents and clinical fellows from 1 January 2020 to 31 June 2020 (n = 2543) were eligible for this study. Eligible trainees were invited through email, text messages and phone calls to complete an online survey that collected information about sociodemographic, occupational and community factors related to SARS-CoV-2 infection, medical history and SARS-CoV-2 test results, as detailed previously [5]. Self-reported SARS-CoV-2 test results and prospective data on symptoms reported prior to SARS-CoV-2 testing were extracted from Mount Sinai's COVID-19 Employee Health Services (EHS) Registry. A total of 391 physician trainees responded to the survey invitation, out of which 328 trainees had undergone at least one SARS-CoV-2 test at the time of survey completion and were included in the present study. From those, 186 participants also had longitudinal data on symptomatology preceding the laboratory-confirmed SARS-CoV-2 tests available from the COVID-19 EHS registry. The study protocol was approved by the Institutional Review Board at Icahn School of Medicine at Mount Sinai. Written informed electronic consent was obtained from all study participants.

2.2. Mount Sinai Employee COVID-19 Testing and Assessment of SARS-CoV-2 Infection

On 6 March 2020, Mount Sinai's EHS established an online registry for employees to voluntarily report high-risk exposures and daily symptoms of COVID-19. RT-PCR

swabs and IgG antibody testing were available at no cost to all symptomatic employees on 7 April 2020 and to asymptomatic employees by 6 May 2020. Sensitivity and specificity of the Mount Sinai Hospital Clinical Laboratory COVID-19 ELISA antibody test were 92.5% (95% CI: 80.1–97.4%) and 100% (95% CI: 95.1–100%), respectively [5,20]. The sensitivity and specificity of the Roche Cobas RT-PCR test offered were 100% [5,21]. SARS-CoV-2 infection status was assessed by the type of test (RT-PCR, IgG antibody test or both) and whether the results were positive or negative. Among the subset of 186 study participants who had prospective data recorded in the COVID-19 EHS registry, there was 100% agreement between their SARS-CoV-2 test result reported from the laboratory compared to the self-reported SARS-CoV-2 test results collected from participants during the survey [5].

2.3. Assessment of Symptoms

Participants were asked via survey to report the months over the study period they experienced cough, chills, fever, fatigue, myalgia, headache, shortness of breath, sore throat, diarrhea, nausea/vomiting, loss of sense of smell, loss of sense of taste, malaise and runny nose. Self-reported information was collected on the presence or absence of each symptom every month from February 2020 through June 2020. These prospective data were further matched with the symptom information recorded by the COVID-19 EHS registry in real time for 186 participants, before the participants underwent a laboratory-confirmed SARS-CoV-2 test. Symptoms that were assessed in the prospective EHS registry were fever or chills, new onset persistent cough, shortness of breath, fatigue, muscle or body aches, headache, new loss of taste or smell, sore throat, new onset runny nose or nasal congestion not related to allergic rhinitis, nausea or vomiting and diarrhea. Agreement between retrospectively and prospectively collected data was 100% for all reported symptoms in the subset of 186 participants.

2.4. Assessment of Sociodemographic and Occupational Factors

The survey collected additional information regarding sociodemographic (sex, age, race) and occupational factors hypothesized to be associated with SARS-CoV-2 infection, as detailed previously [5]. Among a wide list of occupational factors examined, deployment to care for unfamiliar patient populations during the COVID-19 patient surge, assignment to in-patient medical–surgical units and training in high-risk procedural specialties were associated with increased odds for SARS-CoV-2 in this study population previously [5] and were, therefore, accounted for in the present analysis.

