



Symptoms associated with a COVID-19 infection among a non-hospitalized cohort in Vienna

Nicolas Munsch · Stefanie Gruarin · Jama Nateqi · Thomas Lutz · Michael Binder · Judith H. Aberle · Alistair Martin · Bernhard Knapp 

Received: 21 December 2021 / Accepted: 13 March 2022 / Published online: 13 April 2022
© The Author(s) 2022

Summary

Background Most clinical studies report the symptoms experienced by those infected with coronavirus disease 2019 (COVID-19) via patients already hospitalized. Here we analyzed the symptoms experienced outside of a hospital setting.

Methods The Vienna Social Fund (FSW; Vienna, Austria), the Public Health Services of the City of Vienna (MA15) and the private company Symptoma collaborated to implement Vienna's official online COVID-19 symptom checker. Users answered 12 yes/no questions about symptoms to assess their risk for COVID-19. They could also specify their age and sex, and whether they had contact with someone who tested positive for COVID-19. Depending on the assessed risk of COVID-19 positivity, a SARS-CoV-2 nucleic acid

amplification test (NAAT) was performed. In this publication, we analyzed which factors (symptoms, sex or age) are associated with COVID-19 positivity. We also trained a classifier to correctly predict COVID-19 positivity from the collected data.

Results Between 2 November 2020 and 18 November 2021, 9133 people experiencing COVID-19-like symptoms were assessed as high risk by the chatbot and were subsequently tested by a NAAT. Symptoms significantly associated with a positive COVID-19 test were malaise, fatigue, headache, cough, fever, dysgeusia and hyposmia. Our classifier could successfully predict COVID-19 positivity with an area under the curve (AUC) of 0.74.

Conclusion This study provides reliable COVID-19 symptom statistics based on the general population verified by NAATs.

Supplementary Information The online version of this article (<https://doi.org/10.1007/s00508-022-02028-9>) contains supplementary material, which is available to authorized users.

N. Munsch · A. Martin · B. Knapp (✉)
Science Department, Symptoma GmbH, Vienna, Austria
science@symptoma.com

S. Gruarin · J. Nateqi · T. Lutz
Science Department, Symptoma GmbH, Salzburg, Austria

J. Nateqi
Department of Internal Medicine, Paracelsus Medical University, Salzburg, Austria

M. Binder
Vienna Health Care Company, Vienna, Austria

J. H. Aberle
Center for Virology, Medical University of Vienna, Vienna, Austria

B. Knapp
Faculty Computer Science, University of Applied Sciences Technikum, Vienna, Austria

Keywords Symptom assessment · Self-reported · Symptom checker · Chatbot · Machine learning

Introduction

The frequency of the symptoms associated with coronavirus disease 2019 (COVID-19) is valuable information for health authorities during the pandemic. Such knowledge has been used in a variety of applications including triage recommendations and diagnostics [1–3]; however, most studies reporting on COVID-19 symptom frequencies concern patients in hospital settings [1]. Thus, published symptom frequencies may not reflect those found in the general population. For example, non-hospitalized people experience less severe forms of the disease [1].

To reduce the above sampling bias, some approaches have been implemented to collect COVID-19 self-reported symptoms from the general population. Self-reported symptoms have been collected

via a variety of means, including an official test prioritization questionnaire [4], symptom trackers [5–7], a symptom checker [8] and even Twitter [9]; however, only the symptoms collected via an official test prioritization questionnaire are associated with verified tests results. With other methods, test results are only self-reported. Of the data associated with a verified result, while valuable, most have shortcomings. For example, the data released by the Israeli Ministry of Health does not contain dysgeusia or anosmia as it was not known as a relevant symptom at the early stages of the pandemic [4].

In this work, we describe the COVID-19 symptoms reported by a non-hospitalized cohort in Vienna from 2 November 2020 to 18 November 2021. With a dataset spanning over 9000 users, we analyze the association between these symptoms and the COVID-19 NAAT status. Lastly, we build a classifier to predict COVID-19 positivity from those experiencing flu-like symptoms.

