This is an Accepted Manuscript for Epidemiology & Infection. Subject to change during the editing and production process.

DOI: 10.1017/S0950268824000037

THE PREDICTIVE ROLE OF SYMPTOMS IN COVID-19 DIAGNOSTIC MODELS – A LONGITUDINAL INSIGHT

Olivia Bird,[1,2] Eva P. Galiza,[1] David Neil Baxter,[3] Marta Boffito,[4] Duncan Browne,[5] Fiona Burns,[6] David R. Chadwick,[7] Rebecca Clark,[8] Catherine A. Cosgrove,[1] James Galloway,[9] Anna L. Goodman,[10] Amardeep Heer,[11] Andrew Higham,[12] Shalini Iyengar,[13] Christopher Jeanes,[14] Philip A. Kalra,[15] Christina Kyriakidou,[16] Judy M. Bradley,[17] Chigomezgo Munthali,[18] Angela M. Minassian,[19] Fiona McGill,[20] Patrick Moore,[21,22] Imrozia Munsoor,[23] Helen Nicholls,[24] Orod Osanlou,[25] Jonathan Packham,[26,27] Carol H. Pretswell,[28] Alberto San Francisco Ramos,[1] Dinesh Saralaya,[29] Ray P. Sheridan,[30] Richard Smith,[31] Roy L. Soiza,[32] Pauline A. Swift,[33] Emma C. Thomson,[34] Jeremy Turner,[35] Marianne Elizabeth Viljoen,[36] Paul T. Heath,[*1] and Irina Chis Ster[*37]

* contributed equally

[1]Vaccine Institute, St. George's , University of London and St. George's University Hospitals National Health Service Foundation Trust, London, United Kingdom; [2]Oxford University Hospitals NHS trust; [3]Medical Education, Stockport National Health Service Foundation Trust, Stepping Hill Hospital, Poplar Grove, Stockport, United Kingdom; [4]Chelsea and Westminster Hospital National Health Service Foundation Trust and Faculty of Medicine, Imperial College London, London, United Kingdom; [5]Endocrinology/Diabetes/General Medicine, Royal Cornwall Hospitals National Health Service Trust, Truro, United Kingdom; [6]Faculty of Population Health Sciences, Institute for Global Health, University College London, and Royal Free London National Health Service Foundation Trust, London, United Kingdom; [7]Centre for Clinical Infection, South Tees Hospitals National Health Service Foundation Trust, James Cook University Hospital, Middlesbrough, United Kingdom; [8]Layton Medical Centre, Blackpool, United Kingdom; [9]Centre for Rheumatic Disease, Kings College London, London, United Kingdom; [10]Department of Infectious Diseases, Guy's and St Thomas' National Health Service Foundation Trust, and Medical Research Council Clinical Trials Unit at University College London, London, United Kingdom; [11]Lakeside Healthcare Research, Lakeside Surgeries Corby, Northants, United Kingdom; [12]Gastrointestinal and Liver Services, University Hospitals of Morecambe Bay National Health Service Foundation Trust, Kendal, United Kingdom; [13]Accelerated Enrollment Solutions, Synexus Hexham Dedicated Research Site, Hexham General Hospital, Hexham, United Kingdom; [14]Department of Microbiology, Norfolk and Norwich University Hospitals National Health Service Foundation Trust, Norwich, Norfolk,

United Kingdom; [15]Nephrology, Salford Royal Hospital, Northern Care Alliance National Health Service Foundation Trust, Salford, United Kingdom; [16]Accelerated Enrollment Solutions, Synexus Midlands Dedicated Research Site, Birmingham Research Park, Birmingham, United Kingdom; [17]Dentistry and Biomedical Sciences, School of Medicine, Wellcome-Wolfson Institute for Experimental Medicine, Queen's University of Belfast, Belfast, Northern Ireland, United Kingdom; [18]Accelerated Enrollment Solutions, Synexus Merseyside Dedicated Research Site, Burlington House, Waterloo, Liverpool, United Kingdom; [19]Centre for Clinical Vaccinology and Tropical Medicine, University of Oxford, and Oxford Health National Health Service Foundation Trust, Warneford Hospital, Oxford, United Kingdom; [20]Microbiology, Leeds Teaching Hospitals National Health Service Trust, Leeds, United Kingdom; [21]The Adam Practice, Poole, Dorset, United Kingdom; [22]University Hospital Southampton National Health Service Foundation Trust, Southampton, United Kingdom; [23]Accelerated Enrollment Solutions, Synexus Glasgow Dedicated Research Site, Venture Building, Kelvin Campus, Glasgow, Scotland, United Kingdom; [24]Accelerated Enrollment Solutions, Synexus Wales Dedicated Research Site, Riverside Court Gwaelod-y-Garth, Cardiff, Wales, United Kingdom; [25]School of Medical Sciences (Pharmacology/Pharmacy), Bangor University, and Clinical Pharmacology and Therapeutics/General Internal Medicine, Betsi Cadwaladr University Health Board, Wales, United Kingdom; [26]Academic Unit of Population and Lifespan Sciences, University of Nottingham, Nottingham, United Kingdom; [27]Rheumatology Department, Haywood Hospital, Midlands Partnership National Health Service Foundation Trust, Stafford, United Kingdom; [28]Accelerated Enrollment Solutions, Synexus Lancashire Dedicated Research Site, Matrix Park Buckshaw Village, Chorley, Lancashire, United Kingdom; [29]National Institute for Health Research Patient Recruitment Centre and Bradford Teaching Hospitals National Health Service Foundation Trust, Bradford, United Kingdom; [30]Geriatric Medicine, Royal Devon University Healthcare, Exeter, Devon, United Kingdom; [31]Nephrology, East Suffolk and North Essex National Health Service Foundation Trust, United Kingdom; [32]Aberdeen Royal Infirmary and Ageing Clinical and Experimental Research Group, University of Aberdeen, Aberdeen, Scotland, United Kingdom; [33]Renal Services, Epsom and St Helier University Hospitals National Health Service Trust, London, United Kingdom; [34]School of Infection & Immunity, Medical Research Council-University of Glasgow Centre for Virus Research, and Queen Elizabeth University Hospital, National Health Service Greater Glasgow & Clyde, Glasgow, Scotland, United Kingdom; [35]Diabetes and Endocrinology, Norfolk and Norwich University Hospitals National Health Service Foundation Trust, Norwich, Norfolk, United Kingdom; [36]Accelerated Enrollment Solutions, Synexus Manchester Dedicated Research Site, Kilburn House, Manchester, United Kingdom, [37] Institute of Infection and Immunity, George's, University of London

Corresponding author: ichisste@sgul.ac.uk

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Conflicts of Interest:

CAC reports receiving grant support, paid to her institution, from Novavax, Moderna, GSK.

ALG reports receiving grant support, paid to her institution, from Novavax and entered into a partnership with AstraZeneca for further development of ChAdOx1 nCoV-19. A. L. G. is named as an inventor on a patent covering use of a particular promoter construct that is often used in vectored vaccines and is incorporated in the ChAdOx1 nCoV-19 vaccine and may benefit from royalty income paid to the University of Oxford from sales of this vaccine by AstraZeneca and its sublicensees under the university's revenue sharing policy.

PTH reports receiving grant support, paid to his institution, from Novavax, Pfizer, Moderna, Valneva, Janssen, Astra Zeneca.

ICS declares receiving grant support, paid to her institution, from NIHR and Astra Zeneca.

Other authors report no conflicts of interest.

Disclaimer: the findings and conclusions presented here are the authors and do not necessarily represent the views of Novavax themselves, although the affiliated authors were given the opportunity to review the submission and provide feedback.

Financial support: no specific funding.

Data availability: The data are available upon request and subject to Novavax's permission. Please contact Professor Paul Heath pheath@sgul.ac.uk.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Introduction

The SARS-COV-2 pandemic has contributed to significant global morbidity and mortality. As of the 7th of March 2023, there have been over 759 million cases of COVID-19, including 6.8 million deaths[1]. The burden of disease was greatly felt by all public health organizations, but particularly on healthcare systems which were frequently put under strain as they managed surges of infections[2]. The unprecedented scale and speed of the pandemic, its similarities to influenza and the three major foci of care homes, hospitals and the community, proved to be a challenging combination for devising a standard list of symptoms for COVID-19. Accurate recognition of the symptoms that indicated infection and warranted urgent testing was particularly important in the early stages of the pandemic when Polymerase Chain Reaction (PCR) testing kits were in demand[3].

The gold standard for diagnosing SARS-COV-2 infection is an oropharyngeal/nasal PCR swab, although latterly Lateral Flow Tests are used for rapid diagnosis[4]. In the UK, PCR testing was initially prioritised to those presenting with a new (or worsening) cough, fever, or breathlessness[5]. However other symptoms, such as altered or loss of smell (anosmia) or taste (ageusia), and gastro-intestinal symptoms (such as loss of appetite and diarrhoea) have been associated with COVID-19[6–8]. In a Cochrane Review (2021), mainly based on more severely affected populations (e.g. hospitalised patients), the pooled specificities for anosmia and ageusia were high (90.5%) suggesting these symptoms may be a useful marker for COVID-19[9]. The updated review (2022) concluded that most other individual symptoms had poor diagnostic accuracy[10].

In a study of 483 subjects in Washington DC of whom 42% were healthcare or essential workers, aged between 25-44 years, who retrospectively reported symptoms, 27% were reported to be PCR

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

positive. Wojtusiak et al. concluded that clusters of symptoms are more predictive of COVID-19 than any one specific symptom[11] . In a different study, the same authors also examined the importance of the order of symptom occurrence in deriving a disease diagnostic model[12]. A meta-analysis based on sample data collected from nine established longitudinal cohorts designed a 4-category cross-sectional outcome aiming at capturing characteristics of long COVID in the UK population[13]. Based on questionnaires completed by subsets of participants between July 2020 and September 2021 and self-reported COVID results as well as presence/absence of symptoms, the meta-analysis demonstrated considerable heterogeneity between studies[13].

The observation of previous research is that there is a great deal of variation in data collection methods (e.g. smartphone apps, patient records[14–16]), epidemiological heterogeneity of study populations (e.g. hospitals, Intensive Care Units, care homes[13–15]) and different reporting methods (e.g. self-reports, interviews[17]). As symptoms develop over time, cross-sectional outcomes and retrospectively collected information on symptoms may be difficult to relate to COVID-19 onset which is also known to have a variable incubation period (2-14 days)[18]. The Zoe Health study compared three different symptom-based diagnostic models for SARS-CoV-2 and investigated the effect of demographic variables on the models' performance metrics and found that the discrimination power of all models improved with the number of days of symptoms included, whilst the most relevant symptoms for detecting COVID-19 were anosmia and chest pain[12].

The UK phase 3 Novavax COVID-19 clinical trial was conducted at 33 sites and recruited 15,185 participants[19]. Its primary aim was to evaluate the efficacy and safety of the vaccine. We used the prospectively reported symptoms of possible SARS-CoV-2 infection to assess the discrimination

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

power of individual symptoms and to investigate an optimal combination to generate a diagnostic model for the presence of SARS-CoV-2 infection in the UK population.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Methods

The data for this analysis were provided by Novavax, Inc.[19]. The methods and results of the trial are described elsewhere[19]. Data included are from 28th October 2020 until 28th February 2021.

Monitoring for COVID-19

All participants had a SARS-CoV-2 PCR test performed at recruitment and were tested for symptomatic infection throughout the study. Participants were instructed to contact the study team within 24 hours if they self-assessed COVID-19 symptoms **(Table 1**), triggering a surveillance visit. Throat/nasal swabs were self-collected by participants approximately 24 hours after the onset of symptoms, then daily for up to 3 days. A participant with suspected or confirmed COVID-19 was asked to complete a symptom diary, starting on their first day of symptoms, reporting daily for a minimum of 10 days (even if their symptoms resolved and regardless of SARS-CoV-2 PCR result). Participants with confirmed symptomatic COVID-19, signified by a positive PCR test, continued documenting their symptoms until resolution. Virologic confirmation was performed by PCR assay at the U.K. Department of Health and Social Care laboratories with the TaqPath system (Thermo Fisher Scientific).

Statistical methodology

The main objective was to construct an optimal diagnostic model for COVID-19 based on participants' symptoms and to highlight differences in the dynamics of specific symptoms in groups defined by participants who experienced COVID-19 and those who did not. To extrapolate the results to the UK population we started by plotting and empirically comparing the distribution of age, gender and ethnicity distributions in the sample data to that of the UK population[20–22]. We

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

then used post-stratification techniques for incorporating population demographic distributions[23]. This procedure allowed us to produce estimates generalisable to the UK community population. Weights were derived and assigned to each participant such that the subsequent estimation procedures inflated the effect of under-represented groups (e.g. young ethnic minorities) and depressed the effect of overrepresented groups in the sample (e.g. old White).

