Insights from Twitter about novel COVID-19 symptoms

Chayakrit Krittananawong1*, Bharat Narasimhan2, Hafeez Ul Hassan Virk3, Harish Narasimhan4, Zhen Wang5,6, and W.H. Wilson Tang7

1The Michael E. DeBakey VA Medical Center, Section of Cardiology, Baylor College of Medicine, Houston, TX, USA; 2Department of Internal Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 3Department of Cardiovascular Diseases, Case Western Reserve University, University Hospitals Cleveland Medical Center, Cleveland, OH, USA; 4Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA; 5Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN, USA; 6Division of Health Care Policy and Research, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA; and 7Department of Cardiovascular Medicine, Heart and Vascular Institute, Cleveland Clinic, Cleveland, OH, USA

Online publish-ahead-of-print 23 November 2020

When first identified of coronavirus disease 2019 (COVID-19), fever, cough, shortness of breath, or difficulty breathing often are symptoms of COVID-19. On 27 April 2020, the Centers for Disease and Prevention (CDC) has added six more symptoms (chills, repeated shaking with chills, muscle pain, headache, sore throat, and new loss of taste or smell) to the existing list of symptoms for COVID-19.1 Some cases reported diarrhoea as COVID-19 symptoms. Twitter, an interactive social media platform, has been increasingly utilized in both academic and non-academic purposes. However, scant studies using Twitter have explored its value for the COVID-19 era. In this study, we performed Twitter text mining to analyse self-reported COVID-19 symptoms quantitatively.

Twitter data were collected via the twitter public streaming application programming interface, a more automated data retrieval approach through the application programming interface. Tweets containing the terms ‘covid’, ‘covid-19’, ‘corona’, ‘coronavirus’, COVID-19 related hashtag (e.g. #covid, #covid19, #corona), as well as filtering using ‘positive’, ‘test’, ‘tested’, ‘feel’, ‘I’, ‘we’, ‘my’, ‘us’, ‘our’ between 1 April 2020 and 6 May 2020 were obtained and included. Two investigators (C.K. and B.N.) reviewed each tweet with disagreements resolved by discussion or by involving the co-authors (H.H.V. and H.H.) to adjudicate and establish consensus. All analyses performed in this study relied on public, anonymized data, and adhere to the terms and conditions, terms of use and privacy policies of Twitter. No specific tweets with identification of personal information are included in this report. Tweets captured outside the USA were excluded. Unrelated topics, academic tweets, and retweets were also excluded.

We initially developed codes using a grounded theory approach of self-report symptoms of COVID-19. First, one investigator (C.K.) examined a 35% sample from all messages to build an initial coding framework. Next, the research team (B.N., H.H.V., and H.H.) analysed an additional 30% of all tweets and met to refine the overall coding framework, including tweet category, location, and self-report symptoms. Discrepancies were resolved via discussion among co-authors. If new symptoms emerged, we adjusted the coding framework and re-analysed all tweets following the new coding framework. Once no new categories emerged, reaching saturation, the coding framework was finalized into the final coding framework that accounted for each system (e.g. neurological, pulmonology, cardiovascular, or musculoskeletal systems). In the final coding framework, each tweet was reviewed by two investigators (C.K. and B.N.), and disagreements were resolved by discussion among the co-authors to establish consensus. For selected tweets, random selections were made by each investigator (C.K., H.J., and B.N.) and discussed among the group to determine which tweets to include in the cohort.

Of more than 1 million COVID-19 related tweets, 531 994 tweets were excluded due to retweeting (i.e. resharing of the same message), 254 683 were excluded due to irrelevant, and 374 982 were excluded due to academic tweets. Overall, we analysed 14 698 Tweets associated with COVID-19 symptoms. The most frequently reported symptoms were taste disturbance (22%), anosmia (21%), psychiatric symptoms (17%), fever (15%), cough (5%), shortness of breath (4%), chest pain (4%), muscle pain (3%), and cold (2%) (Figure 1). We also identified uncommon reports of COVID-19 symptoms including psychiatric symptoms (17%), rash (5%), toe symptoms (2%), conjunctivitis (1%), palpitation (3%), and numbness (1%) (Figure 1).