Methods

Data collection

From November 2020, Vienna's online COVID-19 symptom checker provided inhabitants with an initial COVID-19 risk assessment. Depending on the outcome, possible options for further action included a nucleic acid amplification test (NAAT) using the reverse transcription-polymerase chain reaction (RT-PCR) method [10, 11]. The aim was to offer an additional scalable service, complementing the medical telephone health service "1450". The symptom checker is currently available at <https://symptomchecker.fsw.at/>.

The Vienna Social Fund (FSW), the Public Health Services of the City of Vienna (MA15), and the private company Symptoma mutually developed the chatbot based on previous results detailing the accuracy of Symptoma's symptom checker with respect to COVID-19 [3, 12, 13]. During the chatbot conversation, each user was asked the same set of questions and responses were recorded accordingly. A user had to answer a series of 12 yes/no questions about symptoms. These are fever ($>38^{\circ}\text{C}$), cough, dyspnea, sneezing, rhinorrhea, sore throat, malaise, fatigue, diarrhea, headache, hyposmia and dysgeusia. In addition, the user could indicate if, in the last 10 days, there was close contact with a duration longer than 15 min with a person who tested positive for COVID-19. Finally, each user was invited to specify their age and sex. We did not record the exact age of users for data protection issues, but only the age group (see Supplementary Table 1). Symptoms, age group and sex information are used by Symptoma's algorithm to rank COVID-19 against over 20,000 other potential causes [3]. If COVID-19 appeared in the 30 top causes, the user was offered a NAAT [10, 11]. A NAAT was also

offered if the user reported a positive self-test, had returned from abroad, or has a severe medical pre-condition and reported any symptom.

The statistics reported in this paper are based on the combined information of the chatbot conversations and the results of the NAATs. A total of 120,768 users were screened this way between the 2nd of November 2020 and the 18th of November 2021. A total of 88,861 users (73.6%) were eligible for the NAAT, of which 10,089 (11.4%) were tested. Among users who did a NAAT, 956 (9.5%) did not report any symptoms but only close contact with a person who tested positive for COVID-19. These were excluded from our further analyses. The remaining 9133 (90.5%) users, who were both symptomatic and had performed a NAAT, were used in the further analyses.

Data analysis

All data were anonymized prior to analysis. Only sex, age group, the answers to the questions, and the result of the NAAT were collated. We analyzed for each symptom if there was a significant difference between users who tested positive for COVID-19 (C19+) and users who tested negative for COVID-19 (C19-). The *P*-values were calculated by a two-tailed Fisher's exact test and corrected for multiple testing by the Benjamini-Hochberg [14] method.

We analyzed for each symptom if there was a significant difference between male and female among those who tested positive, and if there was a significant difference between male and female among those who tested negative. The *P*-values were calculated by a two-tailed Fisher's exact test and corrected for multiple testing by the Benjamini-Hochberg [14] method for each of the two analyses independently.

In addition, we quantified the association of each symptom pair via a two-tailed Fisher's exact test. Odds ratios (OR) were calculated and *P*-values were corrected for multiple testing by the Benjamini-Hochberg [14] method.

We also quantified the association of the three-wise combinations of symptoms via a Breslow-Day test. The *P*-values were corrected for multiple testing by the Benjamini-Hochberg [14] method.

Lastly, we built a logistic regression model to predict C19+ based on the collected data. A total of 8966 users, who provided an age group and specified their sex as either male or female, were included in this analysis. Sex information is encoded to be 1 for the female and 0 for the male. Age categories were encoded as integers and treated as continuous. Performance was assessed based on the concatenation of the 10 test sets obtained from the cross-validation. We analyzed the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC). Confidence intervals (CI) were calculated by bootstrapping with 3000 repetitions. We then repeated this analysis when including interaction terms in the model.