We constructed a master file which included multiple PCR tests per participant and multiple symptomatic episodes. The resulting data have a hierarchical structure with implications on the subsequent choice of analyses and estimation procedures (details in **Supplementary Information**). Participants were initially grouped by their PCR results, i.e. participants with at least one PCR positive result and those always negative. We reported the frequency and proportion of the symptomatic participants in the two groups. We estimated the probabilities of testing positive given a specific symptomatic episode, and the mean number of reports (or number of days) of a specific symptom within an illness episode. We also investigated the symptom report dynamics and explored the extent to which symptoms were associated with demographics. These analyses identified the main confounder candidates and their potential influence for the subsequent receiver operating characteristic (ROC) analyses.

Non-parametric techniques such as local polynomial smoothing have been used to fit curves on the daily probabilities of the reports in the PCR+ and PCR- participants. A heatmap of daily probabilities of reported symptoms has also been presented in ascending order of their magnitude on the first day in positive patients.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

We assessed the effect of reporting the number of days of each specific symptom on the probability of testing PCR positive (PCR+) vs PCR negative (PCR-), measured as the odds ratios and their 95% CIs. We derived a symptom-based diagnostic model using two-level logistic regression and evaluated the discriminatory power of this model using area under the curve (AUC) as a metric for its discrimination. We also performed a two-stage process ROC analysis[24]. The technique allows for multiple episodes to be associated with an individual, and adjustments using population weights. The result is an estimate of the ROC curve for each specific symptom as a function of age and ethnicity – known as a covariate specific ROC curve[24]. Using these techniques, we have also highlighted the increasing discrimination power of individual symptoms based on the temporally ordered reports restricted to the first 1, 2, 3 to longer than 15 days after the start of the symptomatic illness episode. The effect of age and ethnicity on the discrimination power of individual symptoms were also evaluated. More details in **Supplementary information**.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Results

Data summary

Table 2 displays a simplified picture of the data based on a binary assessment. From 15,139 participants, 317 (2.1%) had a PCR+ episode and 3,320 (21.9%) had at least one symptomatic episode. 8% (266/3320) of the symptomatic population were PCR+ and 84% (266/317) of the PCR+ participants reported symptoms. **Figure 1** displays the age distribution against that of the UK population stratified by gender and ethnicity[20–22]. These data have been used to calculate the weights associated with our analyses.

Table 3 presents demographic data, stratified by PCR status. The comorbidities variable indicates the presence of at least one comorbidity. COVID-19 was directly associated with younger age, i.e. one year increasing in age decreased the OR of COVID-19 by a small yet significant factor of 0.98 ($p<0.001$). Ethnic minorities (excluding white) were twice as likely to test positive than their white counterparts, i.e. OR=1.924 (95%CI (1.169, 3.167)). The other than white category included Asians (n=462 (3.1%)), Black (n=60 (0.4%)) and others (n=153 (1%)).

Summary symptoms data (overall and stratified by PCR status) are presented in **Table 4** and illustrated in **Figure 2.** Runny nose (16.9%) was the most reported symptom in this cohort, followed by cough (14.6%) and tiredness (12.6%). Nausea (5.3%), diarrhoea (4.1%) and anosmia/ageusia (3.6%) were the least reported. This ordering is preserved in PCR- participants; however, in PCR+ participants cough (75.1%) was the most frequent symptom, followed by

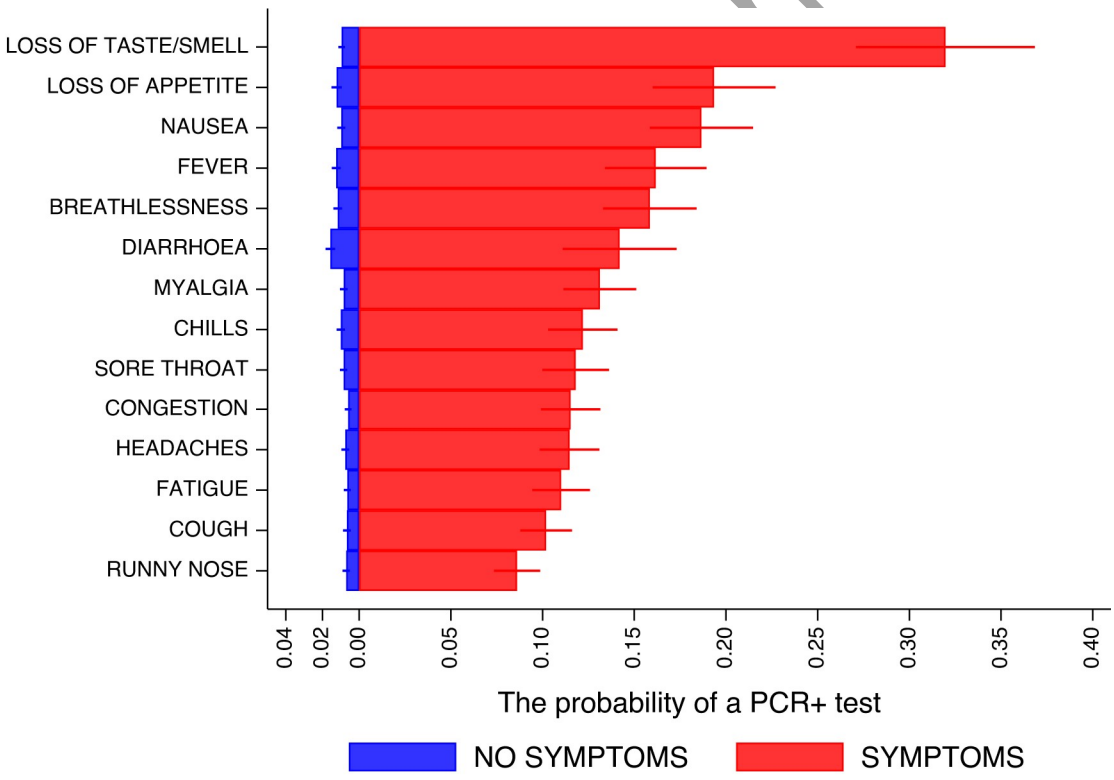
https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

congestion (74.8%) and tiredness (74.4%). Anosmia/ageusia was reported by 53.3% of PCR+ participants vs. 2.5% of PCR- participants.

The probabilities of PCR status by specific symptoms reports

Figure 3 displays the probabilities of testing PCR+ conditioned on each symptom (reported at least once). The prevalence of COVID-19 was 31.9% (27.1%-36.8%) in those reporting anosmia/ageusia and 19.4% (16%-22.7%) for loss of appetite.

The number of specific symptoms' analyses



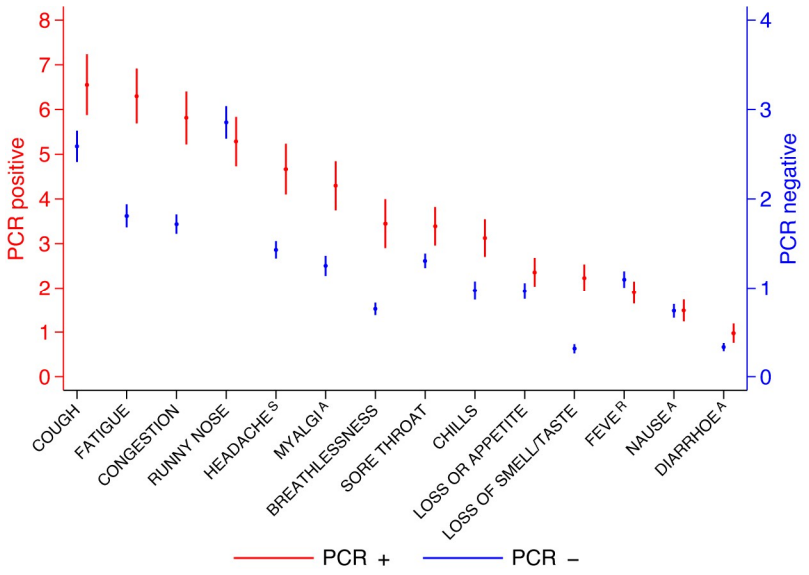
https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure 4 shows the mean number of days (and their 95% CIs) that each specific symptom was reported during a symptomatic episode, stratified by PCR status. PCR+ participants reported a significantly longer duration of specific symptoms compared to PCR- participants. For example, the mean number of days of cough was 6-7 in PCR+ participants and 2-3 in PCR- participants.

Table 5 presents an exploratory analysis on the rate ratios (fold-effects) as measures of associations between the mean number of days of specific symptoms with population characteristics, this has been also analysed in the PCR+ subgroup in **Table 6.** From **Table 5**, we learn that age was directly associated with an increased number of reports of runny nose, cough and loss of appetite, but inversely associated with sore throat and anosmia/ageusia. Women reported 24.3% (95% CI (11.4%, 38.7%)) more headaches than men. Other than white participants reported fewer symptoms than White participants; for runny nose by a factor of 0.76 (95% CI (0.65, 0.89)), cough (by a factor of 0.77 (95% (0.62, 0.95)), and congestion (by a factor of 0.77 (95% (0.62, 0.96)). Increasing BMI was associated with increased reporting of myalgia (p=0.033) and breathlessness (p<0.001). Those with co-morbidities reported 18.5% (95% CI (8.1%, 29.8%)) more days of cough, 16.1% (95% CI (1.9%, 32.2%)) more days of myalgia and 22.4% (95% CI (3.6%, 44.5%)) more days of breathlessness on average, than those without co-morbidities (**Table 5**).

In those with a positive PCR (**Table 6**) many of these trends remained significant, for example, the effect of age on myalgia (p=0.039) and loss of appetite (p=0.012), the effect of gender on headaches (p=0.033), of ethnicity on congestion (p=0.002) and of BMI on breathlessness (p=0.012). Increased BMI was associated with longer duration of cough (p=0.022).

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press



https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

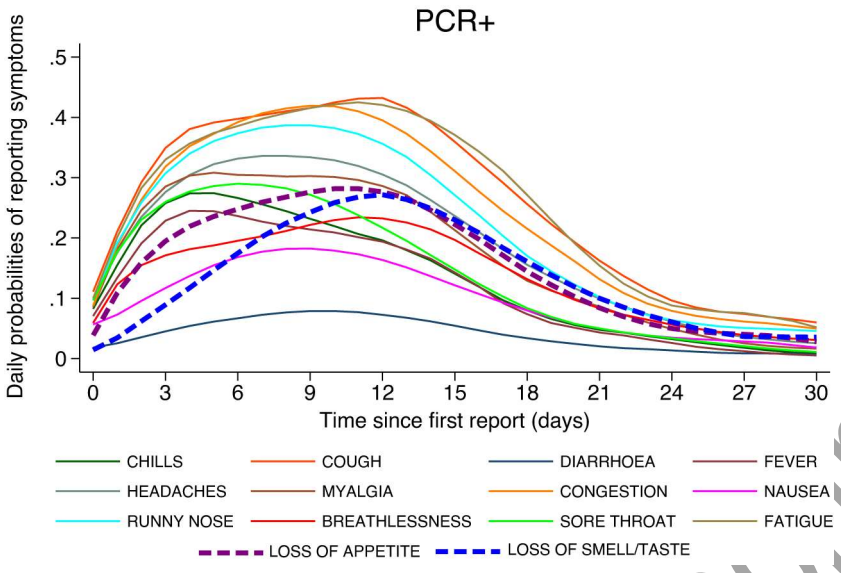
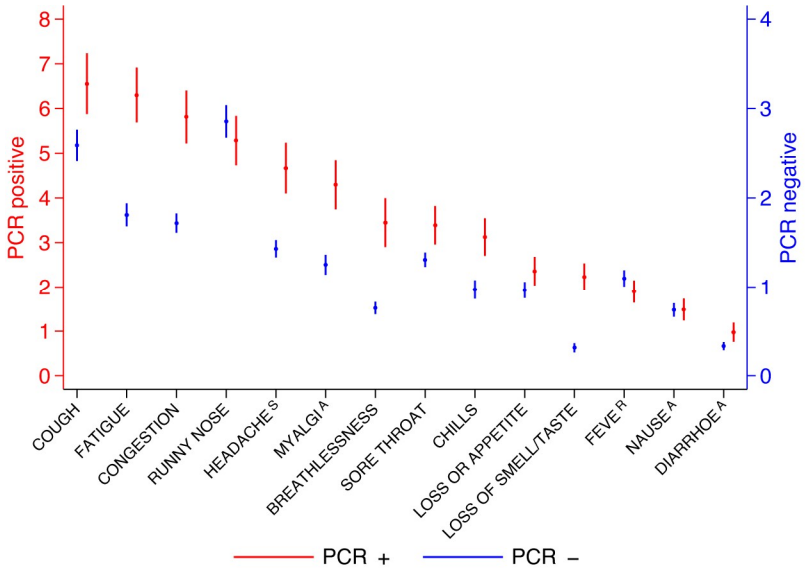


