Real time tracking of self-reported symptoms to predict potential COVID-19 infection

Cristina Menni, PhD[1,6*], Ana M Valdes , PhD [1,2,6], Maxim B Freidin, PhD [1], Carole H. Sudre, PhD[3], Long H. Nguyen, MD [5]; David A. Drew, PhD[5],Sajaysurya Ganesh, PhD[4], Thomas Varsavsky, Phd[3], M. Jorge Cardoso, PhD[3], Julia S El-Sayed Moustafa, PhD[1], Alessia Visconti, PhD[1], Pirro Hysi , PhD[1], Ruth C E Bowyer, PhD[1], Massimo Mangino, PhD[1], Mario Falchi, PhD [1], Jonathan Wolf, BSc[4], Sebastien Ourselin, PhD[3], Andrew T. Chan, MD, MPH[5], Claire J Steves , PhD [1,7], Tim D Spector , MD[1,7*]

Affiliations:

[1] Department of Twin Research and Genetic Epidemiology, King's College London, Westminster Bridge Road, SE17EH London, UK

[2] Academic Rheumatology Clinical Sciences Building, Nottingham City Hospital, Hucknall Road, Nottingham, NG5 1PB, UK

[3] School of Biomedical Engineering & Imaging Sciences, King's College London, SE1 7EU London UK

[4] Zoe Global Limited, 164 Westminster Bridge Road, London SE1 7RW, UK

[5]Clinical & Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, 100 Cambridge Street, Boston, MA, USA

[6] equal contribution

[7] equal contribution

Corresponding authors:

tim.spector@kcl.ac.uk; cristina.menni@kcl.ac.uk

2,618,862 participants reported their potential symptoms of COVID-19 on a smartphone-based app. Among the 18,401 who had undergone a SARS-CoV-2 test, the proportion of participants who reported loss of smell and taste was higher in those with a positive test result (4,668 of 7,178 individuals; 65.03%) than in those with a negative test result (2,436 of 11,223 participants; 21.71%); odds ratio= 6.74[95%CI 6.31; 7.21]. A model combining symptoms to predict likely infection was applied to data from all app users who report symptoms (805,753) and predicted that 140,312 (17.42%) participants are likely to have COVID-19.

Since the outbreak of COVID-19, an acute respiratory illness caused by the novel coronavirus SARS-CoV-2, in China in December 2019 until 21st April 2020, over 2,573,143 cases have been confirmed worldwide (https://www.worldometers.info/coronavirus/). Although many people have presented with flu-like symptoms, widespread population testing is not yet available in most countries, including the US (https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/testing-in-us.html) and the UK[1]. Thus, it is important to identify the combination of symptoms most predictive of COVID-19 infection, to help to guide recommendations for self-isolation and prevent further spread of the disease[2].

Case reports and mainstream media articles from various countries indicate that a number of patients with diagnosed COVID-19 developed anosmia (loss of smell) [3,4]. Mechanisms of action for the SARS-CoV-2 viral infection causing anosmia has been postulated [5,6]. Other studies indicate that a number of infected individuals present anosmia in the absence of other symptoms[7,8] suggesting that this symptom could be used as screening tool to help identify potential mild cases who could be recommended to self-isolate[9].

We investigated whether loss of smell and taste is specific to COVID-19 in 2,618,862 individuals who used an app-based symptom tracker[10] (Methods). The symptom tracker is a free smartphone application that was launched in the UK on Tuesday the 24th March 2020, and in the US on 29th March 2020. It collects data from both asymptomatic and symptomatic individuals and tracks in real

time how the disease progresses by recording self-reported health information on a daily basis, including symptoms, hospitalisation, RT-PCR test outcomes, demographic information and pre-existing medical conditions

Between March 24, 2020 and April 21, 2020, 2,450,569 UK and 168,293 US individuals reported symptoms through the smartphone app. Of the 2,450,569 participants in the UK, 789,083 (32.2%) indicated having one or more potential symptoms of COVID-19 (**Table 1**). 15,638 UK and 2,763 US app users reported having had an RT-PCR SARS-CoV-2 test, and having received the outcome of the test. In the UK cohort, 6,452 participants reported a positive test and 9,186 participants had a negative test. In the cohort from the UK, of the 6,452 participants who tested positive for SARS-CoV-2, 4,178 participants (64.76%) reported loss of smell and taste compared with 2,083 out of 9,186 participants (22.68%) who tested negative (OR[95%CI]= 6.40[5.96; 6.87], P<0.0001, after adjusting for age, sex and BMI). We replicated this result in the US subset who had been tested for SARS-CoV-2 (adjusted OR[95%CI] = 10.01[8.23;12.16], P <0.0001) and combined the adjusted results using inverse variance fixed effect meta-analysis (OR(95%CI] = 6.74[6.31; 7.21], P<0.0001).