Models with and without interaction terms were compared via a one-way ANOVA. All analyses were done in Python 3.8 using the libraries Numpy (1.21.0) [15], Pandas (1.3.4) [16], Scikit-learn (1.0.1), and Statsmodels (0.13.1) [17]. Visualizations were produced using Matplotlib (3.5.1) [18] and Seaborn (0.11.2) [19]. Benjamini-Hochberg-corrected P -values are reported when multiple testing correction was performed.

Results

Symptom frequencies among COVID-19 positive and COVID-19 negative users

Summary statistics of participants and numerical details are given in Supplementary Table 1. Our study cohort consisted of 9133 non-hospitalized persons experiencing flu-like symptoms of whom 2692 (29.5%) tested positive for COVID-19 (C19+) and 6341 (69.4%) tested negative for COVID-19 (C19-). The test was unclear for 100 persons (1.1%). The median group age was 31–40 years for C19+ and 21–30 years for the C19- groups. In Fig. 1 we compared the symptom frequencies between C19+ and C19-. The symptoms most frequently reported by C19+ users were malaise (78.6%), fatigue (73.8%), headache (63.7%), cough (59.8%), and fever (49.7%). Users less frequently reported sore throat (47.8%), close contact with a person who tested positive for COVID-19 (40.7%), rhinorrhea (38.1%), sneezing (33.9%), dysgeusia (28.9%), and hyposmia (26.0%). Dyspnea (15.0%) and diarrhea (11.8%) were rarely reported.

The C19+ users significantly more frequently reported cough ($P < 0.001$), hyposmia ($P < 0.001$), fever ($P < 0.001$), dysgeusia ($P < 0.001$), headache ($P < 0.001$), malaise ($P < 0.001$), fatigue ($P < 0.001$), and close contact with a person who tested positive for COVID-

19 ($P < 0.001$). On the contrary, C19+ users significantly less frequently reported diarrhea ($P < 0.001$), sore throat ($P < 0.001$) and sneezing ($P = 0.01$); however, no significant difference between the C19+ and C19- groups was found for rhinorrhea ($P = 0.12$) and dyspnea ($P = 0.17$).

The largest increase of symptom frequency in C19+ persons was found for close contact with a person who tested positive for COVID-19 (+19.8%), hyposmia (+15.8%) and dysgeusia (+14.1%). The largest decrease in C19+ persons was found for sore throat (-5.4%) and diarrhea (-5.3%).

In both C19+ and C19- groups women reported sore throat ($P < 0.001$), sneezing ($P < 0.001$) and headache ($P < 0.01$) more frequently than men. Men reported fever more frequently ($P < 0.001$). In the C19+ group only, rhinorrhea and dyspnea ($P < 0.01$) were more frequently present for women than for men. In the C19- group only, men reported diarrhea more frequently ($P = 0.02$), while women more frequently reported fatigue ($P < 0.001$) (Supplementary Fig. 1).

Co-occurrence and association of symptoms

The frequency and the association of all pairs of symptoms within the C19+ are indicated in Fig. 2. For 66.7% of the pairs of symptoms, the association was significantly positive. The three highest associations of symptoms within the C19+ group were between dysgeusia and hyposmia (odds ratio, OR=32.05, $P < 0.001$), fatigue and malaise (OR=8.63, $P < 0.001$), and sneezing and rhinorrhea (OR=7.02, $P < 0.001$). Within the C19+ group, 21% reported both dysgeusia and hyposmia, 23% reported both rhinorrhea and sneezing and 66% reported both fatigue and malaise. These associations were also observed within the C19- group (see Supplementary Fig. 2).

Among associations with a OR higher than 1.18, only those between fever and diarrhea (OR=1.28, $P = 0.052$), malaise and dysgeusia (OR=1.22, $P = 0.083$), and dysgeusia and headache (OR=1.19, $P = 0.063$) were not significant. Associations with an OR between 0.86 and 1.16 did not have a significant positive or negative association.

The three strongest significantly negative associations in the C19+ group were between fever and hyposmia (OR=0.56, $P < 0.001$), fever and dysgeusia (OR=0.69, $P < 0.001$), and fever and sneezing (OR=0.71, $P < 0.001$).