Figure 5 and

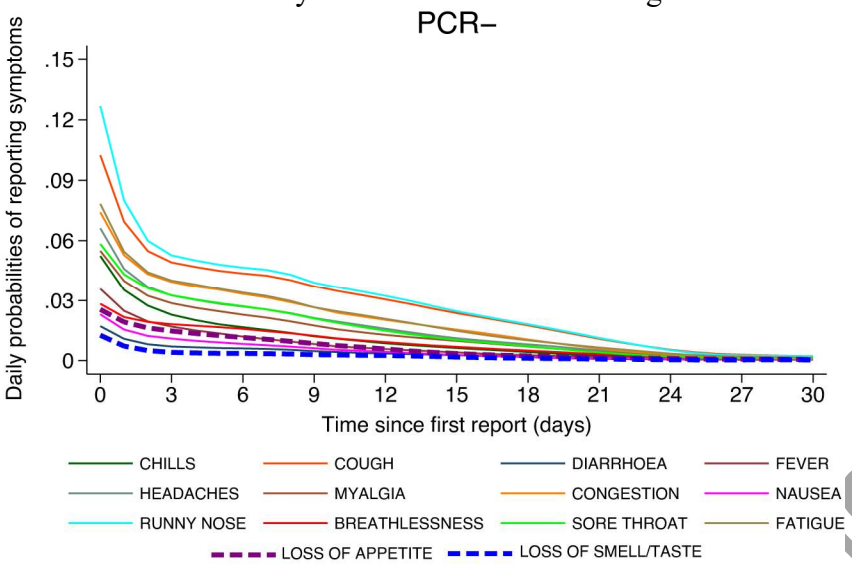
https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure *6* present the daily probabilities of specific symptoms (starting with the first report of any symptom), stratified by PCR result. Whilst these probabilities fall swiftly in PCR- participants (**Figure 6)**, they start more slowly and peak later in those with COVID-19 (



https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure 5**)**. Fever peaked on the 4th day (24%), followed by chills (27%), whilst myalgia (31%) and loss of appetite (28%) peaked on the 5th day. Anosmia/ageusia (27%) and cough (43%) peaked on the 12th day. These findings are also reflected in

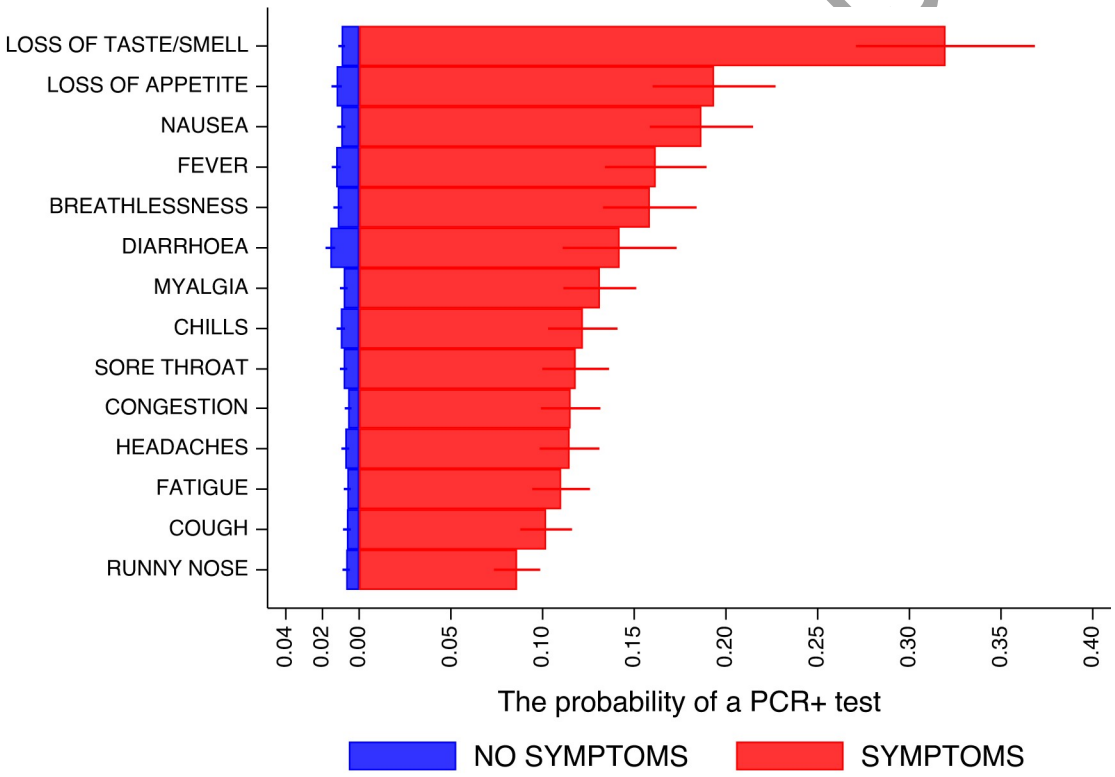


https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure 7; symptoms in PCR negative participants fall rapidly shown by the dark purple, whereas they are later to peak and slower to fade in PCR positive participants, shown by the changing colour scale.

The optimal diagnostic model for testing PCR positive based on symptoms and controlled for population characteristics

Figure 8 presents the effects (ORs) of reporting a specific symptom for 3 days within an episode, on the probability of testing PCR+. The rationale for considering 3-day symptom effect as a meaningful magnitude for the length of reports was inspired by



https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure 4**.** In this figure, all specific symptoms seem to have a mean less than 3 days in PCR- participants. Anosmia/ageusia (OR=14.4 (95%CI 9.2,22.6)), nausea (OR=5.8 (95%CI 4.2,7.9)), loss of appetite (OR=5.6 (95%CI 4.5, 7.2)) and fever (OR=5.4 (95%CI 4.2, 6.97)) have the strongest effects in terms of magnitude and statistical significance.

The most parsimonious model, i.e. the model with the least number of predictors, yet explaining the most variability in the data, is shown in **Table 7.** The model retains anosmia/ageusia (OR=5.2 (95%CI 3.4, 7.9)), loss of appetite (OR=2.3 (95%CI 1.6, 3.3)), fever (OR=1.9 (95%CI 1.4, 2.6)), congestion (OR=1.9 (95%CI 1.5, 2.4)) and cough (OR=1.3 (95%CI 1.1, 1.6)) as key symptoms associated with a PCR+ episode, whilst runny nose (OR=0.7 (95%CI 0.5,0.9)) and chills (OR=0.6 (95%CI 0.4, 0.8)) are associated with testing PCR-. This model has a discrimination power of approximatively 0.86 in terms of AUC but does not account for population weights.

Table 8 presents combinations of symptoms predicting the probabilities of COVID-19 using the optimal model. For example, a white participant of 50 years of age would have over 90% probability of testing PCR+ if s/he reported 3 days of loss of taste and smell, 3 days of loss of appetite, 3 days of fever and 3 days of cough with 1 day of congestion, runny nose and chills.

The discriminatory power of specific symptoms

Figure 9 shows how the discriminatory power of individual symptom evolves if only the first number of days after onset are considered - that is only day 1, only days 1-2, only days 1-3 and so on. Symptoms which peak later such as anosmia/ageusia gain discrimination power as the number of days of reporting increases. For other less specific symptoms, the individual discrimination

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

power remains constant or even declines, for example sore throat peaks very early and then tapers off.

The area under the curve in **Figure 10** shows the discrimination power of each symptom in the model using the maximum likelihood ROC 2-stage regression analysis (uncontrolled for age and ethnicity and population weighted). The higher the AUC, the better the symptom discriminates between PCR+ and PCR-, the steep incline of the curve followed by the flattening line suggests that discrimination is little affected as the number of false positives increases.

When controlled for age and ethnicity, the two-stage ROC model does not quantify their effects on the ROC curve of specific symptoms in a directly interpretable manner, but qualitative conclusions are displayed in **Table 9** and visualised in **Figure 11**. Age and ethnicity affect the ROC curve for each symptom, notably the discriminatory power of anosmia/ageusia decreased with increasing age and is smaller ethnic minorities, compared to White ethnicity.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Discussion

The main objectives of this study were to develop a symptom-based diagnostic model for a PCR-proven SARS-CoV-2 infection, investigate the dynamics of the symptoms and their discrimination power for a potential COVID-19 diagnostic model. Our prospective, longitudinal, real-time collection, together with analytical techniques (post-stratification weights[20–22]), which produce generalizable results to the UK adult community population, provides a better understanding on the dynamics of COVID-19 symptomology. The rather poor engagement of people other than White in COVID-19 clinical trials has been documented[25] but our method overcame this difficulty.

We found a four-month prevalence of COVID-19 of 2.1%, in line with the estimated population prevalence at that time[26]. Of the individual symptoms, anosmia and/or ageusia were the least reported symptoms overall (3.6%); however, participants reporting them for 3 days or more were more likely to test positive for COVID-19 (OR= 14.4 (9.2,22.6)). **Figure 3** presents the probabilities testing positive conditioned on symptoms reports. Also, of those testing positive for SARS-CoV-2, over half (53.3%) reported the presence of anosmia or ageusia (**Figure 2**). Other symptoms such as loss of appetite, a new fever, congestion and cough were strongly associated with a positive result. Fever, cough and anosmia/ageusia have been identified as the strongest candidates for predicting COVID-19 in studies such as a REACT-1 and also in a meta-analysis of 9 studies examining symptoms of COVID-19 and long COVID syndrome[13,17]. The odds of having COVID-19 have been reported as positively associated with shortness of breath (OR=3.1, (95%CI)), although our results do not support it as a 'leading' symptom [13]. On its own runny nose was the most reported symptom (16.9%) in our study, and frequently reported in those with confirmed COVID-19 (72.6%). The participants reporting it were the least likely (8%) to test

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

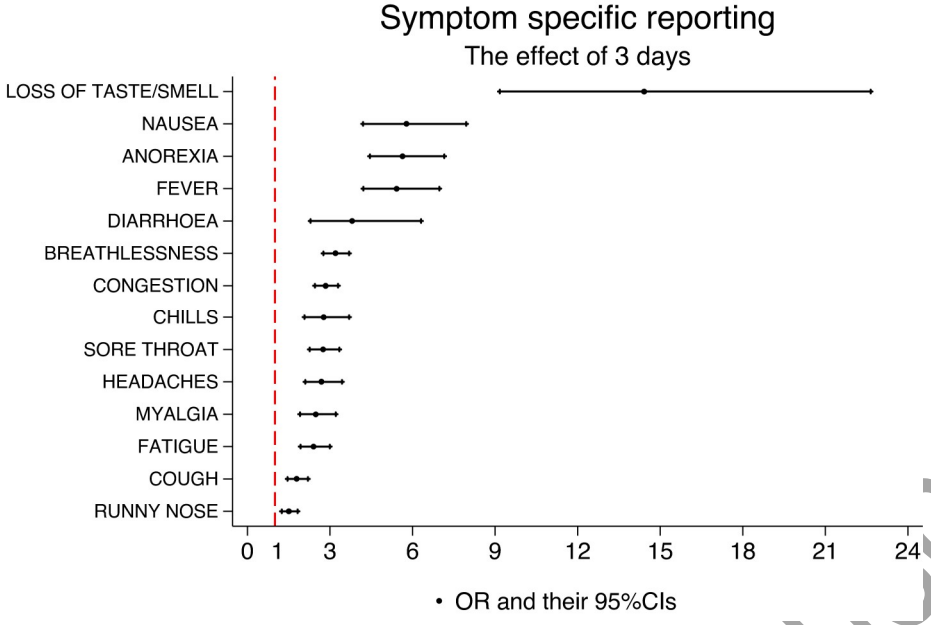
positive for COVID-19 (**Figure 3**), when accounting for the entire episode, and the symptom turned out to have high discriminatory power (AUC=0.83, Figure 9) in ruling out the disease, consistent with other findings[11,17].

Unlike many other studies[6–8,10,16], this research examined the number of days that specific symptoms are reported within an infection episode. We found that PCR+ participants reported a significantly longer duration of specific symptoms per episode, compared with those that were PCR-; cough had the longest duration followed by tiredness whilst runny nose had the longest duration among PCR- participants. We also found that cough, anosmia/ageusia and loss of appetite peaked later in SARS-CoV-2 infection, typically around day 12 (**Figure 5)**. Research in Czechoslovakia demonstrated anosmia and ageusia had a later onset than other symptoms, beginning a median of two or more days after the onset of symptoms, and lasting longer than fever or loss of appetite[27]. These findings are consistent with Wojtusiak et al who found that headaches, chills and cough were more relevant if they occurred at onset, whilst loss of taste and smell and loss of appetite had a higher relevance if they occurred later in the infection[12].