We reran logistic regressions adjusting for age, sex and BMI to identify other symptoms besides anosmia, which might be associated with being infected by SARS-CoV-2. All ten symptoms queried (fever, persistent cough, fatigue, shortness of breath, diarrhoea, delirium, skipped meals, abdominal pain, chest pain and hoarse voice) were associated with testing positive for COVID-19 in the UK cohort, after adjusting for multiple testing (**Figure 1a**). In the US cohort, only loss of smell and taste, fatigue and skipped meals were associated with a positive test result.

We performed stepwise logistic regression in the UK cohort, by randomly dividing it into a training and test sets (ratio 80/20) to identify independent symptoms most strongly correlated to COVID-19 adjusting for age, sex and BMI. A combination of loss of smell and taste, fatigue, persistent cough and loss of appetite resulted in the best model (with the lowest Akaike information criterion [AIC]).

We therefore generated a linear model for symptoms that included loss of smell and taste, fatigue, persistent cough and loss of appetite to get a symptoms prediction model for COVID-19:

Prediction model** = -**1.32 - (0.01 x age) + (0.44 x sex) + (1.75 x loss of smell and taste) + (0.31 x severe or significant persistent cough) + (0.49 x severe fatigue) + (0.39 x skipped meals)

where all symptoms are coded as 1 if the person self-reports the symptom and 0 if not. The sex feature is also binary, with 1 indicative of male participants and 0 representing females. The obtained value is then transformed into predicted probability using exp(X)/(1+exp(X)) transformation followed by assigning cases of "predicted COVID-19" for probabilities > 0.5 and controls for probabilities <0.5.

The prediction model had a sensitivity of 0.65[0.62; 0.67], a specificity of 0.78[0.76; 0.80], and a ROC-AUC 0.78[0.76; 0.82]. Cross-validation ROC-AUC was 0.75[0.74; 0.76] and positive predictive value was 0.69[0.66; 0.71] and negative predictive value was 0.75[0.73; 0.77]. In this model, the strongest predictor was loss of smell and taste, **Figure 1b.** Excluding loss of smell and taste from the model resulted in reduced sensitivity (0.33[0.30,0.35]) but increased specificity (0.84[0.83,0.86]). We also computed the ROC-AUC stratifying for sex and age-groups and found that results were similar in all groups with no significant differences between strata suggesting that our model works similarly within different sex and age groups. We validated the model in the US cohort and found a sensitivity of 0.66 [0.62, 0.69], a specificity of 0.83 [0.82, 0.85], a positive predictive value of 0.58 [0.55, 0.62] and a negative predictive value 0.87 [0.86, 0.89] (**Figure 1c**).

We also queried if the association between loss of smell and taste and COVID-19 was influenced by mainstream media reports. We assessed the correlation between loss of smell and taste and being COVID-19 positive in different date ranges: (i) March 24 to April 3, 2020 following a number of reports in the UK mainstream media (eg: https://www.bbc.co.uk/news/health-52111606) reporting anosmia as a symptom of COVID-19, (ii) the week of April 4-10, 2020, and (iii) from April 11, 2020 to

April 21. In the UK the OR[95%CI] for the associations of self-reported loss of smell and taste and a positive test for COVID-19 across these periods were 4.98[4.47;5.56], 6.64[5.75;7.68] and 10.40[9.08 ;11.91] respectively. However, this association was not found in the US cohorts: Mar24-Apr3 OR [95%CI] =8.13[5.18;12.78], Apr 4-10: 12.30[8.96;16.90] and Apr11-21: 9.13[6.73;12.38].

Finally, we applied the predictive model to the 608,385 UK and US individuals reporting symptoms who had not had a COVID-19 test and we find that according to our model 140,312 of 805,753 participants, 17.42%[14.45%;20.39%], reporting some symptoms are likely to be infected by the virus, representing 5.36% as a proportion of the overall responders to the app.