2.5. Statistical Analysis

All main analyses were performed on the whole data set of 328 participants who self-reported undergoing at least one type of SARS-CoV-2 test over the study period in survey responses. A schematic diagram of performed analyses is shown in Supplementary Figure S1. Differences in symptoms and sociodemographic and occupational factors between SARS-CoV-2 test result groups were examined using Fisher's exact test for categorical variables and a Wilcoxon rank-sum test for continuous variables [5]. The odds ratios (95% CI) for the associations between each symptom and SARS-CoV-2 test result were estimated using Bayesian generalized linear mixed effect regression (BGlmer). Prediction analyses were performed using an extreme gradient boosting (XGBoost) model [22] that was trained exclusively for all the symptoms experienced during the first wave. Percentage Shapley additive explanations (SHAP) [23] scores were used to show the contribution of each component in the prediction model. Two XGBoost models were examined and the accuracy between the two models was compared: (1) including only symptoms as predictors of SARS-CoV-2 test results, and (2) including symptoms and additionally sex, age, race and occupational risk factors that were associated with SARS-CoV-2 test results in previous analyses [5].

Sensitivity analysis was conducted on the data set of 186 participants who had prospective data on symptoms reported prior to SARS-CoV-2 testing and laboratory-confirmed SARS-CoV-2 test results through the EHS COVID-19 registry. Both the XGBoost model

including only symptoms and the XGBoost model including symptoms and other risk factors were run for this subset.

For all statistical analyses, *p*-values were two-sided and the level of statistical significance was set at 0.05. All analyses were conducted using SAS version 9.4 (SAS Institute, Cary, North Carolina) or R version 4.1.0. A few missing data for covariates were imputed using random forests with the "mice" R package. The prediction analysis was conducted using the XGBoost R package.

3. Results

3.1. Participant Characteristics

Study participants had a median age of 31 years (IQR (29, 33) years), 58% identified themselves as females (vs. 42% males) and 63% as of White race (vs. 25% Asians, 8% Blacks and 4% other race). A total of 66 (20%) participants had a positive SARS-CoV-2 test result over the study period. During the study period, the most common symptoms reported by study participants were fatigue (33%), cough (27%), myalgia (24%), headache (22%) and sore throat (22%). Nausea and vomiting were the least reported symptoms (Table 1).

Table 1. Self-reported symptoms, sociodemographic and occupational risk factors by SARS-CoV-2 test status in the study population.

Variable	All Participants (n = 328)	By SARS-CoV-2 Status		*p*-Value
		Negative Test Result (n = 262)	Positive Test Result (n = 66)	
		Symptoms		
Cough, n (%)				
No	238 (73)	208 (87)	30 (13)	<0.001
Yes	90 (27)	54 (60)	36 (40)	
Chills, n (%)				
No	276 (84)	241 (87)	35 (13)	<0.001
Yes	52 (16)	21 (40)	31 (60)	
Fever, n (%)				
No	269 (82)	239 (89)	30 (11)	<0.001
Yes	59 (18)	23 (39)	36 (61)	
Fatigue, n (%)				
No	220 (67)	202 (92)	18 (8)	<0.001
Yes	108 (33)	60 (56)	48 (44)	
Myalgia, n (%)				
No	250 (76)	228 (91)	22 (9)	<0.001
Yes	78 (24)	34 (44)	44 (56)	
Headache, n (%)				
No	256 (78)	219 (86)	37 (14)	<0.001
Yes	72 (22)	43 (60)	29 (40)	
Shortness of breath, n (%)				
No	284 (87)	236 (83)	48 (17)	<0.001
Yes	44 (13)	26 (59)	18 (41)	