Previous research has suggested that individual symptoms are not predictive of COVID-19 on their own. Our analysis has suggested that individual symptoms would not have had sufficient predictive power for COVID-19 early in their occurrence but that this would increase with the

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

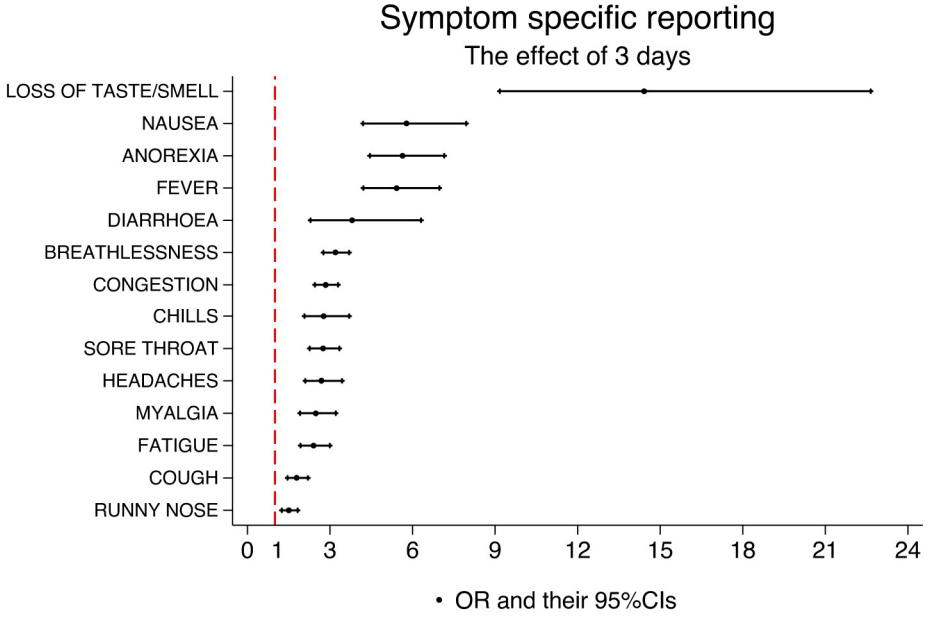
number of days in which they manifest (



https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure *9*). Hence, our final predictive model is based on specific symptomatic episodes, i.e. their entire number of symptomatic days within an episode and adjusted for age and ethnicity. The model retained episodes of anosmia/ageusia, loss of appetite, fever, congestion and cough as all positively associated with testing PCR+, together with runny nose, chills and age as all negatively associated with testing PCR+ (**Table 7**) consistent with other findings[28]. The concept of 3 days as a meaningful magnitude for the length of reports was inspired by Figure 4, in which all symptoms had a mean of less than 3 days in PCR- participants. In light of this, this information may be particularly useful at the time of clinical triage, namely the number of days symptoms have been experienced by subjects presenting for hospital care. The model, based on two-level logistic regression, has a discriminating power of ~86%.

Our ROC analysis showed that the discrimination power of anosmia/ageusia increased from irrelevance during the first few days to exceeding all others after day 9 (



https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure *9*). Our report also showed that the discriminatory power of anosmia/ageusia decreases with age, which may reflect a biological phenomenon associated with aging[29]. Cough alone remained relatively constant in its discrimination power, however PCR- participants also reported prolonged cough. Our data do not support diarrhoea as a candidate symptom of COVID-19.

Two-stage ROC analysis suggests that the prediction power may be less discriminatory in older participants and in those from ethnic minorities, this was true for all symptoms. Comparatively, the Canas et al. model showed better discrimination in participants of normal weight compared to those who were underweight and/or overweight, and in non-healthcare workers and, consistent with our results, found that younger people were more likely to test PCR positive, possibly due to increased social mixing[15]. Our diagnostic model is similar to this model as it identified persistent cough and loss of smell, alongside abdominal pain and myalgia as early features of COVID-19[15]. However, the Canas model had a younger population than our study (mean age 46.7 years vs 53.1 years) and COVID-19 was self-reported, thereby the results are difficult to compare[15]. Moreover, the study reported 'blisters on the feet' and 'eye soreness' as relevant features of COVID-19, the significance of which the paper questions itself[15].

Our estimated prevalences of specific symptoms among both positive and negative groups are higher than those presented in the meta-analysis by Bowyer et al[13]. Although the study participants stem from nine longitudinal cohorts, the data collection is essentially retrospective and cross-sectional. The authors stated a great deal of heterogeneity. Notably the data have been collected during the summer whilst ours were collected during the winter, including Christmas, when transmission intensified, hence we postulate that variation could be attributable to the season. Our

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

prevalence of specific symptoms among PCR+ and PCR- are closest to those from Generation Scotland cohort (access via Bowyer et al. or from University of Edinburgh)[13,30] consistent with our explanation above, given somewhat cooler temperatures in Scotland during the summer. We have retrieved some partial information and appended a relevant comparative Table in the **Supplementary information**.

Though multiple centres participated in the clinical trial, the three level regression techniques did not reveal important differences in the estimates or their standard errors. Variability between the centres was not expected to be significant as the same trial protocol and procedures were used. We have disregarded the effect of the intervention (placebo or vaccine), as preliminary analysis did not show a significant impact on results (data not shown).

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Limitations

Despite the data being gathered prospectively and in real-time, we observed gaps in the daily records, for example, a participant may report fever for 3 consecutive days, then none on the fourth day and then again on the fifth and sixth days. The statistical analysis considered the number of reports (i.e. the number of days with specific symptoms) rather than the whole length of time they were experienced. This may have led to underestimating their effect; however, we are confident that recall bias has been minimalised to a greater extent than if the data had been collected from a retrospectively collected self-report. Asymptomatic infections are likely to be underrepresented in this analysis. As this research set out to explore symptoms of COVID-19 we don't believe this to be a major limitation to our analysis, but it does mean we cannot calculate the true prevalence of COVID-19 infection in the study population, Unfortunately, we also did not benefit from information such as recent contacts or travel/work patterns which could have been useful in building a reliable diagnostic model as suggested by the Cochrane Review paper[10]. At the time of data collection, the circulating strain of SARS-CoV-2 was the alpha variant[31], however omicron has a higher tropism for naso-epithelial cells than pulmonary cells[32] and anosmia has been reported less frequently with the omicron variant[33]. Therefore, care should be taken if applying the model outside our study population.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Conclusion

This research adds to the body of literature on COVID-19 symptoms as an in-depth exploration of symptoms reported by those unaware of their diagnosis at the time of reporting, thereby minimising reporting bias. We found younger participants, and those from ethnic minorities were more likely to test positive for COVID-19 and, consistent with previous research, anosmia and/or ageusia most strongly predict a positive PCR result; however, we have also shown that these symptoms peak late in infection. This calls into question their consideration as early markers of the disease. Similar to other research we found that a cluster of fever, congestion and cough are all positively associated with COVID-19, with PCR positive participants reporting more days of symptoms e.g., cough, than those who were PCR negative. We also found that diarrhoea, runny nose and chills are not indicative of COVID-19. Overall, our model has a discriminating power of 86% to predict COVID-19; although, as anosmia and ageusia often develop later in the infection, our proposed model is unlikely to identify early infections, particularly in the elderly or those from ethnic minorities.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Acknowledgements

2019nCoV-302 Study Group Members

The NVX-CoV2373-2019nCoV-302 clinical trial was a collective group effort across multiple institutions and locations. Below is a list of sites and staff that significantly contributed to the implementation and conduct of the NVX-CoV2373-2019nCoV-302 clinical trial.

Site	Investigators
Aberdeen Royal Infirmary, NHS Grampian	Roy L. Soiza, Robin Brittain-Long, Chiara Scicluna, Carole Edwards, Lynn Mackay, Mariella D'Allesandro, Amy Nicol, Karen Norris, Sandra Mann, Heather Lawrence, Ruth Valentine
Accelerated Enrollment Solutions	Marianne Elizabeth Viljoen, Carol H. Pretswell, Helen Nicholls, Imrozia Munsoor, Agnieszka Meyrick, Christina Kyriakidou, Shalini Iyengar, Arham Jamal, Nick Richards, Helen Price, Bridie Rowbotham, Danielle Bird, Karen Smith, Olga Littler, Kirsty Fielding, Anna Townsend-Rose, Karen Miller, Jessica Davis, Alison Elliot-Garwood, Lauren Trottier, Paul Edwards
Belfast Health and Trust	Margaret McFarland
Betsi Cadwaladr University Health Board	Thomas Eadsforth, Jonathan Heseltine, Nick Heseltine, Rebecca Andrews, Lynne Grundy, Laura Longshaw, Julia Parton
Blandford Group Practice	Katharine Lucy Broad
Bradford Teaching Hospitals NHS Foundation Trust	Karen Regan, Kim Storton, Declan Ryan-Wakeling, Brad Wilson, Malathy Munisamy, John Wright, Anil Shenoy, Beverley English, Lucy Brear
Centre for Clinical Vaccinology and Tropical Medicine, University of Oxford	Paola Cicconi
Chelsea and Westminster Hospital	Marta Boffito, Ana Milinkovic, Ruth Byrne, Roya Movahedi, Rosalie Housman, Julie Logan, Alfredo Soler-Carracedo, Veronica Canuto, Serge Fedele, Candida Fernandez, Liam Sutcliffe
County Durham & Darlington NHS Foundation Trust	Naveed Kara, Ellen Brown, Andrea Kay
Department of Psychiatry, University of Oxford, NIHR Oxford Health Cognitive Health Clinical Research Facility and NIHR Oxford Health Biomedical Research Centre	Andrea Cipriani, Mary-Jane Attenburrow, Katharine A. Smith
Division of Epidemiology and Public Health, University of Nottingham	Jonathan Packham
Dorset Research Hub, Royal Bournemouth Hospital, University Hospitals Dorset NHS Foundation Trust	Geoff Sparrow
East Suffolk and North Essex NHS Foundation Trust	Richard Smith, Josephine M Rosier, Khalid Saja, Nyasha Nago, Brian Camilleri, Anita Immanuel, Mike Hamblin, Rawlings Osagie, Mahalakshmi Mohan
Epsom and St Helier University Hospitals NHS Trust	Hilary Floyd, Suzanne Goddard, Sanjay Mutgi, John Evans, Sean McKeon, Neringa Vilimiene, Rosavic Chicano, Rachel Hayre, Alice Pandaan
Faculty of Health and Life Sciences, Oxford Brookes University	Catherine Henshall
Guy's and St Thomas' NHS Foundation Trust NIHR BRC	Anna Goodman, Cherry Paice, Naimh Spence, Alice Packham, Movin Abeywickrama, Teona Serafimova, Suhail Aslam, Tanveer Bawa, Sonia Serrano, Moncy Mathew, Karen Bisnauthsing, Samantha Broadhead, Grainne Cullen, Jo Salkeld, Henry Fok, Thurkka Rajeswaran, Andrea Mazzella
Health and Care Research Wales	Nicola Williams, Jayne Goodwin
Highcliffe Medical Centre	Zelda Cheng

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Keele University	Toby Helliwell, Adrian Chudyk
Kings College London	Rafaela Giemza, John Lord Villajin, Noah Yogo, Esther Makanju, Pearl Dulawan, Deepak Nagra, April Buazon, Alice Russell, Georgie Bird
Lakeside Healthcare Research, Lakeside Surgery	Amardeep Heer, Rex Sarmiento, Balraj Sanghera, Melanie Mullin, Adam Champion, Aisling Bevan, Kinzah Iqbal, Alshia Johnson
Layton Medical Centre	Rebecca Clark, Sarah Shaw, Steven Shaw, Amanda Chalk, Martin Lovatt, Caroline Lillicrap, Angela Parker, Jan Hansel, Zhi Wong, Galvin Gan, Eyad Tuma
Leeds Teaching Hospitals NHS Trust	Jane Minton, Jennifer Murira, Razan Saman, Alistair Hall, Kyra Holliday, Zara Khan, James Calderwood, George Twigg, Helena Baker, Julie Corrigan, Katy Houseman
Midlands Partnership NHS Foundation Trust	Subhra Raguvanshi, Dominic Heining, Jake Weddell, Liz Glaves, Kim Thompson, Francis Davies, Ruth Lambley Burke
MRC–University of Glasgow Centre for Virus Research, and Queen Elizabeth University Hospital, NHS Greater Glasgow & Clyde, Glasgow, Scotland	Emma C. Thomson
National Institute for Health Research Patient Recruitment Centre and Bradford Teaching Hospitals NHS Foundation Trust, Bradford	Dinesh Saralaya
Newcastle University	Adam Farrier
NIHR Clinical Research Facility, University Hospital Southampton NHS Foundation Trust	Lisa Berry
NIHR Clinical Research Network, Thames Valley and South Midlands, Oxford University Hospitals NHS Foundation Trust	Nancy Hopewell, Leigh Gerdes
NIHR Southampton Clinical Research Facility and NIHR Wessex Local Clinical Research Network, University Hospital Southampton NHS Foundation Trust	Mihaela Pacurar, Saul N Faust
Norfolk and Norwich University Hospital NHS Foundation Trust	Jeremy Turner, Christopher Jeanes, Adele Cooper, Jocelyn Keshet-Price, Lou Coke, Melissa Cambell-Kelly, Ketan Dhatariya, Claire Williams, Georgina Marks, James Sudbury, Lisa Rodolico
Northern Ireland Clinical Research Facility, Queen's University Belfast and Belfast Health and Social Care Trust	Sharon Carr, Roisin Martin, Angelina Madden
Northern Ireland Clinical Research Network	Maurice O'Kane, Paul Biagioni, Sonia McKenna, Alison Clinton
North Tees and Hartlepool NHS Foundation Trust	Justin Carter, Matthew Dewhurst, Bill Wetherill, Rachel Taylor
Oxford Health NHS Foundation Trust, Warneford Hospital	Thandiwe Hoggarth, Katrina Lennon Collins, Marie Chowdhury, Adil Nathoo, Anna Heinen, Jayne E. Starrett, Orla MacDonald, Tokoza Muimo, Claudia Hurducas, Liliana Cifuentes, Sarah McCartney
Quadram Institute	Jane Ewing
Queen Elizabeth University Hospital, NHS Greater Glasgow and Clyde	Guy Mollett, Rachel Blacow, John Haughney, Jonathan MacDonald, John Paul Seenan, Stewart Webb, Colin O'Leary, Scott Muir, Beth White, Neil Ritchie
Queen's University Belfast and Belfast Health and Social Care Trust	Judy Bradley, Daniel F. McAuley, Jonathan Stewart

