





Diagnosis of COVID-19 by exhaled breath analysis using gas chromatography–mass spectrometry

Wadah Ibrahim^{1,2,4}, Rebecca L. Cordell^{3,4}, Michael J. Wilde ³, Matthew Richardson^{1,2}, Liesl Carr^{1,2}, Ananga Sundari Devi Dasi^{1,2}, Beverley Hargadon^{1,2}, Robert C. Free^{1,2}, Paul S. Monks ³, Christopher E. Brightling^{1,2}, Neil J. Greening^{1,2} and Salman Siddiqui^{1,2}, on behalf of the EMBER consortium

¹Dept of Respiratory Sciences, University of Leicester, Leicester, UK. ²Institute for Lung Health, NIHR Leicester Biomedical Research Centre (Respiratory theme), Glenfield Hospital, Leicester, UK. ³School of Chemistry, University of Leicester, Leicester, UK. ⁴These authors contributed equally.

Corresponding author: Salman Siddiqui (ss338@le.ac.uk)



Shareable abstract (@ERSpublications)

GC-MS-detected exhaled breath biomarkers can identify PCR-positive #COVID19 patients with modest accuracy <https://bit.ly/3sYqL9H>

Cite this article as: Ibrahim W, Cordell RL, Wilde MJ, *et al.* Diagnosis of COVID-19 by exhaled breath analysis using gas chromatography–mass spectrometry. *ERJ Open Res* 2021; 7: 00139-2021 [DOI: 10.1183/23120541.00139-2021].

Copyright ©The authors 2021

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact permissions@ersnet.org

Received: 2 March 2021
Accepted: 24 April 2021

Abstract

Background The ongoing coronavirus disease 2019 (COVID-19) pandemic has claimed over two and a half million lives worldwide so far. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection is perceived to be seasonally recurrent, and a rapid noninvasive biomarker to accurately diagnose patients early on in their disease course will be necessary to meet the operational demands for COVID-19 control in the coming years.

Objective The aim of this study was to evaluate the role of exhaled breath volatile biomarkers in identifying patients with suspected or confirmed COVID-19 infection, based on their underlying PCR status and clinical probability.

Methods A prospective, real-world, observational study was carried out, recruiting adult patients with suspected or confirmed COVID-19 infection. Breath samples were collected using a standard breath collection bag, modified with appropriate filters to comply with local infection control recommendations, and samples were analysed using gas chromatography–mass spectrometry (TD-GC-MS).

Results 81 patients were recruited between April 29 and July 10, 2020, of whom 52 out of 81 (64%) tested positive for COVID-19 by reverse transcription–polymerase chain reaction (RT-PCR). A regression analysis identified a set of seven exhaled breath features (benzaldehyde, 1-propanol, 3,6-methylundecane, camphene, beta-cubebene, iodobenzene and an unidentified compound) that separated PCR-positive patients with an area under the curve (AUC): 0.836, sensitivity: 68%, specificity: 85%.

Conclusions GC-MS-detected exhaled breath biomarkers were able to identify PCR-positive COVID-19 patients. External replication of these compounds is warranted to validate these results.

Introduction

The ongoing coronavirus disease 2019 (COVID-19) pandemic has claimed over two and a half million lives worldwide so far by March 2021 [1]. Rapid and accurate diagnostic testing of COVID-19 is invaluable to allow rapid isolation, healthcare pathways and access to definitive therapies. Final diagnosis of COVID-19 infection relies heavily on real-time-reverse-transcriptase polymerase chain reaction (rRT-PCR) positivity [2] as well as clinical symptoms, radiological features and blood biomarkers. The false-negative rate for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RT-PCR remains highly variable with rates reaching up to 67% false negative within the first 5 days post exposure [3]. It is strongly recommended that PCR results are carefully interpreted and correlated with pre-test probability, clinical symptoms, blood biomarkers and digital pathology, particularly when used as a basis for removing precautions put in place to prevent onward transmission.



There are several tests that offer increased turnaround at the expense of reduced sensitivity; these include rapid antigen tests and lateral flow assays [4, 5]. The trade off in sensitivity is justified when mass high throughput testing may be required at frequent intervals, *e.g.* at airports, schools and hospitals; however, this has important implications for patients, healthcare professionals and COVID-19 policymakers, highlighting the imperative of developing early, noninvasive, more sensitive diagnostic tools.

Exhaled breath analysis has attracted notable scientific and clinical interest in recent years. Volatile organic compounds (VOCs) have the potential to mirror various metabolic processes both locally within the respiratory system and systemically, *via* blood circulation [6, 7]. VOCs have been utilised as diagnostic, prognostic and treatment response biomarkers for various respiratory illnesses, including infections [8–11]. The rapid, cost-effective and noninvasive nature make VOCs a strong candidate as a potential COVID-19 biomarker, substantiated by preliminary studies highlighting its diagnostic potential [12, 13].

In this pilot study, we evaluated exhaled breath volatile compounds as potential biomarkers for COVID-19 infection, based on the underlying PCR status and clinical probability.