Table 1. *Cont.*

Variable	All Participants (n = 328)	By SARS-CoV-2 Status		*p*-Value
		Negative Test Result (n = 262)	Positive Test Result (n = 66)	
Sore throat, n (%)				
No	257 (78)	210 (82)	47 (18)	0.13
Yes	71 (22)	52 (73)	19 (27)	
Diarrhea, n (%)				
No	291 (89)	238 (82)	53 (18)	0.02
Yes	37 (11)	24 (65)	13 (35)	
Nausea/vomiting, n (%)				
No	316 (96)	256 (81)	60 (19)	0.02
Yes	12 (4)	6 (50)	6 (50)	
Loss of sense of smell, n (%)				
No	283 (86)	257 (91)	26 (9)	<0.001
Yes	45 (14)	5 (11)	40 (89)	
Loss of sense of taste, n (%)				
No	291 (89)	257 (88)	34 (12)	<0.001
Yes	37 (11)	5 (14)	32 (86)	
Malaise, n (%)				
No	274 (84)	241 (88)	33 (12)	<0.001
Yes	54 (16)	21 (39)	33 (61)	
Runny nose, n (%)				
No	266 (81)	220 (83)	46 (17)	0.01
Yes	62 (19)	42 (68)	20 (32)	
Sociodemographic and Occupational Factors				
Sex, n (%)				
Female	189 (58)	155 (82)	34 (18)	0.26
Male	139 (42)	107 (77)	32 (23)	
Age, years, median (IQR)	31 (29, 33)	31 (29, 33)	30 (28, 33)	0.36
Race, n (%)				
Asian	82 (25)	71 (87)	11 (13)	0.27
Black	26 (8)	19 (73)	7 (27)	
White	202 (63)	156 (77)	46 (23)	
Other	12 (4)	10 (83)	2 (17)	
Missing	6	6	0	
Change in usual patient population, n (%)				
No	296 (90)	230 (78)	66 (22)	0.003
Yes	32 (10)	32 (100)	0 (0)	

Table 1. *Cont.*

Variable	All Participants (n = 328)	By SARS-CoV-2 Status		*p*-Value
		Negative Test Result (n = 262)	Positive Test Result (n = 66)	
Medical–surgical unit, n (%)				
No	106 (32)	89 (84)	17 (16)	0.20
Yes	222 (68)	173 (78)	49 (22)	
Training specialty, n (%)				
High-risk Primary Procedural	52 (16)	32 (62)	20 (38)	0.001
Primary Non-procedural	213 (67)	180 (85)	33 (15)	
Surgery/surgical subspecialty	53 (17)	41 (77)	12 (23)	
Missing	10	9	1	

3.2. Symptoms Associated with SARS-CoV-2 Test Result

Associations between symptoms and SARS-CoV-2 test result did not substantially differ between the crude and multivariable-adjusted BGlmer regression models (Table 2). After adjusting in the models for age, sex, race, change in usual patient population, medical–surgical unit and training specialty, 13 out of 14 symptoms were significantly associated with a positive SARS-CoV-2 test result. The strongest associations were observed for loss of taste (adjusted OR 9.77, 95% CI 9.68–9.87), loss of smell (adjusted OR 9.18, 95% CI 9.11–9.25) and fever (adjusted OR 9.17, 95% CI 2.20–38.3). Other symptoms associated with a positive SARS-CoV-2 test result were cough, chills, fatigue, myalgia, malaise and shortness of breath. The association between pharyngitis or sore throat and SARS-CoV-2 test result was not significant (OR 1.39, 95% CI 0.78–2.48) (Table 2).

Table 2. Unadjusted and adjusted effect estimates (OR, 95% CI) for the associations between self-reported symptoms and SARS-CoV-2 test status.

Symptom	Unadjusted Model		Adjusted Model [2]	
	OR [1]	95% CI [1]	OR [1]	95% CI [1]
cough	2.64	1.50, 4.67	2.99	1.87, 4.76
chills	5.57	1.26, 24.56	5.78	1.05, 31.67
fever	8.15	3.02, 22.00	9.17	2.20, 38.26
fatigue	2.23	1.42, 3.49	2.06	1.46, 2.92
myalgia	3.76	1.09, 12.90	3.38	1.02, 11.18
headache	1.66	1.08, 2.57	1.97	1.27, 3.05
shortness of breath	2.43	1.07, 5.49	3.65	1.92, 6.94
pharyngitis	1.39	0.73, 2.67	1.39	0.78, 2.48
diarrhea	1.90	1.02, 3.52	2.05	1.01, 4.14
nausea/vomiting	4.11	1.31, 12.90	6.31	1.49, 26.65
loss of smell	8.70	8.37, 9.03	9.18	9.11, 9.25

Table 2. *Cont.*

Symptom	Unadjusted Model		Adjusted Model [2]	
	OR [1]	95% CI [1]	OR [1]	95% CI [1]
loss of taste	8.80	8.71, 8.88	9.77	9.68, 9.87
malaise	6.60	3.65, 11.95	4.24	1.20, 14.97
runny nose	1.29	1.01, 1.64	1.54	1.05, 2.25

[1] Abbreviations: OR, odds ratio; CI, confidence interval. [2] All adjusted models were adjusted for sex, age, race, change in usual patient population, medical surgical unit and training specialty.