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Research and Development, NHS Grampian	Chiara Scicluna, Mariella D'Alessandro, Carole Edwards, Lynn MacKay, Amy Nicol, Karen Norris, Heather Lawrence, Sandra Mann, Ruth Valentine
Royal Bournemouth Hospital, University Hospitals Dorset NHS Foundation Trust	Nicki Lakeman, Laura Purandare
Royal Cornwall Hospital NHS Trust	Duncan Browne, David Tucker, Peter Luck, Angharad Everden, Lisa Trembath, Michael Visick, Nick Morley, Laura Reid, Helen Chenoweth, Kirsty Maclean
Royal Devon University Healthcare	Ray P. Sheridan, Tom Burden, Craig Francis Lunt, Shirley Todd, Stephanie Estcourt, Jasmine Marie Pearce, Suzanne Wilkins, Cathryn Love-Rouse
Royal Free London NHS Foundation Trust	Eva Torok-Pollok, Mike Youle, Sara Madge, Natalie Hills, Nikesh Devani, Aarti Nandani, Janet North, Nargis Hemat, Suluma Mohamed
Royal Oldham Hospital, Northern Care Alliance, Greater Manchester	Rachel Newport
Salford Royal Hospital, Northern Care Alliance, Greater Manchester	Philip A. Kalra, Chukwuma Chukwu, Olivia Wickens, Vikki O'Loughlin, Hema Mistry, Louise Harrison, Robert Oliver, Anne-Marie Peers, Jess Zadik, Katie Doyle
South Tees Hospitals NHS Foundation Trust	David R. Chadwick, Kerry Colling, Caroline Wroe, Marie Branch, Alison Chilvers, Sarah Essex, Vicky Hanlon, Helen Dunn, Steven Liggett, Jane Greenaway, Tarn Nozedar
Stafford Town Primary Care Network	Mark Stone
Vaccine Institute, St George's University of London & St George's University Hospitals NHS Foundation Trust	Alberto San Francisco Ramos, Emily Beales, Olivia Bird, Zsofia Danos, Hazel Fofie, Cecilia Hultin, Sabina Ikram, Fran Mabesa, Aoife Mescall, Josyanne Pereira, Jennifer Pearce, Natalina Sutton
St Helens and Knowsley Teaching Hospitals NHS Trust	Emma Snashall
Stockport NHS Foundation Trust, Stepping Hill Hospital	David Neil Baxter, Sara Bennett, Debbie Suggitt, Kerry Hughes, Wiesia Woodyatt, Lynsey Beacon, Alissa Kent, Chris Cooper, Milan Rudic, Simon Tunstall, Matthew Jackson
Swanage Medical Practice	Claire Hombersley
The Adam Practice	Patrick Moore, Rebecca Cutts
University College London	Danielle Solomon, Janet M. North
University Hospitals of Morecambe Bay NHS Foundation Trust	Andrew Higham, Marwan Bukhari, Mohamed Elnaggar, Michelle Glover, Fiona Richardson, Alexandra Dent, Shahzeb Mirza, Rajiv Ark, Jennie Han
University of Exeter Medical School, William Wright House, Royal Devon and Exeter Hospital	Suzy V. Hope, Philip J. Mitchelmore
University of Liverpool	Rostam Osanlou, Thomas Heseltine

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Tables legends

Table 1 Qualifying symptoms of suspected COVID-19.

Table 2 The PCR and symptomatic status of all study participants; 3320 (21.9%) of all participants had at least one symptomatic episode and 317 (2.1%) of all had a PCR+ episode.

Table 3 Cohort demographics stratified by participant PCR status. The ORs measure univariate associations between the PCR status and population characteristics, irrespective of the presence of symptoms.

Table 4 Number (proportions) of participants with specific symptoms, overall and conditioned on the presence/absence of a PCR confirmed episode.

Table 5 The fold-effects of demographics and their 95%CIs on the mean number of days of specific symptoms reported during a symptomatic episode. The estimation uses a Poisson zero inflated model on the number of reports of an episode and allows for multiple episodes with events associated with one participant. This analysis accounts for the length of the event-episode.

Table 6 The fold-effects of demographics and their 95%CIs on the mean number of days of specific symptoms reported during a symptomatic episode restricted to the PCR+ participants. The estimates are the result of fitting a zero-inflated Poisson model on the number of reports within an episode whilst allowing for multiple episodes with events associated with one participant. This analyses also account for the length of the event-episode.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Table 7 The optimal model for PCR+ based on symptoms and population characteristics on a two-level weighted logistic regression analysis. The adjusted effects of three specific reports are shown.

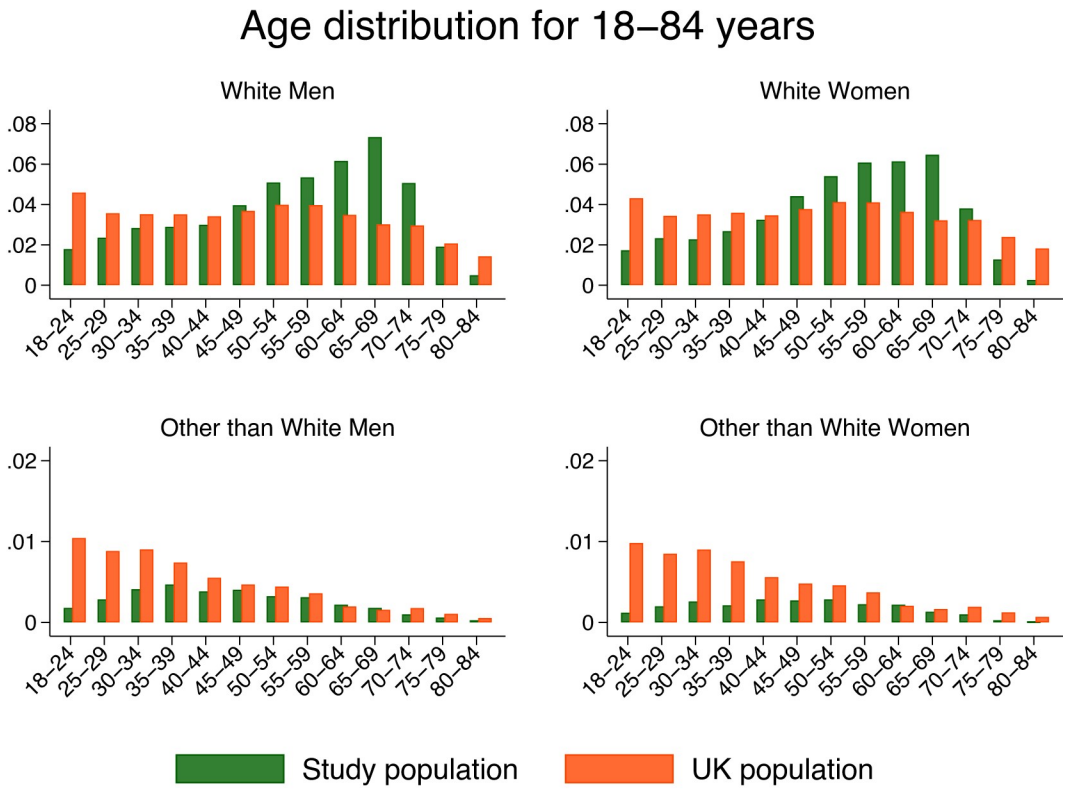
Table 8 Examples of various combination of potential bundles of symptoms and their corresponding probabilities of testing positive as predicted by the optimal model above (age is held at 50 years and the ethnicity is assumed White). That is, a White participant of 50 years of age reporting 1 day of anosomia/ageusia, 3 days of loss of appetite and 3 days of fever and one day of nose congestion and 3 days of cough and 1 day of runny nose and 1 day of chills had 83% chance to test positive (column in bold).

Table 9 The effect of age and ethnicity on the ROC curve and subsequently on discrimination power associated with each classifier in the model. The coefficients are only qualitatively interpreted.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

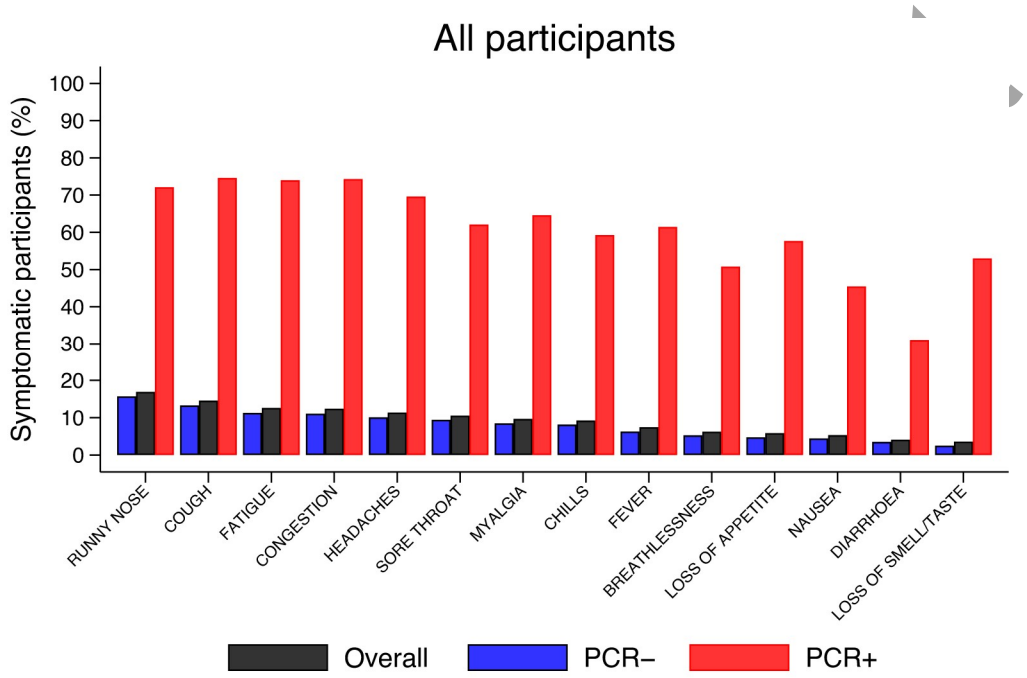
Figures Legends

Figure 1. Age distribution in the study sample compared to that of the UK population, stratified by gender and ethnicity.



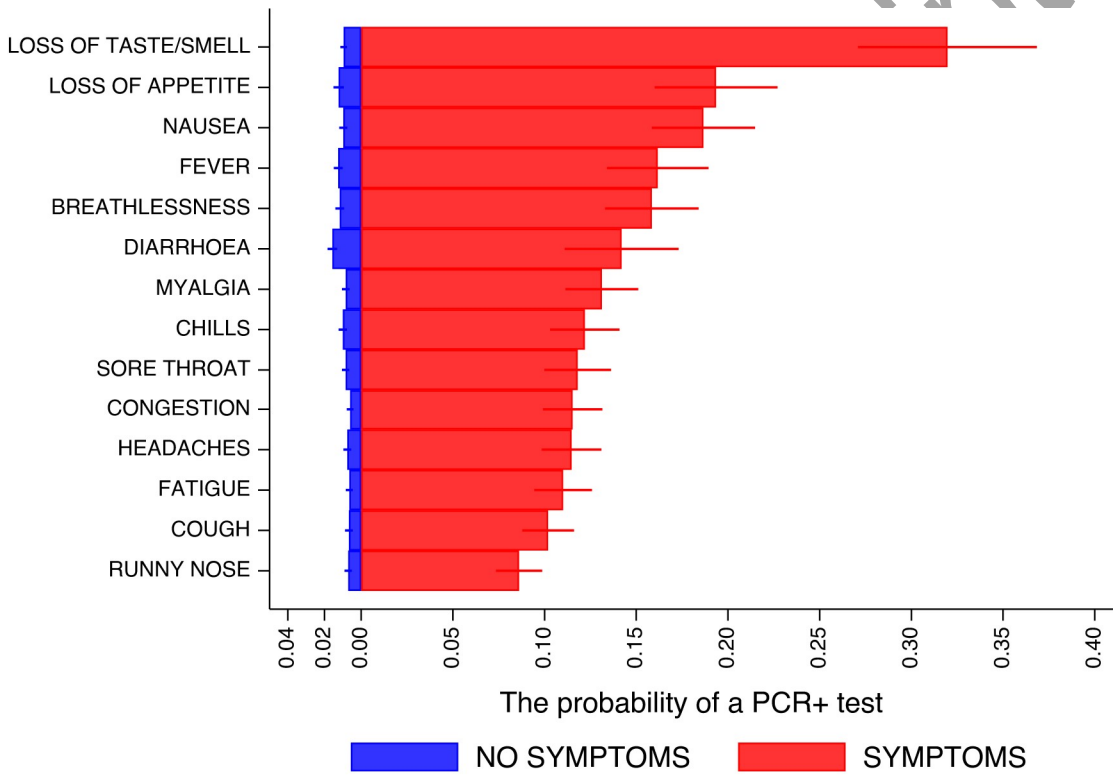
https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure 2. Proportions of participants with specific symptoms, overall and stratified by PCR status, illustrating Table 3 above. For example: overall, 16.9% of all participants reported runny nose at least once but the figure is much higher (72.6%) among PCR+ contrasting with 15.7% among PCR-.