Materials and methods

Study design and participants

This was a prospective, real-world observational study conducted in Leicester, United Kingdom. The National Research Ethics Service Committee East Midlands approved the study protocol (REC number: 16/LO/1747). Patients admitted to hospital with suspected COVID-19 infection were approached and provided informed consent. Those who agreed to take part provided a breath sample within 24 h of admission to hospital, and samples were analysed using thermal desorption coupled gas chromatography–mass spectrometry (TD-GC-MS).

All participants underwent testing for SARS-CoV-2 using PCR of nasopharyngeal swab, as part of the local streaming and clinical care pathways.

Clinical probability

In addition to PCR testing participants were classified by clinical probability of COVID-19 infection. This was with the aim of capturing patients with false-negative PCR swabs or those with delayed presentation who had pneumonitis but were no longer shedding viral load.

We classified the study subjects as high clinical probability of SARS-CoV-2 infection if they fulfilled either two major or one major and three minor diagnostic criteria (table 1).

These criteria were contrived by a panel of senior clinicians, supported by peer-reviewed published evidence [14–20].

Exhaled breath sample collection

Breath sampling was optimised to minimise cross-contamination and infection risk. The breath sampling apparatus included a mouthpiece with integrated HEPA filter (GVS) and a t-piece with directional valves (Intersurgical), connected to a polypropylene shut-off valve with 1/4" push connections (RS Components) *via* a 22F-10 mm connector (Intersurgical). The valve was connected on the other side, *via* a piece of 1/4" PFA tubing of minimal length, to the large port of a 3 L Tedlar bag with dual stainless-steel fittings with a piece of pre-conditioned silicon tubing surrounding the fitting to ensure a secure connection. The tubing and shut-off valves were pre-conditioned by heating for 4 h at 60°C and the Tedlar bags flushed with nitrogen and vacuum evacuated six times prior to sampling to reduce background VOCs. The

TABLE 1 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection diagnostic criteria

Major criteria	Minor criteria
Presence of new bilateral peripheral, middle and lower pulmonary infiltrates on chest radiograph	Low blood lymphocyte count below the lower limit of normal ($<1.0 \times 10^9$ per L)
New continuous cough, high temperature, and loss of sense of smell and or taste	Low blood eosinophil count below the lower limit of normal ($<0.04 \times 10^9$ per L)
	High C-reactive protein level (>50 mg·L ⁻¹)
	New oxygen requirements
	Contact with confirmed COVID-19 in household

pre-assembled, single-use device was handed to the patient at bedside, and patients were asked to perform tidal breathing, opening the shut-off valve when providing the sample and closing it again once the bag was inflated. Average sampling time was less than a minute.

The mouthpiece and all connections above the shut-off valve were disconnected and disposed of on the ward. The remaining bag and valve assembly were wiped down and double bagged before being taken immediately for loading onto sorbent tubes in triplicate (Carbograph 1TD, Markes International Ltd, Llantrisant, UK). The unused fitting on the Tedlar bag was connected to a handheld Elf pump (Escort® Elf, Sigma Aldrich, Dorset, UK) using sealed Tygon tubing, and the sample within the bag drawn through the sorbent tube under negative pressure at a flow rate of $0.5 \text{ L}\cdot\text{min}^{-1}$ for 2 min, loading a total of 1 L of breath onto each of the three sorbent tubes. The sorbent tubes were sealed with brass caps (Swagelok) and stored in the fridge for up to 48 h before being transferred to the lab for analysis. Blank samples were taken from Tedlar bags filled with air (BTCA grade, BOC) at the beginning of each sampling week.

Sample analysis

All sample tubes were loaded with internal standard ($0.6 \mu\text{L}$ of $20 \mu\text{g}\cdot\text{mL}^{-1}$ toluene- d_8 , phenanthrene- d_{10} (Sigma Aldrich) and n-octane- d_{18} (Cambridge Isotope Laboratories) mixture in methanol (Sigma Aldrich)) in a stream of nitrogen at a flow rate of $100 \text{ mL}\cdot\text{min}^{-1}$ for 2 min, purging off the excess solvent.

Reference standards

A $100 \mu\text{g}\cdot\text{mL}^{-1}$ multi component air standard (p/n 4M7537-U, Sigma Aldrich) was used to monitor instrument performance and for peak identity confirmation and was diluted in methanol to give a final concentration of $10 \mu\text{g}\cdot\text{mL}^{-1}$. A $1 \mu\text{L}$ sample of standard was loaded onto sorbent tubes into a stream of N_2 (zero grade, BOC) at $100 \text{ mL}\cdot\text{min}^{-1}$ and purged for 1 min, before addition of internal standard as described above. Benzaldehyde, 1-propanol and octanal (Sigma Aldrich) and SPEX Certiprep Can-Terp Mix (Fisher Scientific, Loughborough, UK) standards were diluted in the same way and loaded onto tubes to confirm the identity of additional selected peaks.

A reference solution used to monitor retention behaviour was prepared from a $40 \text{ mg}\cdot\text{L}^{-1}$ $\text{C}_8\text{--}\text{C}_{20}$ saturated alkanes certified reference material (Sigma Aldrich) combined with a $2000 \mu\text{g}\cdot\text{mL}^{-1}$ aromatics calibration standard (NJDEP EPH 10/08 Rev.2, Thames Restek). The mixture was diluted in methanol (SupraSolv grade, Sigma Aldrich) to give final concentrations of $10 \mu\text{g}\cdot\text{mL}^{-1}$ and $20 \mu\text{g}\cdot\text{mL}^{-1}$ for n-alkanes and aromatics, respectively.