3.3. Symptoms Predicting a Positive SARS-CoV-2 Test Result in the XGBoost Model

In the prediction model including only symptoms, loss of sense of taste, myalgia and loss of sense of smell were the top three predictors of a positive SARS-CoV-2 test result. Other predictors which ranked high were cough and fever (Figure 1). Our symptoms-exclusive prediction model for 328 trainees had an accuracy of 0.878, a sensitivity of 0.922, a specificity of 0.722, a positive predictive value (PPV) of 0.922, a negative predictive value (NPV) of 0.722 and an area under the curve (AUC) of 0.822. Detailed statistics of the prediction model can be found in Supplementary Table S1. Inclusion of sociodemographic and occupational risk factors of SARS-CoV-2 infection in the prediction model did not substantially change the distribution of importance of each symptom, with the highest ranking remaining for loss of sense of smell, myalgia and loss of sense of taste, and increased the accuracy to 0.915, sensitivity to 0.953, specificity to 0.778, PPV to 0.939, NPV to 0.824 and AUC to 0.838 (Supplementary Figure S2 and Supplementary Table S1).

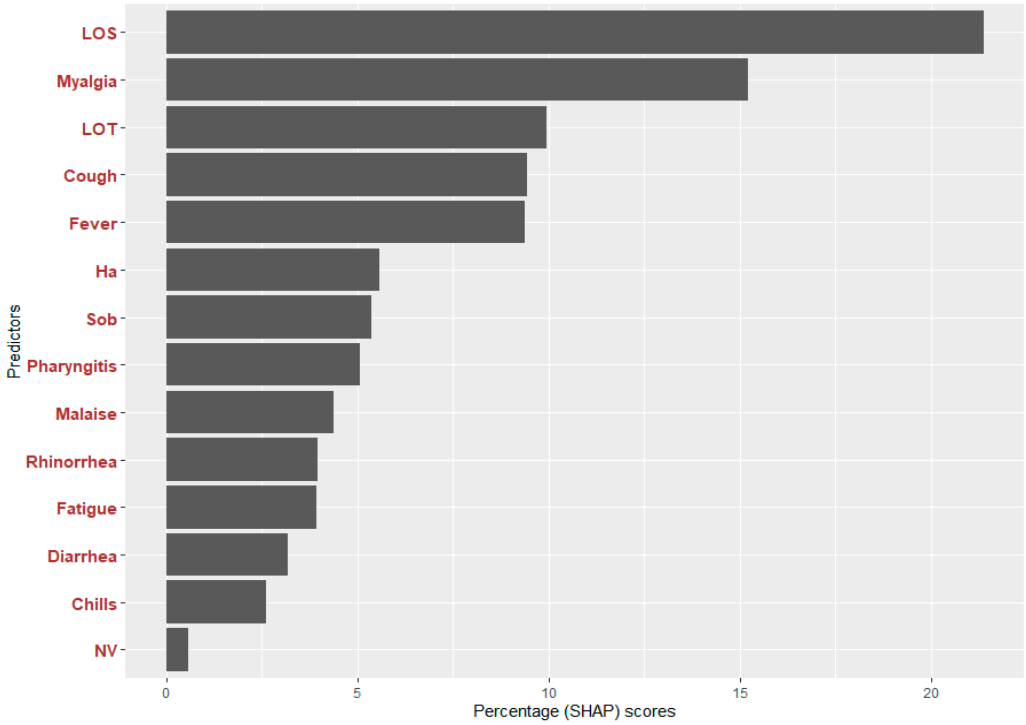

Figure 1. Importance of each predictor's contribution using percentage (SHAP) scores in the prediction model of 328 participants, including only symptoms. Abbreviations: LOS, loss of sense of smell; LOT, loss of sense of taste; Ha, headache; Sob, shortness of breath; NV, nausea/vomiting.