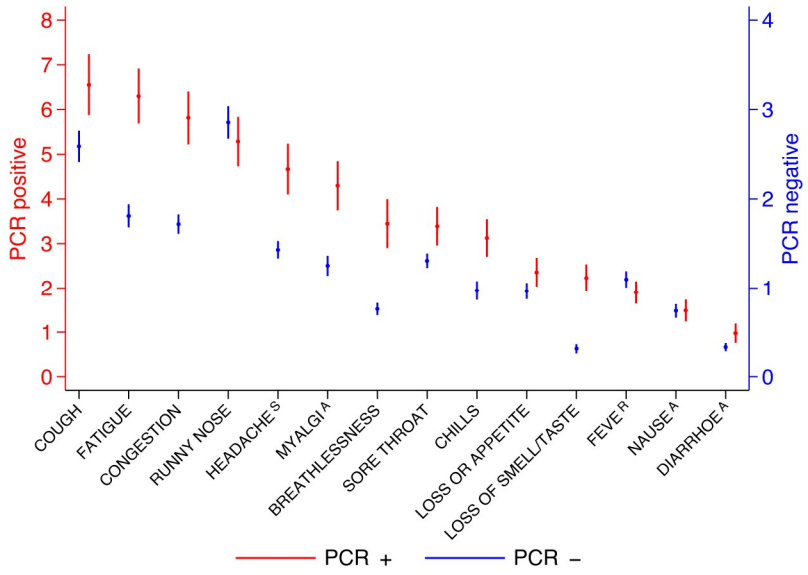
https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure 3. The predicted probabilities of PCR+ status, stratified by the presence of specific symptoms, and their 95%CIs. Predictions related to each specific symptom are unadjusted for the others and are based on a binary regression with robust standard errors accounting for multiple episodes with events associated with a participant. For example, in participants with loss of taste or smell, regardless of the presence or absence of other symptoms, the probability of a positive PCR test is 0.319 or 31.9%.



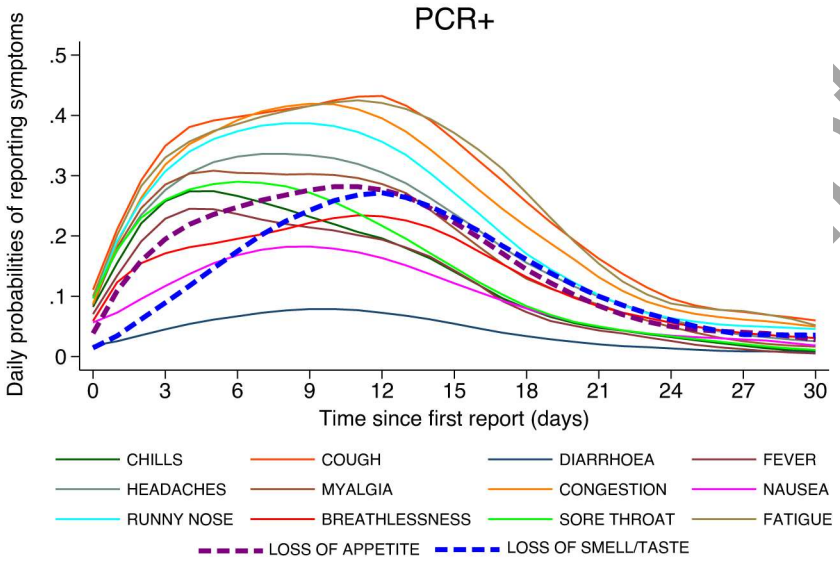
https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure 4. The predicted mean of number of days specific symptoms were reported during an episode and their 95%CIs. The red values (PCR+) are referred to the left axis and the blue values (PCR-) are referred to the right axis. The analysis is restricted to symptomatic participants only. For example, for those participants reporting cough as part of an episode, the mean of the number of days was 6-7 days in PCR+ participants and 2-3 days in PCR-.



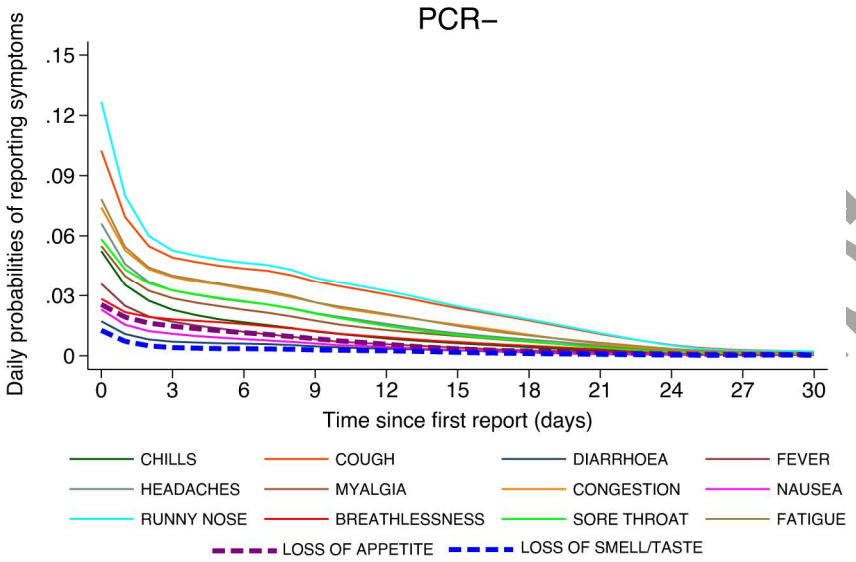
https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure 5 . The daily probabilities of reporting specific symptoms starting with the first report conditioned on PCR+ participants and their corresponding illness episode, i.e. ignoring the symptomatic episodes associated with these participants which were negative. Non-parametric methodology was used to capture the shape of the time series reports.



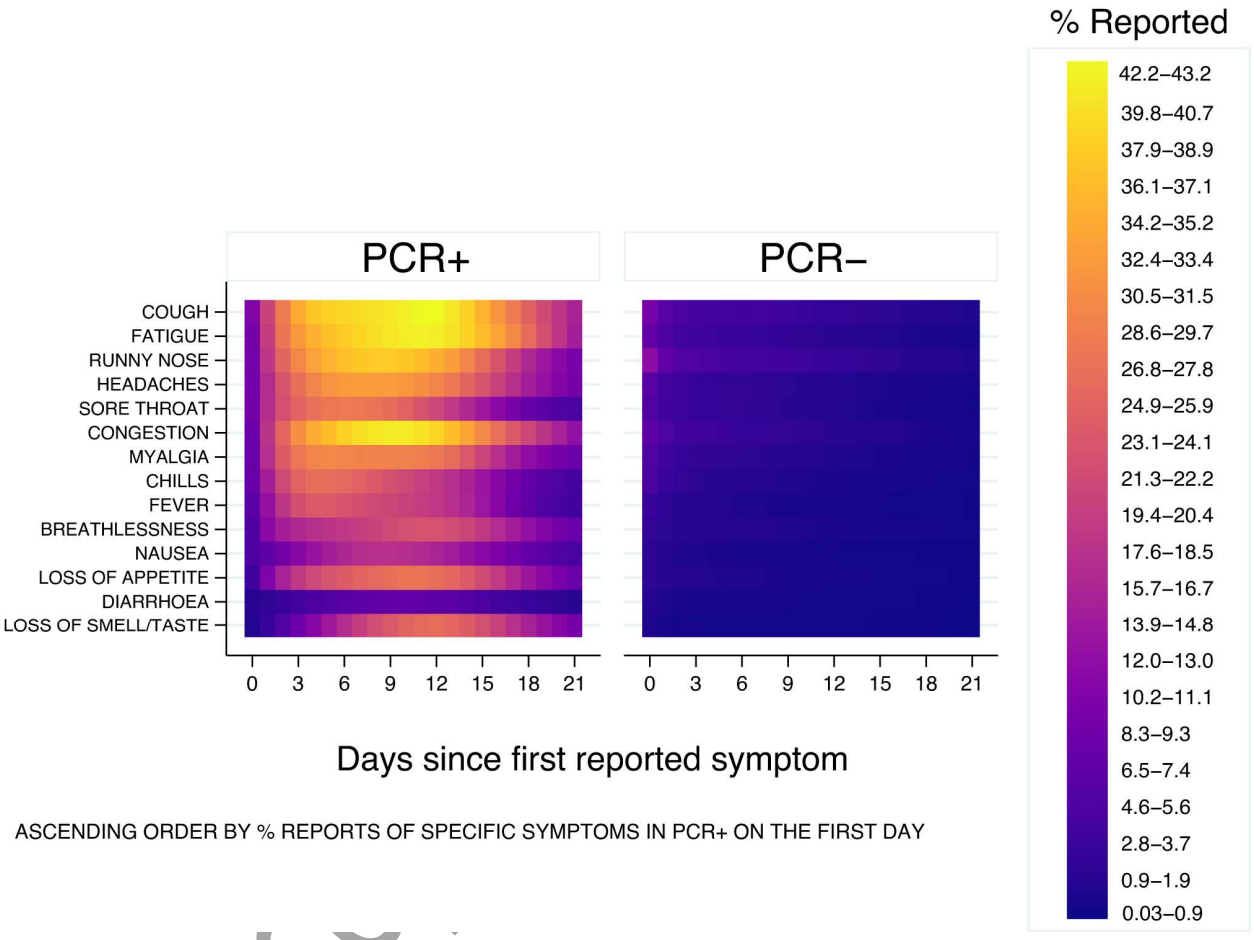
https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure 6. The daily probabilities of reporting specific symptoms starting with the first report in PCR-participants.



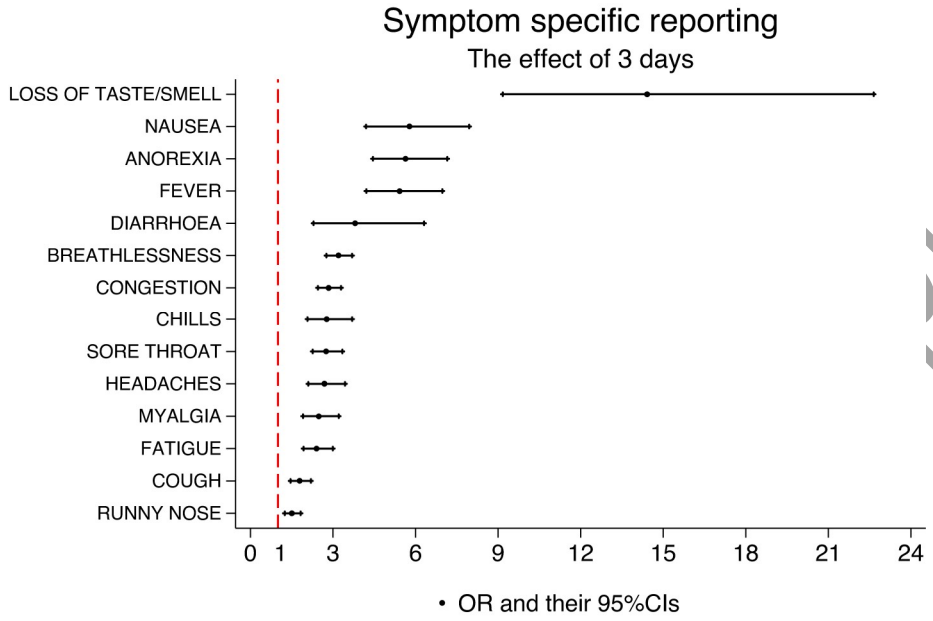
https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure 7. The probabilities of daily occurrences of various symptoms have similar magnitude in both PCR+ and PCR- groups on the first reporting day whilst they peak up later during illness evolution in PCR+ patients and decline in those PCR-, also reflected in the previous figures.