TD-GC-MS analysis

GC-MS analysis was carried out on an Agilent 7820A with 5977B MS (Agilent Technologies Ltd, Stockport, UK). A DB-5MS capillary column $60 \text{ m}\times 0.25 \text{ mm}\times 0.25 \mu\text{m}$ (from Agilent Technologies) was used with a $1 \text{ mL}\cdot\text{min}^{-1}$ column flow rate using helium (N6.0, BOC) as a carrier gas. The GC conditions were as follows: the column starting temperature was 35°C which was then raised to 130°C at $2.8^\circ\text{C}\cdot\text{min}^{-1}$, to 220°C at $4^\circ\text{C}\cdot\text{min}^{-1}$ and then to 320°C at $25^\circ\text{C}\cdot\text{min}^{-1}$, where it was held for a further 10 min.

The mass spectrometer was operated in full scan mode from 40 to 350 amu. The transfer line to the mass spectrometer was heated to 300°C ; the source temperature was maintained at 230°C and the quadrupole at 150°C .

The instrument was interfaced with a Markes TD-100 xr thermal desorption unit (Markes International Ltd). Tubes were pre-purged with carrier gas for 1 min at $50 \text{ mL}\cdot\text{min}^{-1}$ and then desorbed at 300°C for 5 min with a flow of $50 \text{ mL}\cdot\text{min}^{-1}$ onto a “hydrophobic, general purpose” trap (Markes International Ltd) held at -10°C . The trap was then purged for 2 min at $2 \text{ mL}\cdot\text{min}^{-1}$ before being heated at the maximum heating rate to 300°C for 5 min, with a spit flow of $2 \text{ mL}\cdot\text{min}^{-1}$.

Data analysis

A custom library was built in Unknowns (Agilent Technologies) from random selection of 10 patient samples, covering all patient groups, based on their age, PCR status and clinical probability. From this library a method was built in MassHunter Quantitative analysis (Agilent Technologies) using the base peak as quantifier and two qualifier ions for identity confirmation. All peaks were manually screened to ensure correct integration, and siloxane peaks were removed prior to further analysis. Peaks were removed if <60% of samples did not exceed the mean blank integrated peak area plus three standard deviations. Three patient samples were discarded owing to the high levels of exogenous contamination.

Statistical analysis

Feature selection was performed by applying LASSO regression to the final VOCs peak table with the dependent variable as PCR status first, then clinical probability, extracting a set of relevant features for each model.

A logistic regression model was then fitted with the dependent variable as PCR status first, then clinical probability, with the independent variables set as the respective relevant features selected by the LASSO model above. Partial least squares discriminant analysis (PLS-DA) and principal component analysis (PCA) were then applied to the two sets of features, with the dependent variable as SARS-CoV-2 PCR status and clinical probability, respectively.

All analyses were performed using R 4.0.0 [21]. LASSO regression was performed using glmnet [22], PCA was performed using the mixOmics library [23] and PLS-DA was performed using the ropls library [24].

Results

From April 29 to July 10, 2020, a total of 81 adult participants, mean age (\pm SD): 56.5 \pm 15.1, 59% males, with suspected or confirmed COVID-19 infection were recruited. Details of all study participants and their clinical characteristics are detailed in table 2.

Identification of SARS-CoV-2 in breath by PCR status

In total, 52 (64%) of the participants had a positive PCR test at the time of admission (table 2).

A set of seven features were extracted that had non-zero regression coefficients in at least 70 out of 100 runs of 10-fold cross-validation of the LASSO model. Compound identities were confirmed using the Metabolomics Standards Initiative (MSI) [25]. These were: the oxygenates benzaldehyde, 1-propanol (both MSI level 1); a hydrocarbon, 3,6-dimethylundecane (MSI level 2); and two terpenes, camphene and beta-cubebene (MSI level 1 and 2, respectively). Iodobenzene was also extracted, which is likely of

TABLE 2 Clinical characteristics of adult patients recruited with suspected or confirmed coronavirus disease 2019 (COVID-19) infection