3.4. Sensitivity Analysis

Restricted analysis of the 186 study participants with prospective EHS COVID-19 registry data showed similar results with loss of smell and myalgia among the top predictors for positive SARS-CoV-2 test results (Supplementary Figure S3). The statistics of the

prediction models for EHS data can be found in Supplementary Table S1. The prediction model that included symptoms along with other risk factors had slightly better performance using the EHS data set, as was also observed in the main analyses of 328 participants (Supplementary Figure S4).

4. Discussion

In this study of residents and fellows from a large healthcare center in New York City, we found that self-reported symptoms-based screening alone can accurately predict a positive SARS-CoV-2 test result. Among a wide list of symptoms associated with SARS-CoV-2 infection examined, loss of smell, myalgia, loss of taste, cough and fever were found to be top predictors of a positive SARS-CoV-2 test result. Inclusion in the prediction models of sociodemographic (sex, age, race) and occupational risk factors previously shown to increase risk of SARS-CoV-2 infection did not substantially change results, but slightly increased prediction accuracy, suggesting that the combination of symptoms with other potentially known risk factors could further optimize screening of SARS-CoV-2 infection of physician trainees in healthcare settings with remote screening and/or limited testing capacity.

A previous population-based prospective cohort study in Spain using a machine learning approach noted olfactory dysfunction, gustatory dysfunction, fever, dry cough and asthenia (weakness) to be strong predictors of a positive SARS-CoV-2 RT-PCR result; but no association between dyspnea, rhinorrhea and sore throat and a positive test result [24]. Another study analyzed about 42 prospective SARS-CoV-2 studies and also demonstrated that anosmia, ageusia, fatigue, fever and cough were associated with higher odds for SARS-CoV-2 infection [25]. Moreover, they noted that combining symptoms with other sociodemographic (age, gender, etc.) or community risk factors (e.g., travel history) may slightly improve the sensitivity of the prediction model [25]. In our study of young HCWs we observed similar findings, in addition to shortness of breath (dyspnea) and runny nose (rhinorrhea) that were significantly associated with SARS-CoV-2 infection. A meta-analysis of HCW studies also found the occurrence of lack of smell, fever and myalgia to be associated with higher odds of SARS-CoV-2 infection in symptomatic patients, and no significant association for fatigue and sore throat [26]. However, our results demonstrated association between fatigue and a positive SARS-CoV-2 infection in addition to lack of smell, fever and myalgia. We did not find sore throat or pharyngitis to be associated with a SARS-CoV-2 infection, which is in agreement with prior evidence [24–27]. This previous meta-analysis only analyzed the abovementioned five symptoms in association with SARS-CoV-2 infection due to limited data available on symptoms reported in previous studies [26]. One previous study in the UK and USA of 18,401 participants that used smartphone-based apps for symptoms screening also found loss of smell, loss of taste, high temperature, persistent cough and loss of appetite as the top predictors of SARS-CoV-2 infection [9]. In our study, we did not assess loss of appetite, but we identified loss of smell, loss of taste, fever and cough as top predictors of a positive SARS-CoV-2 test in physician trainees. Additionally, results from a few other recent symptom-based COVID-19 screening studies further support our findings that loss of smell, loss of taste, fever, cough and myalgia are important predictors of SARS-CoV-2 infection [11,24,27].

Findings from our study remained robust in sensitivity analyses of a subset of 186 trainees with prospective, real-time data on symptoms reported prior to laboratory-confirmed SARS-CoV-2 test results available from the EHS COVID-19 registry data and, therefore, reverse causation bias is unlikely. Furthermore, we found perfect agreement (100%) between self-reports of SARS-CoV-2 test results and laboratory-confirmed SARS-CoV-2 test results in physician trainees with no evidence of recall bias in survey responses during the study period. Our study sample had a similar age range and race and specialty distributions compared to the total population of eligible residents and fellows for the present study, and therefore results should be more broadly representative of the origin cohort of trainees [5]. Study limitations include the lack of data on loss of appetite previously reported as a potentially important predictor of SARS-CoV-2 infection [9]. We further

assessed symptoms predicting a positive SARS-CoV-2 test result during the first COVID-19 wave and prior to vaccination campaigns. Other factors related to SARS-CoV-2 infection such as specific variants or vaccination status might impact the prediction of a positive SARS-CoV-2 test result. Further research is needed to validate our findings in recent waves with new SARS-CoV-2 variants and after vaccination.