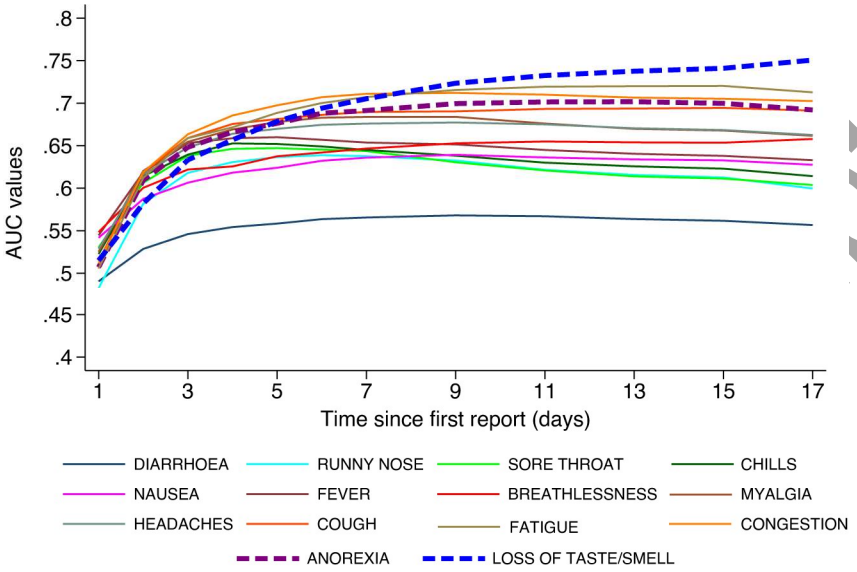
https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure 8. The effect (OR) of reporting a specific symptom for 3 days during an episode, irrespective of other symptoms reported during that episode.



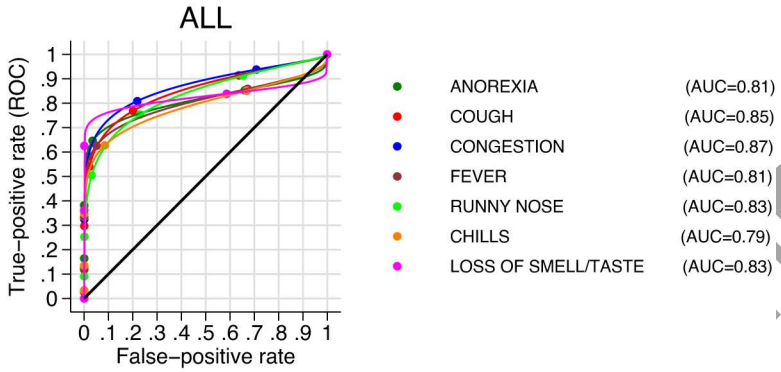
https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure 9. The discrimination power of individual symptoms based on the temporally ordered reports restricted to the first 1, 2, 3 to longer than 15 days after symptomatic illness episode starts.



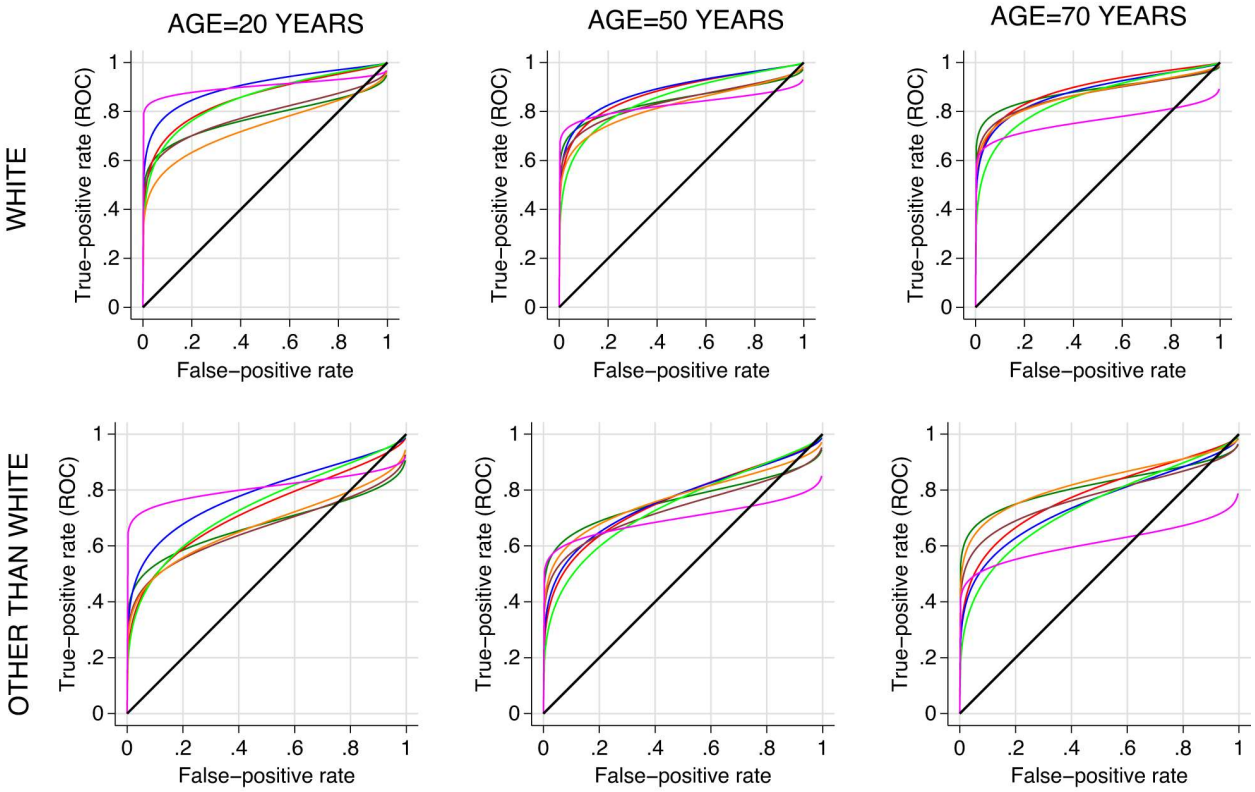
https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure 10. The estimated discrimination power of each classifier. The plot and the AUC estimates follow a maximum likelihood ROC weighted regression analysis uncontrolled for age and ethnicity.



https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

Figure 11. The effect of age and ethnicity on the ROC curve and subsequently on discrimination power associated with each classifier in the model. The colours indicating specific symptom are similar to those displayed in Figure 10.



https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

References

1. World Health Organisation. WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data. Available at: https://covid19.who.int/. Accessed 07 March 2023.
2. Emanuel, E. J. *et al.* Fair Allocation of Scarce Medical Resources in the Time of Covid-19. *The New England Journal of Medicine.* 2020;382:2049-2055.
3. Rossman, H. *et al.* A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. *Nature Medicine.* 2020;26(5):634–638.
4. National Health Service. Main symptoms of coronavirus (COVID-19) - NHS. Available at: https://www.nhs.uk/conditions/coronavirus-covid-19/symptoms/main-symptoms/. Accessed 30 March 2022.
5. Johnson-León, M. *et al.* Executive summary: It's wrong not to test: The case for universal, frequent rapid COVID-19 testing. *eClinicalMedicine*. 2021;33:100759.
6. Lara, B. A. *et al.* Clinical prediction tool to assess the likelihood of a positive sars-cov-2 (covid-19) polymerase chain reaction test in patients with flu-like symptoms. *Western Journal of Emergency Medicine*. 2021;22(3):592–598.
7. Duque, M. P. *et al.* COVID-19 symptoms: A case-control study, Portugal, March-April 2020. *Epidemiology & Infection.* 2021;149:e54.
8. French, N. *et al.* Creating symptom-based criteria for diagnostic testing: a case study based on a multivariate analysis of data collected during the first wave of the COVID-19 pandemic in New Zealand. *BioMed Central Infectious Diseases.* 2021;21(1):1119.
9. Struyf, T. *et al.* Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. *Cochrane Database of Systematic Reviews*. 2021;2(2):CD013665.
10. Struyf, T. *et al.* Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. *Cochrane Database of Systematic Reviews*. 2022;5(5):CD013665.
11. Wojtusiak, J. *et al.* The Role of Symptom Clusters in Triage of COVID-19 Patients. *Quality Management in Health Care.* 2023;32(Supp 1):S21–S28.
12. Wojtusiak, J *et al.* Order of Occurrence of COVID-19 Symptoms. *Quality Management in Health Care.* 2023; 32(Supp 1):S29–S34.
13. Bowyer, R. C. E. *et al.* Characterising patterns of COVID-19 and long COVID symptoms: evidence from nine UK longitudinal studies. *European Journal of Epidemiology.* 2023;38(2):199–210.
14. Drew, D. A. *et al.* Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science.* 2020;368(6497):1362-1367.
15. Canas, L. S. *et al.* Early detection of COVID-19 in the UK using self-reported symptoms: a large-scale, prospective, epidemiological surveillance study. *The Lancet Digital Health.* 2-21;3(9):e587–e598.
16. Menni, C. *et al.* Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine*. 2020;26(7):1037-1040.
17. Elliott, J. *et al.* Predictive symptoms for COVID-19 in the community: REACT-1 study of over 1 million people. *Public Library of Science Medicine.* 2021;18(9):e1003777.
18. Lauer, S. A. *et al.* The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine.* 2020;172(9):577–582.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

19. Heath, P. T. *et al.* Safety and Efficacy of NVX-CoV2373 Covid-19 Vaccine. *New England Journal of Medicine.* 2021;385(13):1172–1183.
20. Office for National Statistics. Population estimates for the UK, England and Wales, Scotland and Northern Ireland - Office for National Statistics. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2020#age-structure-of-the-uk-population. Accessed 27 January 2023.
21. Office for National Statistics. Population estimates by ethnic group, England and Wales - Office for National Statistics. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/datasets/populationestimatesbyethnicgroupenglandandwales. Accessed 27 January 2023.
22. Office for National Statistics. Estimates of the population for the UK, England, Wales, Scotland and Northern Ireland - Office for National Statistics. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland. Accessed 27 January 2023.
23. Little, R. J. A. Post-Stratification: A Modeler's Perspective. *Journal of the American Statistics Association.* 1993;88(423):1001–1012.
24. Alonzo, T. A *et al*. Distribution-free ROC analysis using binary regression techniques. *Biostatistics.* 2002;3(3):421–432.
25. Murali, M. *et al.* Ethnic minority representation in UK COVID-19 trials: systematic review and meta-analysis. *BioMed Central Medicine.* 2023;21:111.
26. Office for National Statistics. Coronavirus (COVID-19) Infection Survey, UK - Office for National Statistics. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/19february2021. Accessed 15 January 2023.
27. Weinbergerova, B. *et al.* COVID-19's natural course among ambulatory monitored outpatients. *Scientific Reports.* 2021;11(1);10124.
28. Rodriguez-Palacios, A. *et al.* Modeling the Onset of Symptoms of COVID-19. *Frontiers in Public Health.* 2020;8:473.
29. Boyce, J. M *et al.* Effects of ageing on smell and taste. *Postgraduate Medical Journal.* 2006;82(966):239-241.
30. Generation Scotland. Access our resources | The University of Edinburgh. Available at: https://www.ed.ac.uk/generation-scotland/for-researchers/access. Accessed 28 October 2023.
31. Public Health England. SARS-CoV-2 variants of concern and variants under investigation in England. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/975742/Variants_of_Concern_VOC_Technical_Briefing_8_England.pdf . Accessed 08 May 2023.
32. Willett, B. J. *et al.* SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. *Nature Microbiology*. 2022;7(8):1161–1179.
33. Menni, C. *et al.* Articles Symptom prevalence, duration, and risk of hospital admission in individuals infected with SARS-CoV-2 during periods of omicron and delta variant dominance: a prospective observational study from the ZOE COVID Study. *The Lancet.* 2022;399(10335):1618-1624.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

- Fever (referred to as FEVER)
- New onset cough (referred to as COUGH)
- New onset or worsening of shortness of breath or difficulty breathing compared to recruitment time (referred to as BREATHLESSNESS)
- New onset fatigue (referred to as FATIGUE)
- New onset generalised muscle or body aches (referred to as MYALGIA)
- New onset headache (referred to as HEADACHES)
- New loss of taste or smell (referred to as LOSS OF TASTE/SMELL)
- New loss of appetite (referred to as ANOREXIA)
- Acute onset of sore throat (referred to as SORE THROAT)
- Acute onset congestion (referred to as CONGESTION)
- Acute onset runny nose (referred to as RUNNY NOSE)
- New onset of chills (referred to as CHILLS)
- New onset of nausea (referred to as NAUSEA)
- New onset of diarrhoea (referred to as DIARRHOEA)

Table 1. Qualifying Symptoms of Suspected COVID-19.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

	Overall		
	PCR-	**PCR+**	**Total**
No symptomatic episode	11768	51	11819
At least one symptomatic episode	3054	266	3320 (21.9%)
	14822	317 (2.1%)	15139

Table 2. The PCR and symptomatic status of all study participants; 3320 (21.9%) of all participants had at least one symptomatic episode and 317 (2.1%) of all had a PCR+ episode.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