The frequency and the association of all three-wise combinations of symptoms within C19+ are indicated in Supplementary Table 2. For 11.9% of the combinations, the OR are significantly different between the strata. Among these combinations, the ones with the highest co-occurrence frequency are the triplets fatigue, malaise and dysgeusia with 21.3%, malaise, headache and dysgeusia with 17.7% and fatigue, headache and dysgeusia with 17.6%. As a reference for these triplets, the expected symptom co-

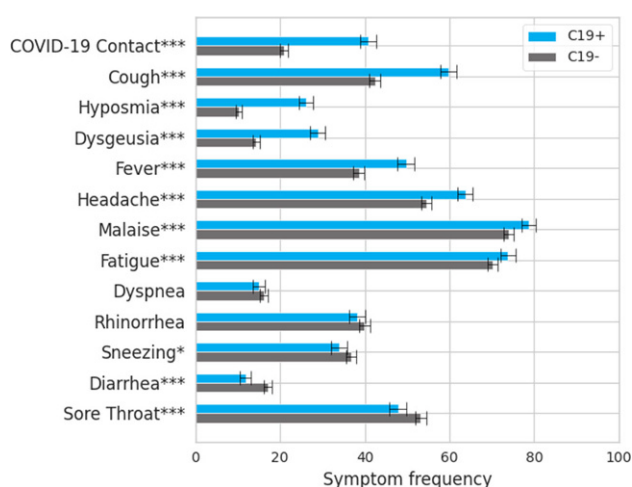
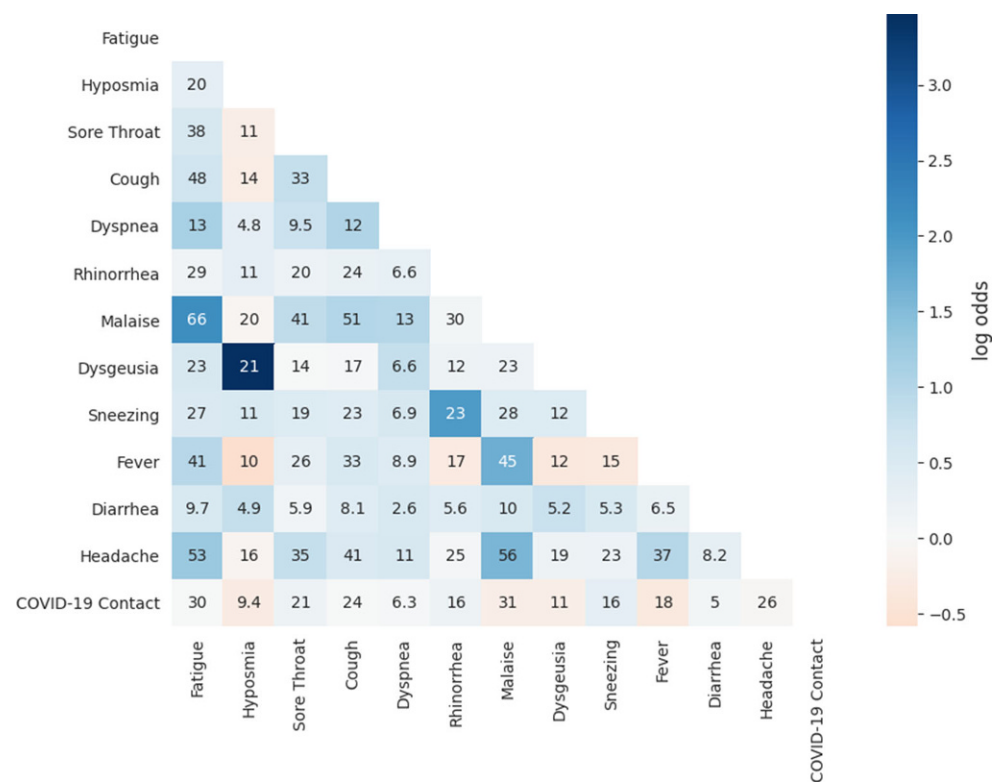


