Title: Proposed Clinical Indicators for Efficient Screening and Testing for COVID-19 Infection from Classification and Regression Trees (CART) Analysis

Authors: Richard K. Zimmerman, MD, MPH, MA,1 Mary Patricia Nowalk, PhD, RDN,1 Todd Bear, PhD,2 Rachel Taber, PhD,2 Theresa M. Sax, BS,3 Heather Eng, BA,3 GK. Balasubramani, PhD3

Affiliations: University of Pittsburgh School of Medicine, Department of Family Medicine1 and Graduate School of Public Health, Department of Behavioral and Community Health Sciences2 and Department of Epidemiology,3 Pittsburgh PA USA

Financial support. This study was supported through a cooperative agreement with the Centers for Disease Control and Prevention (CDC) through grant number U01 IP000467 and the National Institutes of Health grant number 1UL1 TR001857. The US Flu VE Network is supported through cooperative agreements funded by CDC. The findings and conclusions are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. It is subject to the CDC’s Open Access Policy.

Address correspondence to: Richard K. Zimmerman, MD, Suite 520 Schenley Place, 4420 Bayard Street, Pittsburgh PA 15260. Telephone: 412/383-2354; email: zimmrk@upmc.edu.

Word Count: Abstract – 255; Text – 603; Table – 1; Figure – 1

Acknowledgements/Potential Conflict of Interest: Drs. Zimmerman, Nowalk, and Balasubramani and Ms. Eng and Ms. Sax have research funding from Merck & Co, Inc. for unrelated projects. Dr. Zimmerman has funding from Sanofi for an unrelated project. Drs. Bear and Taber have no conflicts to report.
Abstract

**Background:** The introduction and rapid transmission of SARS CoV2 in the United States resulted in implementation of methods to assess, mitigate and contain the resulting COVID-19 disease based on limited knowledge. Screening for testing has been based on symptoms typically observed in inpatients, yet outpatient symptom complexes may differ.

**Methods:** Classification and regression trees (CART) recursive partitioning created a decision tree classifying enrollees into laboratory-confirmed cases and non-cases. Demographic and symptom data from patients ages 18-87 years who were enrolled from March 29-April 26, 2020 were included. Presence or absence of SARS-CoV2 was the target variable.

**Results:** Of 736 tested, 55 were positive for SARS-CoV2. Cases significantly more often reported chills, loss of taste/smell, diarrhea, fever, nausea/vomiting and contact with a COVID-19 case, but less frequently reported shortness of breath and sore throat. A 7-terminal node tree with a sensitivity of 96% and specificity of 53%, and an AUC of 78% was developed. The positive predictive value for this tree was 14% while the negative predictive value was 99%. Almost half (44%) of the participants could be ruled out as likely non-cases without testing.

**Discussion:** Among those referred for testing, negative responses to three questions could classify about half of tested persons with low risk for SARS-CoV2 and would save limited testing resources. These questions are: was the patient in contact with a COVID-19 case? Has the patient experienced 1) a loss of taste or smell; or 2)
nausea or vomiting? The outpatient symptoms of COVID-19 appear to be broader than the well-known inpatient syndrome.
Introduction

The introduction of the SARS CoV2 virus into the United States was earlier than anticipated. Furthermore, the rapid transmission of SARS CoV2 resulted in implementation of methods to assess, mitigate and contain the resulting COVID-19 disease that were based on limited knowledge of its epidemiology. Given potentially insufficient testing supplies, but the need to rapidly identify cases, access to testing was based on symptoms typical of an acute respiratory infection, such as cough, fever, and shortness of breath, as has been reported to be emblematic of COVID-19 infection.\(^1\) A previous report noted that use of cough, fever, shortness of breath and sore throat as the sole screening criteria for COVID-19 infection among health care personnel might have missed 17% of symptomatic cases.\(^2\) Gastrointestinal and non-acute respiratory illness-typical symptoms have been noted.\(^1\) Outpatient symptom complexes may differ from the better-known inpatient set. A data-driven set of clinical indicators for COVID-19 would help to identify outpatient symptoms and those who most benefit from limited testing availability.

Methods

CART recursive partitioning, based on presence or absence of symptoms, was used to create a decision tree to correctly classify enrollees into laboratory-confirmed (RT-PCR) COVID-19 cases.\(^3\) Demographic and symptom data from 736 patients, ages 18-87 years tested for presence of SARS-CoV2 virus from March 29-April 26, 2020 were included.