This study has two main findings. We could first identify the prevalence of common COVID-19 symptoms such as anosmia, fever, cough, difficulty breathing, cold, and chest pain. Second, we identified a large volume of US-based Tweets about novel COVID-19 symptoms such as psychiatric symptoms, rash, toe symptoms, conjunctivitis, or numbness.

Several common symptoms in COVID-19 patients in previous studies have been consistent with our study. A recent meta-analysis

* Corresponding author. Tel: +1 713 798 4951; Email: chayakrit.krittananawong@bcm.edu
© The Author(s) 2020. Published by Oxford University Press on behalf of the European Society of Cardiology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
of 43 studies showed that 16 clinical symptoms among COVID-19 patients including fever (83.3%), cough (60.3%), fatigue (38.0%), myalgia (28.5%), increased sputum production (26.9%), and shortness of breath (24.9%). The present study yielded that Twitter can be used as an alternative to clinical research using self-report symptoms to identify COVID-19 patients. Another meta-analysis of 10 studies showed that the main clinical symptoms of COVID-19 patients include fever (88.5%), cough (68.6%), myalgia or fatigue (35.8%), expectoration (28.2%), and dyspnoea (21.9%), headache, dizziness (12.1%), diarrhoea (4.8%), nausea and vomiting (3.9%).

Most importantly, the present study showed that Twitter could be clinical research using self-report symptoms to identify COVID-19 patients. Although our results may be confounded with large amounts of noise, we found several novel and unusual symptoms after manual review. Unusual symptoms such as psychiatric symptoms, rash, toe symptoms, conjunctivitis, or numbness can be used to warn the public if individuals have those symptoms that may need to get tested and self-quarantine.

This study has certain limitations. First, the present study did not include negative test individuals who reported symptoms (e.g. false-negative tests). Second, Twitter users’ clinical characteristics (i.e. age, socioeconomic, lifestyle, smoking history, comorbidities) are also a limitation of the use of Twitter for our research. Last, we cannot exclude observation bias by the human analysis. However, new artificial intelligence techniques such as semi-supervise text mining or deep reinforcement learning techniques could potentially be used for excluding observation bias by the human analysis.

Twitter may be useful for studying new pandemics, particularly at the beginning of pandemics. Novel symptoms such as psychiatric symptoms, rash, toe symptoms, conjunctivitis, or numbness may be needed close monitoring.

Conflict of interest: C.K. discloses the following relationships – the American College of Cardiology Solution Set Oversight Committee, The Lancet Digital Health (Advisory Board), European Heart Journal - Digital Health (Editorial board) and Journal of the American Heart Association (Editorial board); The Journal of Scientific Innovation in Medicine (Associate Editor). None of the other authors have any disclosures.

Data availability

Data supporting this research belongs to Twitter and cannot be shared without permission from the owner.

References

1. CDC Adds 6 Symptoms To Its COVID-19 List. https://www.npr.org/sections/coronavirus-live-updates/2020/04/27/845321155/cdc-adds-6-symptoms-to-its-covid-19-list (26 October 2020).
2. Fu L, Wang B, Yuan T, Chen X, Ao Y, Fitzpatrick T, Li P, Zhou Y, Lin Y-F, Duan Q, Luo G, Fan S, Liu Y, Feng A, Zou H. Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: a systematic review and meta-analysis. J Infect 2020;80:656–665.
3. Li L-q, Huang T, Wang Y-q, Wang Z-p, Liang Y, Huang T-b, Zhang H-y, Sun W, Wang Y. COVID-19 patients’ clinical characteristics, discharge rate, and fatality rate of meta-analysis. J Med Virol 2020;92:577–583.