Fig. 1 Symptom frequencies in percentage for the C19+ and C19- groups. Error bars indicate the 95% confidence intervals. Significance of the difference between these groups are indicated with one, two, and three asterisks which correspond to a p -value less than 0.05, 0.01, and 0.001 respectively

Fig. 2 Symptom co-occurrence frequencies for the C19+ group. Frequencies are reported in percentage of the C19+ group that report both symptoms. Log odds ratios (LOR) are represented by the color scale. They show the strength of the association. LOR indicates an association when its value is more than 0, a dissociation if lower than 0. The equivalent results for the C19–group are included as Supplementary Fig. 2



frequencies calculated as the product of the individual symptom frequencies are 16.8%, 14.4% and 13.6%, respectively.

Evaluation of a classifier based on symptoms

A classifier was built using the 8966 users who reported symptoms, age, and sex. The ROCs for each test set are shown in Fig. 3. Across the 10 test sets, the logistic regression model predicts with an AUC of 0.74 (95% CI 0.72 to 0.75). The coefficients and the P -values obtained for the logistic regression after training on all the available data are reported in Supplementary Table 3. Possible working points, that being a threshold to which we predict COVID-19 positivity, include a sensitivity of 0.70 and a specificity of 0.65, sensitivity of 0.80 and a specificity of 0.51, or a sensitivity of 0.90 and a specificity of 0.32. We also evaluated the performance when excluding the answer about a contact with a COVID-19 positively tested person. In this setting, the AUC is 0.69 (95% CI 0.68 to 0.70). The ROC curve is included in Supplementary figure 3. Next, we expanded the first model to include all pairwise-interaction terms. It performs significantly better than the model without interaction terms ($P < 0.001$), with an AUC of 0.76 (95% CI 0.74 to 0.77). The ROC curve for the model with interaction terms is shown in Supplementary Fig. 4. All coefficients for this model with interaction terms are reported in the Supplementary Table 4.

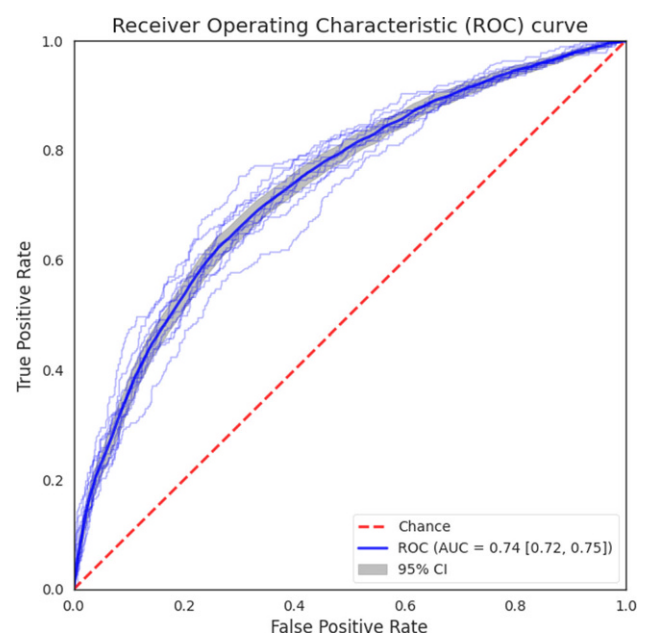


Fig. 3 Receiver operating characteristic (ROC) curve of the logistic regression model when accounting for the contact with COVID-19 case information. The gray band shows the 95% confidence Intervals (CI). The area under the curve (AUC) is provided to summarize the curve. An alternative version of the ROC curve for the logistic regression model without using the contact with COVID-19 case information is included in Supplementary Fig. 3