Symptoms reported at enrollment -- fever, chills, cough, sore throat, shortness of breath, muscle aches, abdominal pain, nausea/vomiting, diarrhea, headache, decrease
or loss of taste or smell and contact with COVID-19 case -- were included in the CART model as categorical predictors. Presence or absence of SARS-CoV2 was the target variable. The tree was built on a ≥10% sample for the parent node and a stopping rule of ≥5% of the sample in the terminal node. A tenfold cross-validation method was used to evaluate reliability. To avoid overfitting, a maximum acceptable difference in risk between the pruned and the sub-tree of one standard error was used for tree pruning. Missing data were handled by surrogate splits. Hosmer-Lemeshow goodness of fit test confirmed the suitability of the trees. The area under the curve (AUC) for receiver operating characteristics, sensitivity, specificity, positive and negative predictive values were assessed. The University of Pittsburgh IRB approved this study.

Results

Of 736 tested, 55 were positive for SARS-CoV2. Those who tested positive significantly more frequently reported chills, loss of taste/smell, diarrhea, fever, nausea/vomiting and contact with a COVID-19 case, but less frequently reported shortness of breath and sore throat (Table). The CART analyses resulted in a 7-terminal node tree with a sensitivity of 96% and specificity of 53%, resulting in an AUC of 78%. The positive predictive value for this tree was 14% while the negative predictive value was 99%. Of clinical utility, three variables (contact with a COVID-19 case, report of loss or diminution of taste or smell and report of nausea or vomiting; followed down the left side of the Figure) would allow almost half (44%) of the participants to be ruled out as likely non-cases without testing.

Discussion
Among those referred for testing, we found negative responses to three questions could classify about half of tested persons who had a minute risk for SARS-CoV2 and would save limited testing resources. These questions are: was the patient in contact with a COVID-19 case? Has the patient experienced 1) a loss of taste or smell; or 2) nausea or vomiting? If COVID-19 case exposure did not occur, and the patient reported neither loss of taste or smell nor nausea or vomiting the patient need not be tested. Alternatively, a clinical decision rule can be electronically programmed to make use of such data. Finally, the outpatient symptoms of COVID-19 appear to be broader than the well-known inpatient syndrome.

References

1. Centers for Disease Control and Prevention. National Center for Immunization and Respiratory Diseases (NCIRD), Division of Viral Diseases


## Table. Presenting Symptoms for Outpatients Attending the COVID-19 Testing Clinic

<table>
<thead>
<tr>
<th>Symptom</th>
<th>COVID-19 cases</th>
<th>Non-cases</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=55 (7.5%)</td>
<td>N=681 (92.5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td></td>
</tr>
<tr>
<td>Abdominal pain</td>
<td>20%</td>
<td>27%</td>
<td>0.286</td>
</tr>
<tr>
<td>Chills</td>
<td>80%</td>
<td>64%</td>
<td>0.016</td>
</tr>
<tr>
<td>Cough</td>
<td>85%</td>
<td>87%</td>
<td>0.662</td>
</tr>
<tr>
<td>Decrease or loss of taste or smell</td>
<td>72%</td>
<td>25%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>53%</td>
<td>38%</td>
<td>0.028</td>
</tr>
<tr>
<td>Fever</td>
<td>86%</td>
<td>68%</td>
<td>0.006</td>
</tr>
<tr>
<td>Headache</td>
<td>86%</td>
<td>82%</td>
<td>0.497</td>
</tr>
<tr>
<td>Muscle aches</td>
<td>66%</td>
<td>67%</td>
<td>0.849</td>
</tr>
<tr>
<td>Nausea/Vomiting</td>
<td>20%</td>
<td>10%</td>
<td>0.013</td>
</tr>
<tr>
<td>Shortness of breath</td>
<td>53%</td>
<td>66%</td>
<td>0.046</td>
</tr>
<tr>
<td>Sore throat</td>
<td>39%</td>
<td>66%</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Contact with COVID-19 case</td>
<td>70%</td>
<td>21%</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
Figure

Decision Tree for Symptoms to Determine Testing Need for Outpatient COVID-19

N = 736

Contact with confirmed COVID case?

Yes N = 211
83% negative 17% positive
Recommend testing

No N = 525

Loss of taste/smell?

No N = 362

Nausea?

No N = 325
99% negative 1% positive
No testing

Yes N = 37
89% negative 11% positive
Consider testing

Body Aches?

Yes N = 163
87% negative 13% positive
Consider testing

No N = 55

Shortness of breath?

No N = 86

Chills?

Yes N = 70
90% negative 10% positive
Consider testing

No N = 16
100% negative
No testing