	All patients	PCR positive	PCR negative
Subjects n	81	52	29
Baseline demographics			
Age years	56.5 \pm 15.1	53.6 \pm 14.5	61.7 \pm 14.9
Male	48/81 (59)	29/52 (55)	19/29 (65)
Caucasian	42/81 (51)	19/52(36)	23/29 (79)
Current smoker	7/81 (8.6)	4/52 (7)	3/29 (10)
Height m	1.68 \pm 0.08	1.66 \pm 0.08	1.72 \pm 0.07
Weight kg	88.6 \pm 25.8	87.7 \pm 24.9	90.2 \pm 27.7
BMI kg·m ⁻²	31.0 \pm 8.1	31.6 \pm 7.7	30.1 \pm 8.7
Comorbidities			
Diabetes	21/81 (25.9)	14/52 (27)	7/29 (24)
Chronic pulmonary disease	13/81 (16.0)	5/52 (9)	8/29 (27)
Cardiac disease	6/81 (7.4)	5/52 (9)	1/29 (3)
Admission observations			
Temperature >37.8°C	25/81 (30.8)	15/52(28)	10/29 (34)
Respiratory rate >25 min ⁻¹	13/81 (16.0)	9/52 (17)	4/29 (13)
Requiring supportive oxygen	38/81 (46.9)	30/52 (57)	8/29 (27)
Blood biomarkers			
Lymphocyte count $\times 10^9$ per L	0.98 (0.7–1.3)	1.0 (0.7–1.3)	0.8 (0.7–1.2)
Eosinophil count $\times 10^9$ per L	0.02 (0.01–0.08)	0.02 (0.01–0.06)	0.06 (0.02–0.1)
C-reactive protein mg·dL ⁻¹	71.0 (41.5–152.5)	68.5 (47–119)	77.0 (32–228)
Imaging			
Presence of bilateral consolidation	67/81 (82)	47/52 (90)	20/29 (69)
Length of hospital stay days	5.9 \pm 5.4	6.5 \pm 6.3	5.0 \pm 3.7
30-day mortality n (%)	9 (11)	7 (8)	2 (2.4)

Data are presented as mean \pm SD, n/N (%) or median (interquartile range), unless otherwise stated. BMI: body mass index.

exogenous origin, and a compound that was unable to be identified owing to co-elution with a much larger peak.

The distribution of the first discriminant function and the first two principal components are shown in figure 1a–b. Wilcoxon test was applied to the first discriminant function and the first PCA scores (figure 1c–d).

The AUC for the first discriminant function score was 0.83 (95% CI 0.74–0.92), sensitivity was 0.68 (95% CI 0.55–0.80), specificity was 0.85 (95% CI 0.72–0.98), positive predictive value (PPV) was 0.89 (95% CI 0.79–0.99) and negative predictive value (NPV) was 0.60 (95% CI 0.44–0.75) (figure 1e). The AUC for the first PCA was 0.79 (95% CI 0.69–0.90), sensitivity was 0.70 (95% CI 0.57–0.82), specificity was 0.78 (95% CI 0.63–0.93), PPV was 0.85 (95% CI 0.74–0.96) and NPV was 0.59 (95% CI 0.43–0.75) (figure 1f).

Identification of SARS-CoV-2 in breath by clinical probability

Fifty-five out of 81 (68%) subjects were classified as “high clinical probability” for COVID-19 of whom 38 out of 55 (69%) have subsequently had positive PCR results.

A set of 11 features were extracted that had non-zero regression coefficients in at least 10 out of 100 runs of 10-fold cross-validation of the LASSO model. The feature selection threshold of 10 out of 100 runs of 10-fold cross-validation is weak, suggesting that these features are less likely to be stable.

The 11 features were a mixture of hydrocarbons – cyclohexene (identified to MSI level 2), 3-heptene (MSI level 2), pentadecane (MSI level 1) and 4-ethenyl-1,2-dimethyl-benzene (MSI level 2); oxygenates – octanal (MSI level 1), benzaldehyde (MSI level 1), 2,2-dimethyl 1-propanol (MSI level 2), 1-propanol (MSI level 1), acetoin (MSI level 2) and acetic acid methyl ester (MSI level 2); and the chloro-carbon tetrachloroethylene (MSI level 2).

The distribution of the first discriminant function and the first two principal components are shown in figure 2a–b. Wilcoxon test was applied to the first discriminant function and the first PCA scores (figure 2c–d).

The AUC for the first discriminant function score was 0.65 (95% CI 0.52–0.78), PPV was 0.82 (95% CI 0.70–0.95) and NPV was 0.41 (95% CI 0.27–0.56) (figure 2e). The AUC for the first PCA was 0.55 (95% CI 0.40–0.70), PPV was 0.81 (95% CI 0.69–0.92) and NPV was 0.44 (95% CI 0.28–0.60) (figure 2f).

Discussion

In this pilot study we evaluated the validity of using exhaled breath VOCs in identifying patients with COVID-19 infection based on their underlying PCR status and clinical probability, using GC-MS. This study provided proof of concept for exhaled breath measurement in patients with severe COVID-19 infection, while maintaining infection control standard precautions. Furthermore, it demonstrated that VOC biomarker profiling can identify COVID-19 patients based on their underlying PCR status with good accuracy.

This study is the first to incorporate GC-MS, the gold standard in breath/VOC discovery analysis, recruiting the largest number of COVID-19 patients to date using an advanced offline technology. The high analytical standards applied in our study and the increased peak capacity resulted in the detection of 389 features that exceed background levels by 3 standard deviations in 60% of samples.

With the current level of understanding of COVID-19 pathophysiology, there are several indicators that SARS-CoV-2 infection would be detectable in the exhaled breath VOC pattern. The disease has been reported to cause a systemic inflammatory response [26], which supports the hypothesis that metabolism is influenced in more than one way and the distribution of breath volatile metabolites is substantially changed [27].