We report that loss of smell and taste is a potential predictor of COVID-19 in addition to other, more established, symptoms of high temperature and a new, persistent cough. COVID-19 appears to cause problems of smell receptors in line with many other respiratory viruses, including previous coronaviruses thought to account for 10-15% of cases of anosmia[7,9].
We also identify a combination of symptoms including anosmia, fatigue, persistent cough, and loss of appetite that together might identify individuals with COVID-19.

A major limitation of the current study is the self-report nature of the data included, which cannot replace physiological assessments of olfactory and gustatory function, or nucleotide-based testing for SARS-CoV-2. Both false negative and false negative reports could be included in the data set[11], and in addition, because of the way the questions are asked, gustatory and olfactory losses are conflated. Second, at present, we do not know whether anosmia was acquired prior to or following other COVID-19 symptoms, or during the illness or afterwards. This information could become available as currently healthy users track symptom development over time. As more accurate tests become available, we have the ability to optimise our model. One caveat of our study is that the

individuals on which the model was trained are not representative of the general population because performing tests for SARS-CoV-2 is not random. Testing is more likely to be done if an individual develops severe symptoms requiring hospitalisation, if an individual has been known to have had contact with people who have tested positive for SARS-CoV-2 infectionCOVID-19, in health workers, and if an individual has travelled in an area of high risk of exposure. Therefore, our results may overestimate the number of expected positive cases of SARS-Cov2 infection**.** Additionally, volunteers using the app are a self-selected group who might not be fully representative of the general population. Another limitation is the potential effect that mainstream media coverage of loss of smell and taste and COVID-19 might have had on app responses. We find that these reports might have influenced UK responders where there is a temporal trend in the strength of the association, but not in the US cohort, therefore we conclude that regardless of any bias introduced by mainstream media reports, the association between COVID-19 and loss of smell and taste remains strong.

Our work suggests that loss of sense of smell and taste could be included as part of routine screening for COVID-19 and should be added to the symptom list currently developed by the World Health Organization (www.who.int/health-topics/coronavirus). A detailed study on the natural history of broader COVID-19 symptoms, especially according to timing and frequency, will help to understand the usefulness of symptom tracking and modelling and to identify likely clusters of infection.

References

1 Whittington, A. M. *et al.* Coronavirus: rolling out community testig for covid-19 in the NHS. *BMJ opinion* (2020).

2 Rossman, H. *et al.* A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. *Nat Med*, doi:10.1038/s41591-020-0857-9 (2020).

3 Gane, S. B., Kelly, C. & Hopkins, C. Isolated sudden onset anosmia in COVID-19 infection. A novel syndrome? *Rhinology*, doi:10.4193/Rhin20.114 (2020).

4 Iacobucci, G. Sixty seconds on . . . anosmia. *BMJ* **368**, m1202, doi:10.1136/bmj.m1202 (2020).

5 Brann, D., Tsukahara, T., Weinreb, C., Logan, D. W. & Datta, S. R. On-neural expression of SARS-CoV-2 entry genes in the olfactory epithelium suggests mechanisms underlying anosmia in COVID-19 patients. *Preprint at h*[*ttps://www.biorxiv.org/content/10.1101/2020.03.25.009084v1.full.pdf*](https://www.biorxiv.org/content/10.1101/2020.03.25.009084v1.full.pdf) (2020).

6 Sungnak, W. *et al.* SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat Med*, doi:10.1038/s41591-020-0868-6 (2020).

7 Hopkiins, C. & Kumar, N. Loss of sense of smell as marker of COVID-19 infection. *ENTUK* (2020).

8 Eliezer, M. *et al.* Sudden and Complete Olfactory Loss Function as a Possible Symptom of COVID-19. *JAMA Otolaryngology-HEad &Neck Surgery* (2020).

9 Spinato, G. *et al.* Alterations in Smell or Taste in Mildly Symptomatic Outpatients With SARS-CoV-2 Infection. *JAMA*, doi:10.1001/jama.2020.6771 (2020).

10 Drew, D. A. *et al.* Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Preprint at h*[*ttps://www.medrxiv.org/content/10.1101/2020.04.02.20051334v1,*](https://www.medrxiv.org/content/10.1101/2020.04.02.20051334v1) (2020).

11 Oleszkiewicz, A., Kunkel, F., Larsson, M. & Hummel, T. Consequences of undetected olfactory loss for human chemosensory communication and well-being. *Philos Trans R Soc Lond B Biol Sci* **375**, 20190265, doi:10.1098/rstb.2019.0265 (2020).