VARIABLE	Summary/ Category	ALL	PCR+	PCR-	OR	p-value	95%CI - low	95%CI - high
		15139	**317**	**14822**				
AGE (years)	Mean/SD Median (IQR) Min-Max	53.1/14.9 55(42, 65) 18-84	49.2/13.6 51(38, 60) 18-79	53.2/14.9 55(43, 65) 18-84	0.983	**<0.001**	0.975	0.991
GENDER	Male Female	7,808(51.6%) 7,331(48.4%)	152(48.0%) 165(52.1%)	7,656(51.6%) 7,166(48.4%)	1.086	0.550	.829	1.423
ETHNICITY	White BAME Missing	14280 (94.3%) 675(4.5%) 184 (1.2%)	288(90.9%) 26(8.2%) 3(0.95%)	13992(94.4%) 649(4.4%) 181(1.2%)	1.924	**0.010**	1.169	3.167
BMI	Mean/SD Median (IQR) Min-Max Missing	27.6/5.3 26.7(23.9-30.4) 15.1-55 412(2.7)	28.2/5.6 27.1(24.1-31.6) 16.8-53 7(2.5)	27.6/5.3 26.7(23.9-30.4) 15.1-55 405(2.7)	1.003	0.845	.976	1.030
BMI>30	No Yes Missing	10777(71.2%) 3950(26.1%) 412(2.7%)	216(68.1%) 94(29.7%) 7(2.2%)	10561(71.3%) 3856(26.0%) 405(2.7%)	1.002	0.991	.759	1.321
Presence of comorbidities	No Yes	8372 (55.3%) 6767 (44.7%)	177(55.8%) 140 (44.2%)	8195(55.3%) 6627(44.7%)	0.816	0.128	.628	1.060

Table 3. Cohort demographics stratified by participant PCR status. The ORs measure univariate associations between the PCR status and population characteristics, irrespective of the presence of symptoms.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

	ALL		PCR+		PCR-	
SYMPTOMS	**15139**		**317**		**14822**	
	Number	Proportion	Number	Proportion	Number	Proportion
RUNNY NOSE	2559	16.9%	230	72.6%	2329	15.7%
COUGH	2205	14.6%	238	75.1%	1967	13.3%
FATIGUE	1908	12.6%	236	74.4%	1672	11.3%
CONGESTION	1878	12.4%	237	74.8%	1641	11.1%
HEADACHES	1718	11.3%	222	70.0%	1496	10.1%
SORE THROAT	1595	10.5%	198	62.5%	1397	9.4%
MYALGIA	1463	9.7%	206	65.0%	1257	8.5%
CHILLS	1398	9.2%	189	59.6%	1209	8.2%
FEVER	1128	7.5%	196	61.8%	932	6.3%
BREATHLESSNESS	945	6.2%	162	51.1%	783	5.3%
ANOREXIA	887	5.9%	184	58.0%	703	4.7%
NAUSEA	806	5.3%	145	45.7%	661	4.5%
DIARRHOEA	620	4.1%	99	31.2%	521	3.5%
LOSS OF SMELL/TASTE	541	3.6%	169	53.3%	372	2.5%

Table 4. Number (proportions) of participants with specific symptoms, overall and conditioned on the presence/absence of a PCR confirmed episode.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

	AGE				GENDER				ETHNICITY				BMI				COMORBIDITES			
	RR	p-value	95%CI Low-High		RR	p-value	95%CI Low-High		RR	p-value	95%CI Low-High		RR	p-value	95%CI Low-High		RR	p-value	95%CI Low-High	
RUNNY NOSE	1.005	**<0.001**	1.003	1.008	0.960	0.301	0.888	1.037	0.758	**0.001**	0.645	0.891	0.999	0.765	0.992	1.006	1.076	0.060	0.997	1.162
COUGH	1.006	**<0.001**	1.003	1.009	0.966	0.477	0.879	1.062	0.769	**0.014**	0.624	0.949	1.005	0.140	0.998	1.013	1.185	**<0.001**	1.081	1.298
FATIGUE	0.998	0.307	0.994	1.002	1.040	0.447	0.940	1.151	0.885	0.210	0.731	1.071	1.004	0.315	0.996	1.012	1.081	0.118	0.980	1.192
CONGESTION	0.997	0.182	0.994	1.001	1.049	0.370	0.945	1.164	0.773	**0.020**	0.622	0.961	1.005	0.214	0.997	1.014	1.050	0.352	0.948	1.162
HEADACHES	0.998	0.265	0.994	1.002	1.243	**<0.001**	1.114	1.387	0.894	0.386	0.694	1.152	1.005	0.329	0.995	1.014	1.047	0.408	0.939	1.167
SORE THROAT	0.995	**0.021**	0.990	0.999	1.056	0.419	0.925	1.205	0.887	0.488	0.633	1.244	0.998	0.781	0.988	1.009	1.039	0.567	0.912	1.182
MYALGIA	1.005	0.060	1.000	1.010	0.975	0.713	0.853	1.115	0.820	0.101	0.647	1.040	1.012	**0.033**	1.001	1.023	1.161	**0.025**	1.019	1.322
CHILLS	1.001	0.786	0.994	1.008	1.043	0.661	0.865	1.257	0.751	0.080	0.545	1.035	1.003	0.718	0.989	1.017	1.102	0.282	0.923	1.315
FEVER	0.999	0.842	0.994	1.005	1.052	0.573	0.881	1.257	0.852	0.453	0.561	1.295	1.014	0.033	1.001	1.027	1.111	0.227	0.936	1.319
BREATHLESSNESS	1.001	0.740	0.995	1.007	1.043	0.606	0.888	1.226	0.684	0.059	0.461	1.014	1.024	**<0.001**	1.012	1.037	1.224	**0.017**	1.036	1.445
ANOREXIA	1.009	**0.018**	1.001	1.016	1.021	0.822	0.849	1.228	0.720	0.139	0.466	1.113	1.012	0.066	0.999	1.025	1.184	0.070	0.986	1.420
NAUSEA	1.002	0.499	0.995	1.009	1.106	0.375	0.885	1.382	0.759	0.309	0.446	1.291	1.008	0.273	0.994	1.023	1.061	0.588	0.856	1.317
DIARRHOEA	0.997	0.494	0.990	1.005	0.900	0.426	0.696	1.166	1.168	0.475	0.763	1.787	1.008	0.416	0.989	1.026	1.137	0.328	0.879	1.471
LOSS OF SMELL/TASTE	0.989	**0.018**	0.979	0.998	1.239	0.112	0.951	1.614	0.894	0.763	0.433	1.847	0.994	0.656	0.966	1.022	0.802	0.105	0.614	1.047

Table 5. The fold-effects of demographics and their 95%CIs on the mean number of days of specific symptoms reported during a symptomatic episode. The estimation uses a Poisson zero inflated model on the number of reports of an episode and allows for multiple episodes with events associated with one participant. The analyses also account for the length of the event-episode.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

	AGE				GENDER				ETHNICITY				BMI				COMORBIDITES			
	RR	p-value	95%CI Low-High		RR	p-value	95%CI Low-High		RR	p-value	95%CI Low-High		RR	p-value	95%CI Low-High		RR	p-value	95%CI Low-High	
RUNNY NOSE	1.003	0.390	0.996	1.011	1.006	0.950	0.831	1.219	0.608	0.050	0.369	1.001	1.008	0.416	0.989	1.028	0.982	0.851	0.814	1.185
COUGH	1.005	0.134	0.999	1.011	1.070	0.449	0.898	1.275	0.681	0.063	0.455	1.020	1.016	**0.022**	1.002	1.030	1.151	0.092	0.977	1.355
FATIGUE	1.006	0.054	1.000	1.012	1.047	0.593	0.885	1.238	0.816	0.197	0.599	1.112	1.006	0.557	0.985	1.028	1.012	0.886	0.862	1.187
CONGESTION	0.997	0.469	0.990	1.005	1.126	0.203	0.938	1.353	0.570	**0.002**	0.400	0.812	1.014	0.112	0.997	1.030	1.064	0.500	0.889	1.274
HEADACHES	1.001	0.820	0.993	1.009	1.255	**0.033**	1.018	1.546	0.734	0.139	0.488	1.106	1.012	0.259	0.991	1.034	1.006	0.956	0.820	1.233
SORE THROAT	0.998	0.751	0.989	1.008	0.940	0.664	0.711	1.243	0.748	0.341	0.412	1.359	1.012	0.337	0.987	1.038	1.115	0.415	0.858	1.450
MYALGIA	1.010	**0.039**	1.000	1.019	0.992	0.949	0.786	1.254	0.685	0.115	0.427	1.097	1.022	0.076	0.998	1.046	1.167	0.178	0.932	1.462
CHILLS	1.008	0.225	0.995	1.021	0.915	0.554	0.681	1.229	0.695	0.282	0.359	1.347	1.014	0.432	0.980	1.049	0.945	0.706	0.705	1.266
FEVER	1.004	0.389	0.995	1.014	0.955	0.726	0.737	1.237	0.797	0.332	0.503	1.262	1.005	0.686	0.982	1.028	0.939	0.627	0.730	1.209
BREATHLESSNESS	1.009	0.137	0.997	1.022	1.069	0.675	0.783	1.460	0.376	0.071	0.130	1.085	1.031	**0.012**	1.007	1.056	1.449	0.019	1.064	1.973
LOSS OF APPETITE	1.016	**0.012**	1.004	1.029	0.968	0.835	0.717	1.309	0.746	0.403	0.376	1.482	1.011	0.488	0.981	1.042	1.025	0.867	0.767	1.370
NAUSEA	1.008	0.110	0.998	1.018	0.953	0.762	0.699	1.299	0.759	0.416	0.391	1.475	1.009	0.541	0.981	1.037	1.019	0.898	0.762	1.364
DIARRHOEA	1.001	0.951	0.981	1.020	0.828	0.479	0.491	1.396	1.540	0.324	0.653	3.630	0.982	0.621	0.916	1.054	0.990	0.971	0.586	1.674
LOSS OF SMELL/TASTE	0.998	0.747	0.985	1.011	1.093	0.527	0.830	1.440	0.734	0.453	0.328	1.646	1.010	0.521	0.980	1.041	0.975	0.858	0.741	1.284

Table 6. The fold-effects of demographics and their 95%CIs on mean number of days of specific symptoms reported during a symptomatic episode restricted to the PCR+ participants. The estimates are the result of fitting a zero-inflated Poisson model on the number of reports within an episode whilst allowing for multiple episodes with events associated with one participant. The analyses also account for the length of the event-episode.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

The optimal diagnostic model for PCR+ based on symptoms and population characteristics

VARIABLE	OR	p-value	95%CI - L	95%CI - H
LOSS OF TASTE AND SMELL	5.181	0.000	3.400	7.894
LOSS OF APPETITE	2.323	0.000	1.643	3.283
FEVER	1.880	0.000	1.385	2.552
CONGESTION	1.875	0.000	1.464	2.402
COUGH	1.338	0.004	1.098	1.631
RUNNY NOSE	0.662	0.004	0.500	0.877
CHILLS	0.578	0.000	0.443	0.753
AGE	0.988	0.024	0.977	0.998
BAME vs. WHITE	2.434	0.001	1.406	4.214

Table 7. The optimal model based on a two-level weighted logistic regression model. The adjusted effects of three specific reports are shown.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

VARIABLE	NUMBER OF REPORTS																	
LOSS OF TASTE AND SMELL	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
LOSS OF APPETITE	1	1	1	2	2	2	3	3	3	1	1	1	1	1	1	3	3	3
FEVER	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	3	3	3
NOSE CONGESTION	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
COUGH	1	1	1	1	1	1	1	1	1	2	2	2	1	1	1	3	3	3
RUNNY NOSE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CHILLS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PROBABILITY OF PCR+	0.60	0.72	0.82	0.66	0.78	0.86	0.73	0.82	0.89	0.62	0.74	0.83	0.65	0.76	0.85	0.83	0.89	0.94

Table 8. Examples of various combination of potential bundles of symptoms and their corresponding probabilities of testing positive as predicted by the optimal model above (age is held at 50 years and the ethnicity is assumed White).

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press

SYMPTOMS		COEFFICIENT	P-VALUE	95%CI - LOW	95%CI -HIGH
LOSS OF TASTE/SMELL	BAME vs. White	-0.436	**0.041**	-0.853	-0.019
	Age	-0.012	**0.011**	-0.021	-0.003
ANOREXIA	BAME vs. WHITE	-0.312	0.116	-0.701	0.077
	Age	0.009	0.053	0.000	0.018
FEVER	BAME vs. WHITE	-0.390	**0.040**	-0.761	-0.018
	Age	0.007	0.109	-0.002	0.016
CONGESTION	BAME vs. WHITE	-0.556	**0.016**	-1.007	-0.105
	Age	-0.003	0.583	-0.012	0.007
COUGH	BAME vs. WHITE	-0.521	**0.028**	-0.986	-0.055
	Age	0.004	0.408	-0.005	0.014
RUNNY NOSE	BAME vs. WHITE	-0.467	**0.034**	-0.897	-0.036
	Age	0.000	0.998	-0.009	0.009
CHILLS	BAME vs. WHITE	-0.191	0.316	-0.564	0.182
	Age	0.010	**0.023**	0.001	0.019

Table 9. The effect of age and ethnicity on the ROC curve and subsequently on discrimination power associated with each classifier in the model. The coefficients are only qualitatively interpreted.

https://doi.org/10.1017/S0950268824000037 Published online by Cambridge University Press