Estimating the efficacy of symptom-based screening for COVID-19

Alison Callahan1,3, Ethan Steinberg1,3, Jason A. Fries1,3, Saurabh Gombar4, Birju Patel5, Conor K. Corbin1 and Nigam H. Shah1

There is substantial interest in using presenting symptoms to prioritize testing for COVID-19 and establish symptom-based surveillance. However, little is currently known about the specificity of COVID-19 symptoms. To assess the feasibility of symptom-based screening for COVID-19, we used data from tests for common respiratory viruses and SARS-CoV-2 in our health system to measure the ability to correctly classify virus test results based on presenting symptoms. Based on these results, symptom-based screening may not be an effective strategy to identify individuals who should be tested for SARS-CoV-2 infection or to obtain a leading indicator of new COVID-19 cases.

npj Digital Medicine (2020) 3:95 ; https://doi.org/10.1038/s41746-020-0300-0

INTRODUCTION

There is substantial interest in developing symptom-based screening to prioritize who should be tested for SARS-CoV-2 infection and to establish symptom-based surveillance to provide an early indicator of new COVID-19 cases1–3. However, the degree to which presenting symptoms are reliable indicators of SARS-CoV-2 infection is unknown4. Therefore, it is crucial to determine whether symptom-based screening to prioritize testing is feasible. To assess the feasibility of using symptom-based screening to assign a probability of SARS-CoV-2 infection, we first quantified the ability to correctly predict results of tests for common respiratory viruses observed to frequently co-infect patients positive for SARS-CoV-2 at Stanford Health Care5, using symptoms mentioned in clinical notes at the time of the test order. After establishing a baseline for the performance of machine learning models to correctly classify common respiratory virus infections6, we then trained a similar model for SARS-CoV-2 test results5.

RESULTS AND DISCUSSION

Performance of models to predict respiratory virus test results

For the respiratory viruses examined, area under the receiver operator curve (AUROC) on the test set ranged from 0.60 to 0.77 (Table 1). Two non-SARS-CoV-2 viruses (influenza type A, and RSV) were moderately predictable given presenting symptoms. For example, mentions of coughing, wheezing and rhinorrhea were features with high importance for the RSV model. However, SARS-CoV-2 and the remaining common respiratory viruses (adenovirus, rhinovirus, metapneumovirus and parainfluenza) were not highly predictable, with average AUROCs below 0.70.

These results suggest that, for both SARS-CoV-2 and other commonly diagnosed respiratory viral infections, the presenting symptoms at the time of the test order may not provide sufficient information to correctly classify whether a given patient will test positive for that virus. Prior studies of presenting symptoms and case definitions for influenza found that information on presenting symptoms alone is not sufficient to accurately diagnose influenza, or distinguish it from other influenza-like illnesses8–12, and our results support this finding. Though our Influenza Virus A model had one of the higher AUROCs we observed, it is not sufficient for use in a clinical setting.

Limitations

There are several limitations to this work. Firstly, we did not include information on the duration of reported symptoms as features in our models, because duration is often not reliably described in clinical notes and therefore difficult to ascertain13. The sensitivity and specificity of SARS-CoV-2 PCR tests is highest in the first few days that symptoms present14. It is possible that the data used in our analysis included patients tested later in the course of SARS-CoV-2 infection, which would diminish classifier performance. Secondly, the number of SARS-CoV-2 positive cases in our data was small, which is fortunate for our population's health, but creates a sample size limitation which can adversely impact machine learning model performance. Lastly, the prevalence of positive cases in those tested for SARS-CoV-2 infection is dependent on the health system's protocols used to decide whether to test a patient; this may impact the applicability of our findings as testing protocols evolve.

We note that clinicians' knowledge of the signs and symptoms of COVID-19 are rapidly evolving, such as recent findings that a substantial fraction of patients experience gastrointestinal symptoms15 and dermatologic symptoms16, and the prevalence of these presentations are still being characterized. The CDC has recently updated their list of COVID-19 related symptoms to include loss of taste or smell, headache, and chills with fever17. Documentation of such emerging symptoms will increase in the clinical notes of tested patients. Therefore, as part of Stanford Health Care's response to COVID-19, we continue to collect data on patients tested for SARS-CoV-2 and to profile their presenting symptoms8,18. Doing so will allow us to assess the effect of improved symptom characterization and additional data on the performance of models to identify SARS-CoV-2 infections in an ongoing manner.

Summary

In summary, our current findings indicate that symptom-based screening may not be an effective strategy to quantify an
individual’s likelihood of having COVID-19. The non-specific nature of the symptoms, and the fact that co-infections with other respiratory viruses are common⁵, might limit the utility of symptom-based screening strategies to prioritize testing and the use of symptom surveys as a leading indicator of new COVID-19 cases in a region¹⁴⁻¹⁵.

**METHODS**

Patient cohort selection

Patients included in our study were those tested for adenovirus, influenza virus A, metapneumovirus, parainfluenza virus, respiratory syncytial virus (RSV), rhinovirus, and SARS-CoV-2 (Table 2). Historical data on patients tested for other respiratory viruses were collected between September 2019 and March 30, 2020, and also included results of tests ordered as part of respiratory pathogen panels⁵. Only the contents of the emergency department or urgent care panels associated with the order of a patient’s virus test was used as input for training the models, in order to emulate the information that would be available in a real-life usage setting.