12 LeDell, E., Petersen, M. & van der Laan, M. Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates. *Electron J Stat* **9**, 1583-1607, doi:10.1214/15-Ejs1035 (2015).

Acknowledgments

We express our sincere thanks to all the participants of the app, including study volunteers enrolled in cohorts within the Coronovirus Pandemic Epidemiology (COPE) consortium. We thank the staff of Zoe Global Limited, the Department of Twin Research at King's College London, the Clinical & Translational Epidemiology Unit at Massachusetts General Hospital, for their tireless work in contributing to the running of the study and data collection. We would like to thank Eran Segal and his lab for helpful input. This work was supported by Zoe Global Limited. The Dept of Twin Research receives support from grants from the Wellcome Trust (212904/Z/18/Z) and the Medical Research Council (MRC)/British Heart Foundation Ancestry and Biological Informative Markers for Stratification of Hypertension (AIMHY; MR/M016560/1)., European Union, Chronic Disease Research Foundation (CDRF), Zoe Global Ltd, NIH and the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. C.M. is funded by the Chronic Disease Research Foundation and by the MRC Aim-Hy project grant, AMV is supported by the National Institute for Health Research Nottingham Biomedical Research Centre. C.H.S. is an Alzheimer's Society Junior Fellowship AS-JF-17-011; SO and M.J.C. are funded by the Wellcome/EPSRC Centre for Medical Engineering (WT203148/Z/16/Z), Wellcome Flagship Programme (WT213038/Z/18/Z). A.T.C. is the Stuart and Suzanne Steele MGH Research Scholar and is a Team Leader for the Stand Up to Cancer Foundation. A.T.C., L.H.N., and D.A.D. are supported by an Evergrande COVID-19 Response Fund Reward through the Massachusetts Consortium on Pathogen Readiness (MassCPR).

Authors' contribution

***Conceived and designed the experiments*:** C.M., A.M.V., C.J.S., T.D.S.; ***Analyzed the data*:** C.M., M.B.F., C.H.S., S.G., T.V., M.J.C., SO, H.N., A.M.V.. ***Contributed reagents/materials/analysis tools**:*

T.W., J.C., M.B.F., S.G., A.V., J.S.E.-M., P.H., M.M., M.F., S.E., D.A.D., A.T.C., L.H.N. ***Wrote the manuscript***: C.M., A.M.V. ***Revised the manuscript***: all

Competing interests' declaration: T.D.S., A.M.V. are consultants to Zoe Global Ltd ("Zoe"). S.G. and J.W. are employee of Zoe Global Limited. Other authors have no conflict of interest to declare.

Data availability

Data collected in the app is being shared with other health researchers through the NHS funded Health Data Research UK (HDRUK)/SAIL consortium, housed in the UK Secure Research Platform (UKSeRP) in Swansea. Anonymised data is available to be shared with bonafide researchers HDRUK according to their protocols in the public interest. See https://healthdatagateway.org/detail/9b604483-9cdc-41b2-b82c-14ee3dd705f6.

US investigators are encouraged to coordinate data requests through the COPE Consortium (www.monganinstitute.org/cope-consortium). Data updates can be found on https://covid.joinzoe.com

Code availability

The app code is public at https://github.com/zoe/covid-tracker-react-native.

Table 1. Characteristics of the study population. Results are presented as (%) for dichotomous traits, as mean(SD) for continuous traits

	Tested for SARS-CoV-2				*Untested for SARS-CoV-2*	
	UK		**US**		**UK**	**US**
	Tested positive	**Tested negative**	**Tested positive**	**Tested negative**		
N	6,452	9,186	726	2,037	2,434,931	165,530
%Females	71.88%	76.40%	78.10%	82.28%	64.24%	68.67%
Age, yrs	41.25(12.18)	41.87(12.14)	44.65(14.31)	47.25(13.80)	43.38 (14.69)	53.00 (16.68)
BMI, kg/m²	27.34(6.06)	27.04(5.91)	28.04(6.91)	27.71(636)	26.39 (5.44)	27.08 (5.87)
Answered questions on symptoms	6,452	9,186	726	2,037	773,539	32,214
Loss of smell and taste	64.76%	22.68%	67.49%	17.33%	9.5%	12.1%
Fatigue	29.80%	15.56%	23.42%	21.50%	8.71%	8.17%
Shortness of breath	15.27%	9.37%	13.50%	12.03%	4.52%	3.65%
Fever	34.34%	23.93%	27.74%	26.25%	9.31%	4.78%
Persistent cough	56.73%	45.56%	44.77%	41.24%	30.17%	22.25%
Diarrhoea	25.95%	19.24%	32.92%	29.06%	19.33%	23.26%
Delirium	17.87%	13.25%	23.55%	23.47%	15.26%	16.25%
Skipped meals	42.03%	24.93%	47.25%	36.08%	19.12%	21.42%
Abdominal pain	21.31%	18.24%	20.94%	21.06%	3.70%	15.94%
Chest pain	42.73%	39.17%	38.57%	41.92%	18.65%	24.74%
Hoarse voice	31.70%	25.85%	26.86%	25.72%	7.86%	13.79%