The discriminatory features identified in our study belonged to various chemical groups and, with the exception of the unidentified peak, were present in higher concentrations in the breath of PCR-positive patients. These included 1-propanol, a previously observed marker of bacterial pneumonia [28], lung cancer [29, 30] and asthma [10]; and 3,6-dimethylundecane, a likely derivative of lipid peroxidation and potentially background air [31]. The concentrations in breath of numerous alkanes have been reported to have altered levels in a range of respiratory diseases including asthma [10, 32] and COPD [33].

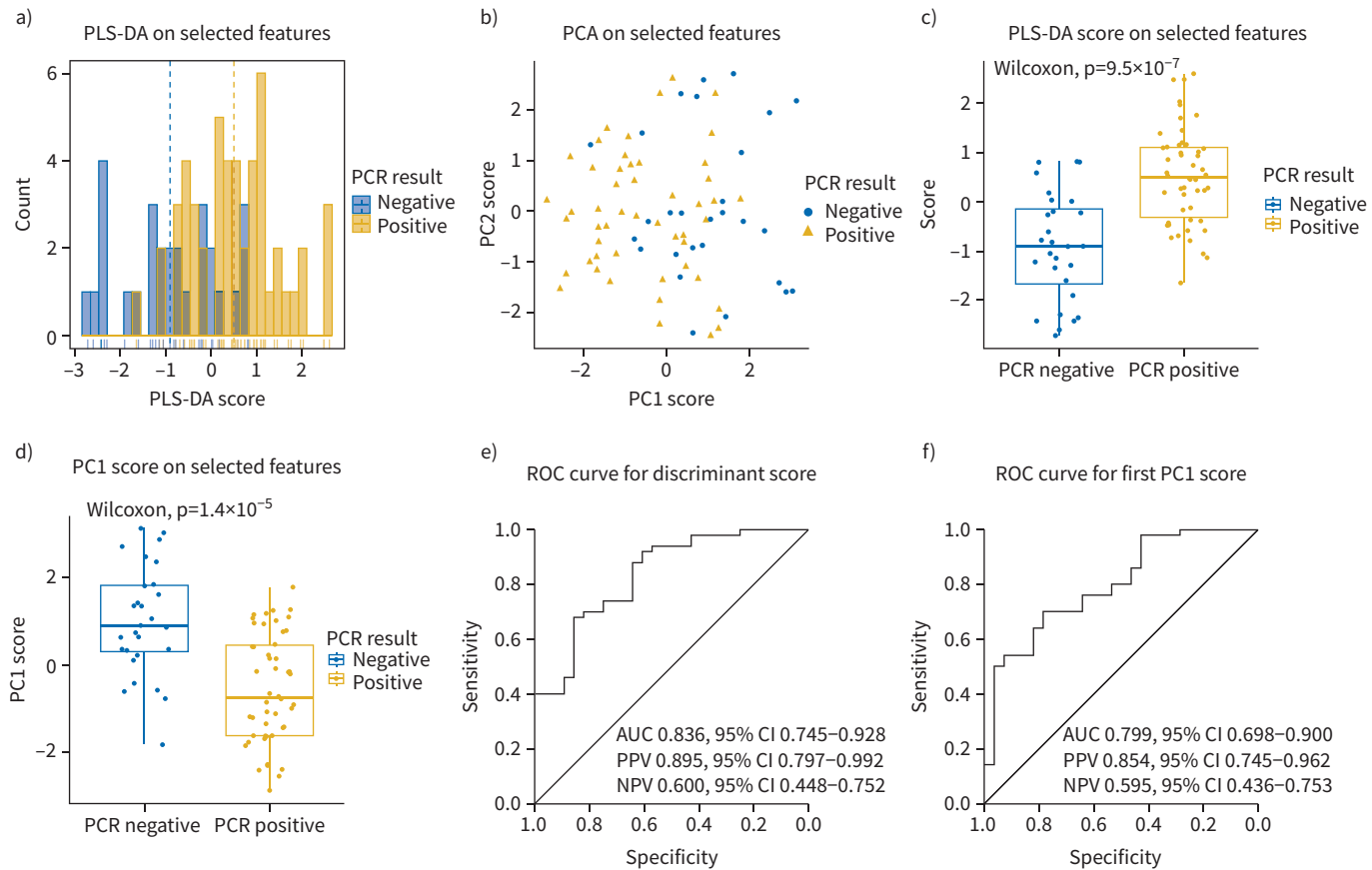


FIGURE 1 a) Partial least squares discriminant analysis (PLS-DA) on the seven extracted breath features with the response variable set as PCR status. b) Principal component analysis (PCA) on the seven extracted breath features with the response variable set as PCR status. c) Wilcoxon test applied to the first discriminant function. d) Wilcoxon test on the first PCA scores. e) Receiver operating characteristic (ROC) curve for the discriminant score. The area under the curve (AUC) for the first discriminant function score: 0.83 (95% CI 0.74–0.92), sensitivity: 0.68 (95% CI 0.55–0.80), specificity: 0.85 (95% CI 0.72–0.98), positive predictive value (PPV): 0.89 (95% CI 0.79–0.99) and negative predictive value (NPV): 0.60 (95% CI 0.44–0.75). f) ROC curves for first principal component score (PC1). The AUC for the first PCA: 0.79 (95% CI 0.69–0.90), sensitivity: 0.70 (95% CI 0.57–0.82), specificity: 0.78 (95% CI 0.63–0.93), PPV: 0.85 (95% CI 0.74–0.96) and NPV: 0.59 (95% CI 0.43–0.75).