Discussion

To the best of our knowledge, this is the first study reporting symptoms associated with COVID-19 of the general population presenting symptoms, tested by NAATs. The drop-out rate of users who reported symptoms without doing a test (91.6%) is slightly lower than in similar studies (95.9% [8], 97.5% [7] and 99.3% [2]). Our results showed a cough frequency of 59.8% which falls within the 95% CI of the symptom frequencies reported in literature (range 59.8–74.1%) [1]. Similar agreements were found for hyposmia (26.0% vs. 17.7%–41.3%), dysgeusia (28.9% vs. 12.4–43.5%), diarrhea (11.8% vs. 7.6–17.4%) and fever (49.7% vs. 35.0–71.7%). The frequency of dyspnea (15.0%) was slightly below the 95% CI reported (16.6–35.5%), while the frequencies of headache (63.7% vs. 9.2–43.5%), fatigue (73.8% vs. 22.1–53.6%) and sore throat (47.8% vs. 13.5–31.6%) were higher [1]. This discrepancy might be explained by the difference in interpreting the symptoms between patient and physician [20]. For example, fatigue was questioned by the associated chatbot, as “Have you been feeling particularly fatigued or dull lately?” (in German: “Fühlen Sie sich neuerdings besonders ermüdet oder matt?”).

The results also show that the symptoms experienced by C19+ significantly differ, except for dyspnea, to those experienced by C19-. This agrees with other studies that also reported hyposmia, dysgeusia, and fever as significantly increased in C19+ persons [6, 21, 22]. Further, the large relative difference of hyposmia (+155%) and dysgeusia (+101%) frequencies for C19+ in comparison to C19- users suggests that hyposmia and dysgeusia are specific but not sensitive, i.e. their presence strongly suggests the user is C19+ but no conclusion can be drawn from their absence.

The association of dyspnea with a COVID-19 positivity, not observed in the present analysis, was also not found by Menni et al. [2]. This might be due to a sample bias as dyspnea is often a late symptom of an infection while chatbot users might rather be at an earlier stage of infection [23]. Alternatively, dyspnea can be a distressing symptom and affected individuals might rather call an emergency hotline instead of using a chatbot [24]. Additionally, results show symptoms were reported in different frequencies by men and women, which could be caused by sex-specific differences in the clinical course [25, 26].

As shown in previous studies, the high correlation found between dysgeusia and hyposmia indicates that these pairs of symptoms frequently occur together [27]. The same holds true for rhinorrhea and sneezing [28]. The high correlation between fatigue and malaise might be explained by the fact that fatigue is a subjective symptom of malaise [29]. The latter pair of symptoms also has a high co-occurrence frequency, which might be explained by the correlation, and the non-specific nature of these symptoms [29].

The AUC of our predictor (0.74) is in the range of the performance of the symptom-based COVID-19 predictor described in the literature. Other reported AUCs were as 0.68 [21], 0.74 [2] and 0.90 [4]. The considerably higher AUC of the latter predictor is explained by the inclusion of many asymptomatic patients who did not report any contact with a COVID-19 infected person. These patients, as expected, are mostly C19-, thereby inflating performance. Predicting COVID-19 positivity from patients who do not report any symptoms or contact is not considered within our study, which only deals with symptomatic people.

Our study has limitations. First, self-reported symptoms are, by definition, not assessed by a medical professional which leads to inconsistencies. Second, there is selection bias because people with a low risk of being C19+ were not offered a test (see methods). Another selection bias is the potential underrepresentation of subgroups with reduced access to the technology. For example, Nguyen et al. show that the usage of chatbots significantly decreases with age and low education level [30]. Third, users can experience additional symptoms after completing the session. These symptoms were not recorded and included in the present study. This leads us to believe, as discussed previously, that our cohort is focused on the early onset of COVID-19. Lastly, the consideration of the NAAT as a ground truth has been criticized due to its low sensitivity [31].

In addition to the above limitations, our sampling period aligns with Austria's COVID-19 vaccination campaign as well as the emergence of new variants. Both of these factors, namely being vaccinated or being infected with a variant, have the potential to alter the symptoms experienced. For example, it has been reported that vaccinations reduce the number of symptoms experienced [32]. In contrast, the emergence of variant B.1.1.7 (Alpha), was reported to not affect the symptoms experienced [33]. In our study, we did not find any significant changes in the symptom frequencies over time (see Supplementary Fig. 5).