Figure Legends

Figure 1. a. Association between symptoms and SARS-CoV-2 infection in 15,638 UK and 2,763 US participants who were tested via RT-PCR, OR(95%CI). **b.** ROC-curve for prediction in the UK test set of the risk of positive test for SARS-CoV-2 using self-reported symptoms and traits: persistent cough, fatigue, skipped meals, loss of smell and taste, sex, age. **c** ROC-curve for prediction in the US validation set of the risk of positive test for SARS-CoV-2 using self-reported symptoms and traits: persistent cough, fatigue, skipped meals, loss of smell and taste, sex, age. AUC=area under the curve; SE=sensitivity; SP=specificity; PPV=positive predictive value; NPV= negative predictive value.

Online Methods

Study setting and participants

The COVID Symptom Tracker smartphone based app was developed by Zoe Global Limited, King's College London, and Massachusetts General Hospital, and was launched in the UK on Tuesday the 24th March 2020, and in the US on 29th March 2020 and after three weeks has reached 2,618,862 users. It enables capture of self-reported information related to COVID-19 infections as described previously[10]. The survey questions are available in **Supplemental Table 1**. On first use, the app records self-reported location, age, and core health risk factors. With continued use and notifications, participants provide daily updates on symptoms, health care visits, COVID-19 testing results, and if they are self-quarantining or seeking health care, including the level of intervention and related outcomes. Individuals without apparent symptoms are also encouraged to use the app.

Ethics: The Ethics for the app has been approved by KCL ethics Committee and all users provided consent for non-commercial use. An informal consultation with TwinsUK members over email and social media prior to the app having been launched found that they were overwhelmingly supportive of the project. The US protocol was approved by the Partners Human Research Committee.

Statistical analysis

Data from the app were downloaded into a server and only records where the self-reported characteristics fell within the following ranges were utilised for further analyses: age between 16 (18 in the US) and 90; height (cm) between 110 and 220; weight (kg) between 40 and 200; BMI(kg/m^2) between 14 and 45; and temperature (in C) between 35 and 42. The individuals whose data was included to develop and test the prediction model were those who completed the report for symptoms in the app, who declared to have had a SARS-CoV-2 RT-PCR test and to have received the result for the test. Only individuals who answered at least 9 of the 10 symptoms and answered about loss of smell and taste were included.