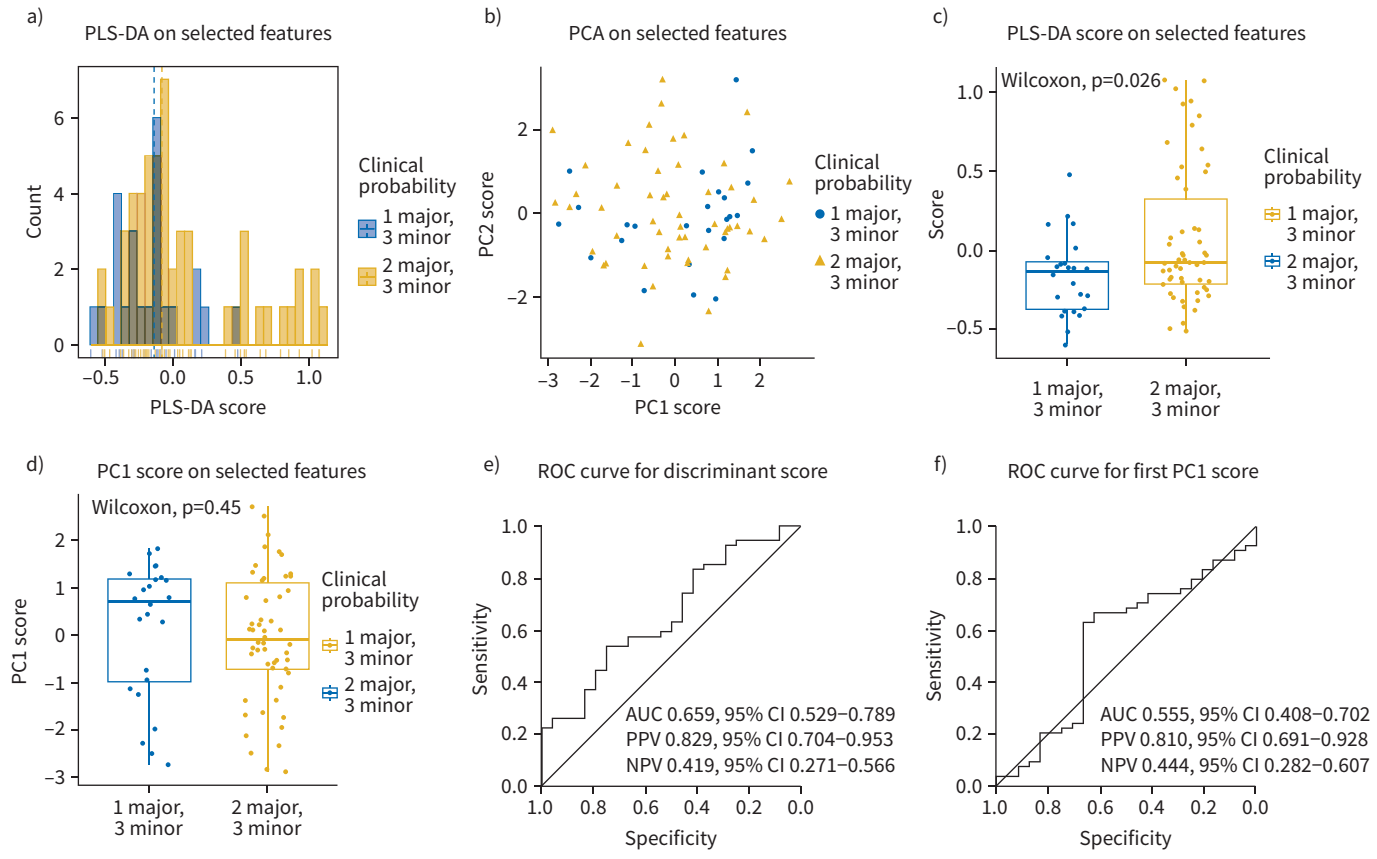


FIGURE 2 a) Partial least squares discriminant analysis (PLS-DA) on the 11 extracted breath features with the response variable set as clinical probability. b) Principal component analysis (PCA) on the 11 extracted breath features with the response variable set as clinical probability. c) Wilcoxon test applied to the first discriminant function. d) Wilcoxon test on the first PCA scores. e) The area under the curve (AUC) for the first discriminant function score, based on clinical probability: 0.65 (95% CI 0.52–0.78), positive predictive value (PPV): 0.82 (95% CI 0.70–0.95) and negative predictive value (NPV): 0.41 (95% CI 0.27–0.56). f) The AUC for the first PCA was 0.55 (95% CI 0.40–0.70), PPV was 0.81 (95% CI 0.69–0.92) and NPV was 0.44 (95% CI 0.28–0.60).

Benzaldehyde is normally considered to be an exogenous VOC and is ubiquitous in indoor air. However, studies have shown that it can be useful as a breath biomarker [34, 35]. The two terpenes, camphene and beta-cubebene, are also generally considered to be exogenous in nature; however, their differential metabolism within certain disease groups has shown promising relevance as a diagnostic aid [33, 36].