5. Conclusions

Our study focused specifically on a young and generally healthy group of residents and fellows of a large healthcare system and found that loss of smell, myalgia, loss of taste, cough and fever can serve as important predictors for symptom-based screening of SARS-CoV-2 infection in healthcare settings with limited testing capacity. Moreover, the predictive value of this method can further be enhanced by inclusion of other risk factors, such as sociodemographic and occupational risk factors known to be associated with SARS-CoV-2 infection risk in HCW populations. These findings can be helpful in certain health centers with remote screening and testing shortages, whenever IgG antibody or a reverse transcriptase polymerase chain reaction test is not available.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/covid3050049/s1, Figure S1: Flow chart of all the main analyses performed; Figure S2: Importance of each predictor's contribution using percentage (SHAP) scores in the prediction model among whole data set (n = 328), including symptoms and other covariates; Figure S3: Comparing the importance of each predictor's contribution using percentage (SHAP) scores in the two prediction models among EHS data set; Figure S4: ROC comparison of prediction models using different predictors and different data set; Table S1: Comparison of statistics in three prediction models.

Author Contributions: T.P.C., V.M., K.O. and D.V. conceived and designed the study. T.P.C., M.Y., V.M. and D.V. had complete access to the data. M.Y. performed the statistical analyses with guidance from V.M. and D.V. T.P.C. wrote the first manuscript draft under the supervision of D.V. I.N. oversaw the implementation of Mount Sinai Health System's Employee Health COVID-19 registry and facilitated interpretation of the analysis. B.C. and I.N. provided information on Mount Sinai Health System's SARS-CoV-2 testing and infection prevention protocols. I.M.L. facilitated data access. All authors (T.P.C., M.Y., V.M., B.K., R.F.K., A.O., B.C., I.M.L., I.N., K.O. and D.V.) contributed to data interpretation, critically revised the manuscript for intellectual content and approved its submission. D.V. supervised all stages of this study and assumes responsibility for the integrity of the data and the accuracy of the analysis. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Institute of Environmental Health Studies (P30ES023515).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the Icahn School of Medicine at Mount Sinai.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available by request from the corresponding author. The data are not publicly available due to privacy restrictions.