Feature engineering

Each patient’s test-associated note was processed using a rule-based NLP pipeline⁶ to extract mentions of medical concepts. Extracted concepts were classified to filter out negated terms, to identify relative timing of the concept (e.g. a current condition or past condition), and to identify note sections, in order to filter out concepts that were not noted at initial observation but added later in the course of documentation. All concepts were derived from the 2018AA SNOMED CT US vocabulary with the semantic group DISORDER⁵, maintained as part of the National Library of Medicine Unified Medical Language System.

Extracted concepts were encoded as binary variables to indicate their presence or absence in a given patient’s clinical note, after filtering to keep only the non-historical, positive mentions about the patient and restricting to signs and symptoms based on SNOMED CT US semantic types. The outcome was whether the virus test returned positive or negative for infection with the tested pathogen.

---

**Table 1. AUROCs of logistic regression models for each respiratory virus with 95% confidence intervals.**

| Virus             | AUROC      |
|-------------------|------------|
| Adenovirus        | 0.68 (0.60–0.76) |
| Influenza virus A | 0.73 (0.68–0.77) |
| Metapneumovirus   | 0.64 (0.57–0.71) |
| Parainfluenza virus | 0.60 (0.53–0.68) |
| RSV               | 0.77 (0.73–0.80) |
| Rhinovirus        | 0.62 (0.58–0.66) |
| SARS-CoV-2        | 0.64 (0.49–0.79) |

**Table 2. Total number of patients and number of positive and negative test results used to develop models for each virus test.**

| Virus             | Patients | Positive tests | Negative tests |
|-------------------|----------|---------------|----------------|
| Adenovirus        | 10,911   | 218           | 10,693         |
| Influenza virus A | 8642     | 980           | 7662           |
| Metapneumovirus   | 9504     | 360           | 9144           |
| Parainfluenza virus | 9507   | 351           | 9156           |
| RSV               | 13,525   | 1059          | 12,466         |
| Rhinovirus        | 8305     | 1387          | 6918           |
| SARS-CoV-2        | 895      | 64            | 831            |

**REFERENCES**

1. Goode, L. Facebook and Google survey data may help map Covid-19’s spread. Wired. https://www.wired.com/2020/03/social-media-surveys-facebook-google-map-covid-19/

2. Gostic, K., Gomez, A. C., Mummah, R. O., Kucharski, A. J. & Lloyd-Smith, J. O. Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19. Elite 9, e55570, https://doi.org/10.7554/eLife.55570. (2020).

3. Chow, E. J. et al. Symptom screening at illness onset of health care personnel with SARS-CoV-2 infection in King County, Washington. JAMA. https://doi.org/10.1001/jama.2020.6637 (2020).

4. Hoehl, S. et al. Evidence of SARS-CoV-2 infection in returning travelers from Wuhan, China. N. Engl. J. Med. 382, 1278–1280 (2020).

5. Kim, D., Quinn, J., Pinsky, B., Shah, N. H. & Brown, J. Rates of co-infection between SARS-CoV-2 and other respiratory pathogens. JAMA. https://doi.org/10.1001/jama.2020.6266 (2020).

6. Shah, N. Estimating the feasibility of symptom-based classification of COVID-19. https://medium.com/nimag/estimating-the-feasibility-of-symptom-based-classification-of-covid-19-cdb1a1ef1950 (2020).

7. Shah, N. Profiling presenting symptoms of patients screened for SARS-CoV-2. https://medium.com/nimag/an-ehr-derived-summary-of-the-presenting-symptoms-of-patients-screened-for-sars-cov-2-910cebbb2b9 (2020).

8. Call, S. A., Vollenweider, M. A., Hornung, C. A., Simel, D. L. & McKinney, W. Does this patient have influenza? JAMA 293, 987–997 (2005).

9. Michiels, B., Thomas, I., Van Royen, P. & Coenen, S. Clinical prediction rules combining signs, symptoms and epidemiological context to distinguish influenza from influenza-like illnesses in primary care: a cross sectional study. BMC Fam. Pract. 12, 4 (2011).

10. Conway, N. T. et al. Clinical predictors of influenza in young children: the limitations of ‘influenza-like illness’. J. Pediatr. Infect. Dis. Soc. 2, 21–29 (2013).

11. van Vugt, S. F. et al. Validity of a clinical model to predict influenza in patients presenting with symptoms of lower respiratory tract infection in primary care. Fam. Pract. 32, 408–414 (2015).

12. Dominguez, A. et al. Usefulness of clinical definitions of influenza for public health surveillance purposes. Viruses 12(1), 95. https://doi.org/10.3390/v12010095 (2020).

13. Hripcsak, G., Albers, D. J. & Perotte, A. Parameterizing time in electronic health record studies. J. Am. Med. Inform. Assoc. 22, 794–804 (2015).

14. Sethuraman, N., Jeremiah, S. S. & Ryo, A. Interpreting diagnostic tests for SARS-CoV-2. JAMA. https://doi.org/10.1001/jama.2020.2359 (2020).
ACKNOWLEDGEMENTS

We acknowledge support from the Bill & Melinda Gates Foundation, award # INV-017214.

AUTHOR CONTRIBUTIONS

Authors E.S. and A.C. conceived of the study, and were equal contributors and co-first authors. Authors E.S., J.A.F. and C.K.C. conducted data processing, model development, and analysis, supervised by authors A.C. and N.H.S. Author S.G. contributed to data acquisition efforts. Author B.P. contributed to manual review of observed symptoms in SARS-CoV-2 tested patients. Authors A.C. and N.H.S. drafted the manuscript, which was edited and approved by all authors.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at https://doi.org/10.1038/s41746-020-0300-0.

Correspondence and requests for materials should be addressed to A.C.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.