In conclusion, we have analyzed in depth the COVID-19 symptoms reported by a non-hospitalized cohort in Vienna over the past year. Data were systematically collected and results were automatically associated with a NAAT. To date, no other work features a general European population in combination with systematic data collection. For this reason, we believe that this work provides excellent new insights into the characteristics of COVID-19.

Acknowledgements We would like to thank the Vienna Social Fund (FSW), Public Health Services of the City of Vienna (MA15) and the Vienna Health Authority for providing data and making this publication possible.

Funding This study has received funding from the European Union's Horizon 2020 research and innovation programme

under grant agreement No 830017 and by the Austrian Research Promotion Agency under grant agreement No 880939 (supported by the Federal Ministries Republic of Austria for Digital and Economic Affairs and Climate Action, Environment, Energy, Mobility, Innovation and Technology).

Funding Open access funding provided by Fachhochschule Technikum Wien.

Declarations

Conflict of interest N. Munsch, S. Gruarin, J. Nateqi, T. Lutz, A. Martin and B. Knapp are current or past employees of Symptoma GmbH. J. Nateqi and T. Lutz hold shares of Symptoma. M. Binder and J. H. Aberle declare that they have no competing interests.

Ethical standards This study was exempted from ethics review by the ethics commission of the city of Vienna (MA15-EK/21-037-VK_NZ). All individuals using the chatbot agreed that their data will be used in an anonymized way.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Struyf T, Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeflang MM, 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*. 2021;2:CD13665.
- Menni C. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med*. 2020;26:8.
- Martin A, Nateqi J, Gruarin S, Munsch N, Abdurahmane I, Zobel M, et al. An artificial intelligence-based first-line defence against COVID-19: digitally screening citizens for risks via a chatbot. *Sci Rep*. 2020;10:19012.
- Zoabi Y, Deri-Rozov S, Shomron N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *npj Digit Med*. 2021;4:1–5.
- Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, et al. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science*. 2020;368:1362–7.
- Dreyer NA, Reynolds M, DeFilippo Mack C, Brinkley E, Petruski-Ivleva N, Hawaldar K, et al. Self-reported symptoms from exposure to Covid-19 provide support to clinical diagnosis, triage and prognosis: An exploratory analysis. *Travel Med Infect Dis*. 2020;38:101909.
- Zens M, Brammertz A, Herpich J, Südkamp N, Hinterseer M. App-based tracking of self-reported COVID-19 symptoms: analysis of questionnaire data. *J Med Internet Res*. 2020;22:e21956.
- Perlman A, Vodonos Zilberg A, Bak P, Dreyfuss M, Leventer-Roberts M, Vurembrand Y, et al. Characteristics and symptoms of app users seeking COVID-19—Related digital health information and remote services: retrospective cohort study. *J Med Internet Res*. 2020;22:e23197.
- Sarker A, Lakamana S, Hogg-Bremer W, Xie A, Al-Garadi MA, Yang Y-C. Self-reported COVID-19 symptoms on Twitter: An analysis and a research resource. *J Am Med Inform Assoc*. 2020;27:1310–5.
- Corman V, Bleicker T, Brünink S, Drosten C, Landt O, Koopmans M, et al. Diagnostic detection of 2019-nCoV by real-time RT-PCR. 2020. <https://www.who.int/docs/default-source/coronaviruse/protocol-v2-1.pdf>. Accessed 13 Dec 2021.
- cobas® SARS-CoV-2 Test. <https://diagnostics.roche.com/global/en/products/params/cobas-sars-cov-2-test.html>. Accessed 13 Dec 2021.
- Munsch N, Martin A, Gruarin S, Nateqi J, Abdurahmane I, Weingartner-Ortner R, et al. Diagnostic accuracy of web-based COVID-19 symptom checkers: Comparison study. *J Med Internet Res*. 2020;22:e21299.
- Nateqi J, Lin S, Krobath H, Gruarin S, Lutz T, Dvorak T, et al. From symptom to diagnosis-symptom checkers re-evaluated : Are symptom checkers finally sufficient and accurate to use? An update from the ENT perspective. *HNO*. 2019;67:334–42.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;57:289–300.
- Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature*. 2020;585:357–62.
- Reback J, jbrockmendel, McKinney W, den Bossche JV, Augspurger T, Cloud P, et al. pandas-dev/pandas: Pandas 1.3.4. Zenodo. 2021; <https://doi.org/10.5281/zenodo.5574486>.
- Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with Python. Proceedings of the 9th Python in Science Conference. 2010. <https://conference.scipy.org/proceedings/scipy2010/seabold.html>. Accessed 13 Dec 2021.
- Caswell TA, Droettboom M, Lee A, de Andrade ES, Hoffmann T, Hunter J, et al. matplotlib/matplotlib: REL: v3.5.1. Zenodo. 2021; <https://doi.org/10.1109/MCSE.2007.55>.
- Waskom ML. seaborn: Statistical data visualization. *J Open Source Softw*. 2021;6:3021.
- Favrat B, Cornuz J. Evaluation of fatigue. <https://bestpractice.bmj.com/topics/en-us/571>. Accessed 13 Dec 2021.
- Dantas LF, Peres IT, Bastos LSL, Marchesi JE, de Souza GFG, Gelli JGM, et al. App-based symptom tracking to optimize SARS-CoV-2 testing strategy using machine learning. *PLoS ONE*. 2021;16:e248920.
- Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. *Nat Med*. 2021;27:73–7.
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395:497–506.
- Parshall MB, Schwartzstein RM, Adams L, Banzett RB, Manning HL, Bourbeau J, et al. An official American Thoracic Society statement: update on the mechanisms, assessment, and management of dyspnea. *Am J Respir Crit Care Med*. 2012;185:435–52.
- Peckham H, de Gruijter NM, Raine C, Radziszewska A, Ciurtin C, Wedderburn LR, et al. Male sex identified by