Acknowledgments: We are deeply grateful to all Mount Sinai Health System resident physicians and fellows for their selfless dedication in responding to the COVID-19 pandemic and to all study participants for their generous contribution to this research.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. CDC Confirms First Case of Coronavirus in the United States. 2020. Available online: https://www.cbsnews.com/news/coronavirus-centers-for-disease-control-first-case-united-states/ (accessed on 26 October 2021).
2. Holshue, M.L.; DeBolt, C.; Lindquist, S.; Lofy, K.H.; Wiesman, J.; Bruce, H.; Spitters, C.; Ericson, K.; Wilkerson, S.; Tural, A.; et al. First Case of 2019 Novel Coronavirus in the United States. *N. Engl. J. Med.* **2020**, *382*, 929–936. [CrossRef] [PubMed]
3. Kopecki, D. New York City Confirms First Coronavirus Case. CNBC. Available online: https://www.cnbc.com/2020/03/01/first-coronavirus-case-confirmed-in-new-york-city.html (accessed on 2 March 2021).
4. At Novel Coronavirus Briefing, Governor Cuomo Declares State of Emergency to Contain Spread of Virus. 2020. Available online: https://www.governor.ny.gov/news/novel-coronavirus-briefing-governor-cuomo-declares-state-emergency-contain-spread-virus (accessed on 26 October 2021).
5. Pawloski, K.R.; Kolod, B.; Khan, R.F.; Midya, V.; Chen, T.; Oduwole, A.; Camins, B.; Colicino, E.; Leitman, I.M.; Nabeel, I.; et al. Factors Associated with SARS-CoV-2 Infection in Physician Trainees in New York City during the First COVID-19 Wave. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5274. [CrossRef]
6. Nguyen, L.H.; Drew, D.A.; Graham, M.S.; Joshi, A.D.; Guo, C.G.; Ma, W.; Mehta, R.S.; Warner, E.T.; Sikavi, D.R.; Lo, C.H.; et al. Risk of COVID-19 among front-line health-care workers and the general community: A prospective cohort study. *Lancet Public Health* **2020**, *5*, e475–e483. [CrossRef] [PubMed]
7. Kim, H.; Hegde, S.; LaFiura, C.; Raghavan, M.; Sun, N.; Cheng, S.; Rebholz, C.M.; Seidelmann, S.B. Access to personal protective equipment in exposed healthcare workers and COVID-19 illness, severity, symptoms and duration: A population-based case-control study in six countries. *BMJ Glob. Health* **2021**, *6*, e004611. [CrossRef]
8. CDCMMWR. Characteristics of Health Care Personnel with COVID-19—United States, February 12–April 9, 2020. *MMWR Morb. Mortal. Wkly. Rep.* **2020**, *69*, 477. [CrossRef]
9. Menni, C.; Valdes, A.M.; Freidin, M.B.; Sudre, C.H.; Nguyen, L.H.; Drew, D.A.; Ganesh, S.; Varsavsky, T.; Cardoso, M.J.; El-Sayed Moustafa, J.S.; et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat. Med.* **2020**, *26*, 1037–1040. [CrossRef]
10. Rossman, H.; Keshet, A.; Shilo, S.; Gavrieli, A.; Bauman, T.; Cohen, O.; Shelly, E.; Balicer, R.; Geiger, B.; Dor, Y.; et al. A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. *Nat. Med.* **2020**, *26*, 634–638. [CrossRef] [PubMed]
11. Ramírez Varela, A.; Moreno López, S.; Contreras-Arrieta, S.; Tamayo-Cabeza, G.; Restrepo-Restrepo, S.; Sarmiento-Barbieri, I.; Caballero-Díaz, Y.; Jorge Hernandez-Florez, L.; Mario González, J.; Salas-Zapata, L.; et al. Prediction of SARS-COV-2 infection with a symptoms-based model to aid public health decision making in Latin America and other low and middle income settings. *Prev. Med. Rep.* **2022**, *27*, 101798. [CrossRef] [PubMed]
12. Iversen, K.; Bundgaard, H.; Hasselbalch, R.B.; Kristensen, J.H.; Nielsen, P.B.; Pries-Heje, M.; Knudsen, A.D.; Christensen, C.E.; Fogh, K.; Norsk, J.B.; et al. Risk of COVID-19 in health-care workers in Denmark: An observational cohort study. *Lancet Infect. Dis.* **2020**, *20*, 1401–1408. [CrossRef]
13. Gholami, M.; Fawad, I.; Shadan, S.; Rowaiee, R.; Ghanem, H.; Hassan Khamis, A.; Ho, S.B. COVID-19 and healthcare workers: A systematic review and meta-analysis. *Int. J. Infect. Dis.* **2021**, *104*, 335–346. [CrossRef]
14. Baker, J.M.; Nelson, K.N.; Overton, E.; Lopman, B.A.; Lash, T.L.; Photakis, M.; Jacob, J.T.; Roback, J.D.; Fridkin, S.K.; Steinberg, J.P. Quantification of Occupational and Community Risk Factors for SARS-CoV-2 Seropositivity Among Health Care Workers in a Large U.S. Health Care System. *Ann. Intern. Med.* **2021**, *174*, 649–654. [CrossRef] [PubMed]
15. Dzinamarira, T.; Murewanhema, G.; Mhango, M.; Iradukunda, P.G.; Chitungo, I.; Mashora, M.; Makanda, P.; Atwine, J.; Chimene, M.; Mbunge, E.; et al. COVID-19 Prevalence among Healthcare Workers. A Systematic Review and Meta-Analysis. *Int. J. Environ. Res. Public Health* **2021**, *19*, 146. [CrossRef] [PubMed]
16. Hashmi, H.A.S.; Asif, H.M. Early Detection and Assessment of Covid-19. *Front. Med.* **2020**, *7*, 311. [CrossRef]
17. Bastiani, L.; Fortunato, L.; Pieroni, S.; Bianchi, F.; Adorni, F.; Prinelli, F.; Giacomelli, A.; Pagani, G.; Maggi, S.; Trevisan, C.; et al. Rapid COVID-19 Screening Based on Self-Reported Symptoms: Psychometric Assessment and Validation of the EPICOVID19 Short Diagnostic Scale. *J. Med. Internet Res.* **2021**, *23*, e23897. [CrossRef]
18. Marcus, G.M.; Olgin, J.E.; Peyser, N.D.; Vittinghoff, E.; Yang, V.; Joyce, S.; Avram, R.; Tison, G.H.; Wen, D.; Butcher, X.; et al. Predictors of incident viral symptoms ascertained in the era of COVID-19. *PLoS ONE* **2021**, *16*, e0253120. [CrossRef]
19. Kennedy, B.; Fitipaldi, H.; Hammar, U.; Maziarz, M.; Tsereteli, N.; Oskolkov, N.; Varotsis, G.; Franks, C.A.; Nguyen, D.; Spiliopoulos, L.; et al. App-based COVID-19 syndromic surveillance and prediction of hospital admissions in COVID Symptom Study Sweden. *Nat. Commun.* **2022**, *13*, 2110. [CrossRef]
20. Center for Devices and Radiological Health. EUA Authorized Serology Test Performance. U.S. Food and Drug Administration. 2020. Available online: https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/eua-authorized-serology-test-performance (accessed on 26 July 2022).
21. Covid-19 Diagnostics in Context-Harvard University. 2020. Available online: https://csb.mgh.harvard.edu/covid (accessed on 26 July 2022).
22. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. *arXiv* **2022**, arXiv:1603.02754.
23. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef]