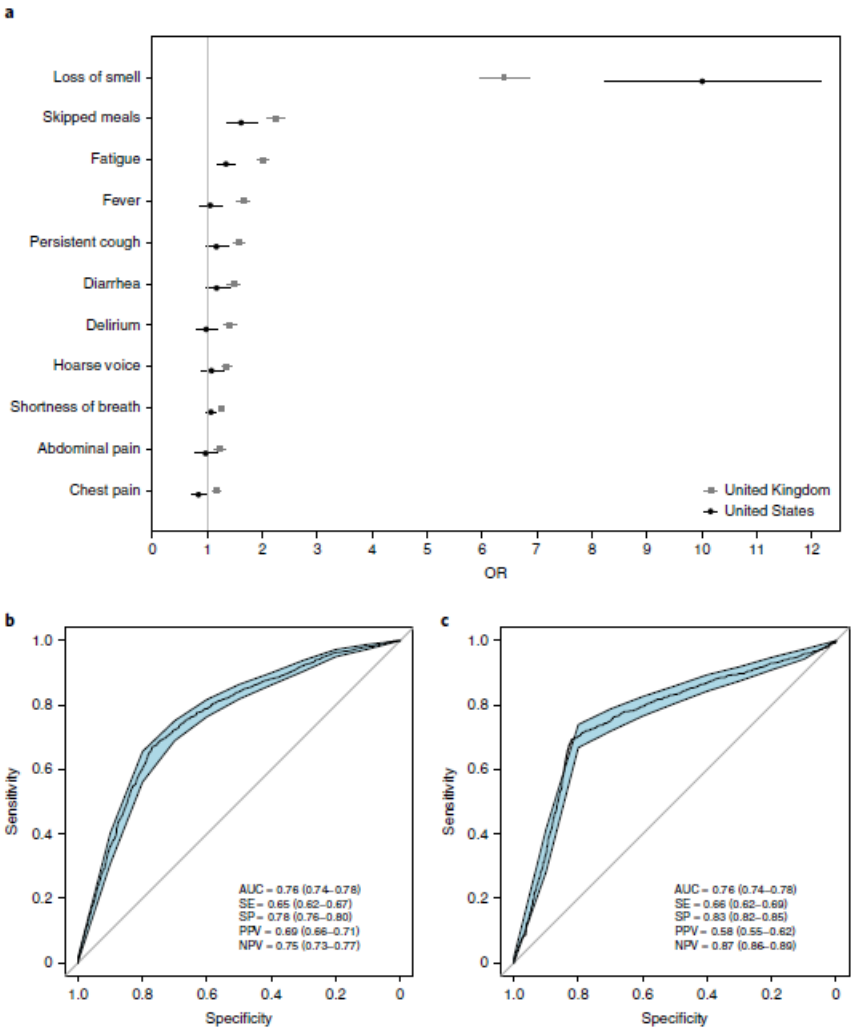
Baseline characteristics are presented as the number (percentage) for categorical variables and the mean (standard deviation) for continuous variables. Multivariate logistic regression adjusting for age, sex and BMI was applied to investigate the correlation between loss of smell and taste and COVID-19 in 15,368 UK users of the symptom tracker app who were also tested in the lab for SARS-CoV-2 (6,452 UK individual tested positive and 9,186 tested negative). Results were replicated in 726 US individual tested positive and 2,037 US individual tested negative. We then randomly split the UK sample into training and test sets with 80/20 ratio. In the training set, we performed stepwise logistic regression combining forward and backward algorithms, to identify other symptoms associated to COVID-19 independently of loss of smell and taste. We included in the model ten other symptoms (including fever, persistent cough, fatigue, shortness of breath, diarrhoea, delirium, skipped meals, abdominal pain, chest pain and hoarse voice) as well as age, sex, and BMI and chose, as the best model, the one with the lowest AIC. We then assessed the performance of the model both it the test set and via 10-fold cross-validation in the whole UK sample of 15,638 individuals using the R package cvAUC[12]. We further validated the prediction model in the US cohort.

For our predictive model, using the R packages pROC and epiR, we further computed the area under curve (AUC) i.e. the overall diagnostic performance of the model, the sensitivity ("positivity in disease") i.e. the proportion of subjects who have the target condition (reference standard positive) and give positive test results, and the specificity ("negativity in health") i.e. the proportion of subjects without a SARS-CoV-2 RT-PCR test that give negative model results.

Finally, we applied the predictive model to the 805,753 individuals reporting symptoms who had not had a SARS-CoV-2 test in order to estimate the percentage of individuals reporting some COVID-19 symptoms likely to be infected by the virus. The proportion of estimated infections was calculated repeatedly by sampling the dataset (with replacement) to get the 95% confidence intervals.

Reporting Summary

Further information on research design is available in the Life Sciences Reporting Summary linked to this article.


a
Loss of smell
Skipped meals
Fatigue
Fever
Persistent cough
Diarrhea
Delirium
Hoarse voice
Shortness of breath
Abdominal pain
Chest pain
United Kingdom
United States
0 1 2 3 4 5 6 7 8 9 10 11 12
OR
b
Sensitivity
1.0 0.8 0.6 0.4 0.2 0
AUC = 0.76 (0.74–0.78)
SE = 0.65 (0.62–0.67)
SP = 0.78 (0.76–0.80)
PPV = 0.69 (0.66–0.71)
NPV = 0.75 (0.73–0.77)
1.0 0.8 0.6 0.4 0.2 0
Specificity
c
Sensitivity
1.0 0.8 0.6 0.4 0.2 0
AUC = 0.76 (0.74–0.78)
SE = 0.66 (0.62–0.69)
SP = 0.83 (0.82–0.85)
PPV = 0.58 (0.55–0.62)
NPV = 0.87 (0.86–0.89)
1.0 0.8 0.6 0.4 0.2 0
Specificity