- global COVID-19 meta-analysis as a risk factor for death and ICU admission. *Nat Commun.* 2020;11:6317.
26. Pradhan A, Olsson P-E. Sex differences in severity and mortality from COVID-19: are males more vulnerable? *Biol Sex Differ.* 2020;11:53.
 27. Carignan A, Valiquette L, Grenier C, Musonera JB, Nken-gurutse D, Marcil-Héguy A, et al. Anosmia and dysgeusia associated with SARS-CoV-2 infection: an age-matched case—Control study. *CMAJ.* 2020;192:E702–7.
 28. Eccles R. Understanding the symptoms of the common cold and influenza. *Lancet Infect Dis.* 2005;5:718–25.
 29. Sharpe M, Wilks D. Fatigue. *BMJ.* 2002;325:480–3.
 30. Nguyen A, Mosadeghi S, Almario CV. Persistent digital divide in access to and use of the Internet as a resource for health information: Results from a California population-based study. *Int J Med Inf.* 2017;103:49–54.
 31. Kanji JN, Zelyas N, MacDonald C, Pabbaraju K, Khan MN, Prasad A, et al. False negative rate of COVID-19 PCR testing: A discordant testing analysis. *Virology.* 2021;18:13.
 32. Antonelli M, Penfold RS, Merino J, Sudre CH, Molteni E, Berry S, et al. Risk factors and disease profile of post-vaccination SARS-CoV-2 infection in UK users of the COVID Symptom Study app: a prospective, community-based, nested, case-control study. *Lancet Infect Dis.* 2021; [https://doi.org/10.1016/S1473-3099\(21\)00460-6](https://doi.org/10.1016/S1473-3099(21)00460-6).
 33. Graham MS, Sudre CH, May A, Antonelli M, Murray B, Varsavsky T, et al. Changes in symptomatology, reinfection, and transmissibility associated with the SARS-CoV-2 variant B.1.1.7: an ecological study. *Lancet Public Health.* 2021;6:e335–45.

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