Two other published studies examined the diagnostic potential of volatile breath markers in COVID-19. RUSZKIEWICZ *et al.* [12] were able to differentiate COVID-19 patients from other cardio-respiratory conditions using a compact GC coupled to ion mobility spectrometry (GC-IMS) in 31 COVID-19 patients from a total of 98 participants across two centres. Multivariate analysis identified seven VOCs from a total of 80 features across all samples, including aldehydes (ethanal, octanal), ketones (acetone, butanone) and methanol as the main compounds driving this separation. In an attempt to validate these compounds in our cohort, a logistic regression model was fitted with the dependant variable as PCR-positive/negative and the independent variables as the identified compounds from the Ruszkiewicz paper. None of the identified compounds held any significant discriminatory value in our cohort. This was not unexpected given the inherent differences in recruitment strategy, sampling procedure, ionisation selectivity and sensitivity of the two techniques, without prior optimisation. GRASSIN-DELYLE *et al.* [13] used real-time, online, proton transfer reaction time-of-flight mass spectrometry (PTR-MS) to measure breath VOCs of 40 ventilated patients with severe COVID-19 or non-COVID-19 acute respiratory distress syndrome, diagnosed by PCR. The prominent VOCs observed in COVID-19 patients were assigned to methylpent-2-enal, 2,4-octadiene, 1-chloroheptane and nonanal; however, the lack of chromatographic separation in this technique makes definitive chemical speciation and, subsequently, meaningful study comparison difficult to achieve.

While RT-PCR-SARS-CoV-2 remains the most widely used COVID-19 diagnostic test, its limited sensitivity, particularly in the early course of the disease, illustrates its inadequacy as an isolated COVID-19 diagnostic test [37–39], which highlights the importance of incorporating clinical probability as an essential component in making a definitive diagnosis. Nonetheless, the peculiarities of SARS-CoV-2 infection do not allow for the development of undisputed clinical criteria for a COVID-19 diagnosis, making our VOC biomarker results, based on clinical probability alone, difficult to interpret.

The diagnostic value of breath VOCs is expectedly limited in high prevalence cohorts with increased pre-test probability. However, once appropriately validated in a low COVID prevalence cohort, exhaled breath VOCs promise to be a useful tool in ruling out COVID-19 infection at the point of admission. This would facilitate extrication from isolation spaces, preserving infection control resources and preventing onward nosocomial transmission.

There are limitations associated with this study that must be taken into consideration when interpreting the findings. Firstly the sample size was relatively small: patient sampling during the first wave was largely dependent on clinical practicality and increasing the sample size was not possible owing to reduced hospital admissions in response to Public Health England recommendations. Secondly, the reference standard test used was RT-PCR: while PCR was the best available diagnostic test at the time of testing, it will have still resulted in some false-negative cases. It is also worth noting that with relatively small-size and complex datasets, overfitting can lead to poor model performance; however, to mitigate for this, feature selection was carried out by the application of LASSO regression to the peak table. External replication of these results is needed in similar populations with suspected severe COVID-19 infection to validate these findings.

Conclusions

This proof-of-concept study demonstrated the potential role of breath testing in COVID-19 diagnostics. The next stage is a large-scale diagnostic accuracy study in a population with a more realistic prevalence of COVID-19 infection to determine the clinical applicability of such a test.

Conflict of interest: None declared.

References

- 1 WHO. WHO coronavirus disease 2019 (COVID-19) situation report-101. 2020.
- 2 Corman VM, Landt O, Kaiser M, *et al.* Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* 2020; 25: 23–30.
- 3 Kucirka LM, Lauer SA, Laeyendecker O, *et al.* Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure. *Ann Intern Med* 2020; 173: 262–267.
- 4 Broughton JP, Deng X, Yu G, *et al.* CRISPR-Cas12-based detection of SARS-CoV-2. *Nat Biotechnol* 2020; 38: 870–874.