24. Azeli, Y.; Fernández, A.; Capriles, F.; Rojewski, W.; Lopez-Madrid, V.; Sabaté-Lissner, D.; Serrano, R.M.; Rey-Reñones, C.; Civit, M.; Casellas, J.; et al. A machine learning COVID-19 mass screening based on symptoms and a simple olfactory test. *Sci. Rep.* **2022**, *12*, 15622. [CrossRef]
25. Struyf, T.; Deeks, J.J.; Dinnes, J.; Takwoingi, Y.; Davenport, C.; Leeflang, M.M.; Spijker, R.; Hooft, L.; Emperador, D.; Domen, J.; et al. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. *Cochrane Database Syst. Rev.* **2022**, *5*, CD013665. [CrossRef]
26. Gómez-Ochoa, S.A.; Franco, O.H.; Rojas, L.Z.; Raguindin, P.F.; Roa-Díaz, Z.M.; Wyssmann, B.M.; Guevara, S.L.R.; Echeverría, L.E.; Glisic, M.; Muka, T. COVID-19 in Health-Care Workers: A Living Systematic Review and Meta-Analysis of Prevalence, Risk Factors, Clinical Characteristics, and Outcomes. *Am. J. Epidemiol.* **2021**, *190*, 161–175. [CrossRef] [PubMed]
27. Savoy, M.; Kopp, B.; Chaouch, A.; Cohidon, C.; Gouveia, A.; Lombardo, P.; Maeder, M.; Payot, S.; Perdrix, J.; Schwarz, J.; et al. Diagnostic performance of individual symptoms to predict SARS-COV-2 RT-PCR positivity and symptom persistence among suspects presenting in primary care during the first wave of covid-19. *Infect. Dis. Rep.* **2023**, *15*, 112–124. [CrossRef] [PubMed]