- 5 Yamayoshi S, Sakai-Tagawa Y, Koga M, *et al.* Comparison of rapid antigen tests for COVID-19. *Viruses* 2020; 12: 1420.
- 6 Hashoul D, Haick H. Sensors for detecting pulmonary diseases from exhaled breath. *Eur Respir Rev* 2019; 28: 190011.
- 7 van de Kant KDG, van der Sande LJTM, Jöbsis Q, *et al.* Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review. *Respir Res* 2012; 13: 117.
- 8 Neerinx AH, Geurts BP, Habets MFJ, *et al.* Identification of *Pseudomonas aeruginosa* and *Aspergillus fumigatus* mono- and co-cultures based on volatile biomarker combinations. *J Breath Res* 2016; 10: 016002.
- 9 Janssens E, Lamote K, Van Meerbeeck J, *et al.* Breath analysis by ion mobility spectrometry allows to discriminate COPD from lung cancer patients. *Eur Respir J* 2018; 52: Suppl. 62, PA1759.
- 10 Schleich FN, Zanella D, Stefanuto P-H, *et al.* Exhaled volatile organic compounds are able to discriminate between neutrophilic and eosinophilic asthma. *Am J Respir Crit Care Med* 2019; 199: A2373.
- 11 Ibrahim W, Carr L, Cordell R, *et al.* Breathomics for the clinician: the use of volatile organic compounds in respiratory diseases. *Thorax* 2021; 76: 514–521.
- 12 Ruszkiewicz DM, Sanders D, O'Brien R, *et al.* Diagnosis of COVID-19 by analysis of breath with gas chromatography-ion mobility spectrometry – a feasibility study. *EClinicalMedicine* 2020; 29: 100609.
- 13 Grassin-Delyle S, Roquencourt C, Moine P, *et al.* Metabolomics of exhaled breath in critically ill COVID-19 patients: a pilot study. *EBioMedicine* 2021; 63: 103154.
- 14 Durrani M, Haq IU, Kalsoom U, *et al.* Chest X-rays findings in COVID 19 patients at a University Teaching Hospital – a descriptive study. *Pak J Med Sci* 2020; 36: S22-S26.
- 15 Cadegiani FA, Zimmerman RA, Campello de Souza B, *et al.* The AndroCoV clinical scoring for COVID-19 diagnosis: a prompt, feasible, costless, and highly sensitive diagnostic tool for COVID-19 based on a 1757-patient cohort. *Cureus* 2021; 13: e12565.
- 16 Liu Z, Long W, Tu M, *et al.* Lymphocyte subset (CD4+, CD8+) counts reflect the severity of infection and predict the clinical outcomes in patients with COVID-19. *J Infect* 2020; 81: 318–356.
- 17 Lindsley AW, Schwartz JT, Rothenberg ME. Eosinophil responses during COVID-19 infections and coronavirus vaccination. *J Allergy Clin Immunol* 2020; 146: 1–7.
- 18 Luo X, Zhou W, Yan X, *et al.* Prognostic value of C-reactive protein in patients with coronavirus 2019. *Clin Infect Dis* 2020; 71: 2174–2179.
- 19 Wang L. C-reactive protein levels in the early stage of COVID-19. *Médecine et Maladies Infectieuses* 2020; 50: 332–334.
- 20 Jing QL, Liu MJ, Zhang ZB, *et al.* Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. *Lancet Infect Dis* 2020; 20: 1141–1150.
- 21 Team RC. R: A Language and Environment for Statistical Computing. 2013.
- 22 Friedman JH, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010; 33: 1–22.
- 23 Rohart F, Gautier B, Singh A, *et al.* mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol* 2017; 13: e1005752.
- 24 Thévenot EA, Roux A, Xu Y, *et al.* Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *J Proteome Res* 2015; 14: 3322–3335.
- 25 Sumner LW, Amberg A, Barrett D, *et al.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 2007; 3: 211–221.
- 26 Schurink B, Roos E, Radonic T, *et al.* Viral presence and immunopathology in patients with lethal COVID-19: a prospective autopsy cohort study. *Lancet Microbe* 2020; 1: e290–e299.
- 27 Walker HJ, Burrell MM. Could breath analysis by MS could be a solution to rapid, non-invasive testing for COVID-19? *Bioanalysis* 2020; 12: 1213–1217.
- 28 van Oort P, de Bruin S, Weda H, *et al.* Exhaled breath metabolomics for the diagnosis of pneumonia in intubated and mechanically ventilated ICU-patients. *Eur Respir J* 2017; 50: Suppl. 61, OA4653.
- 29 Ulanowska A, Kowalkowski T, Trawińska E, *et al.* The application of statistical methods using VOCs to identify patients with lung cancer. *J Breath Res* 2011; 5: 046008.
- 30 Ligor M, Ligor T, Bajtarevic A, *et al.* Determination of volatile organic compounds in exhaled breath of patients with lung cancer using solid phase microextraction and gas chromatography mass spectrometry. *Clin Chem Lab Med* 2009; 47: 550–560.
- 31 Phillips M, Greenberg J, Cataneo RN. Effect of age on the profile of alkanes in normal human breath. *Free Radic Res* 2000; 33: 57–63.
- 32 Ibrahim B, Basanta M, Cadden P, *et al.* Non-invasive phenotyping using exhaled volatile organic compounds in asthma. *Thorax* 2011; 66: 804–809.
- 33 Cazzola M, Segreti A, Capuano R, *et al.* Analysis of exhaled breath fingerprints and volatile organic compounds in COPD. *COPD Res Pract* 2015; 1: 7.

- 34 Bouza M, Gonzalez-Soto J, Pereiro R, *et al.* Exhaled breath and oral cavity VOCs as potential biomarkers in oral cancer patients. *J Breath Res* 2017; 11: 016015.
- 35 Bajtarevic A, Ager C, Pienz M, *et al.* Noninvasive detection of lung cancer by analysis of exhaled breath. *BMC Cancer* 2009; 9: 348.
- 36 Gaude E, Nakhleh MK, Patassini S, *et al.* Targeted breath analysis: exogenous volatile organic compounds (EVOC) as metabolic pathway-specific probes. *J Breath Res* 2019; 13: 032001.
- 37 Woloshin S, Patel N, Kesselheim AS. False negative tests for SARS-CoV-2 infection: challenges and implications. *N Engl J Med* 2020; 383: e38.
- 38 Zhou Y, Pei F, Ji M, *et al.* Sensitivity evaluation of 2019 novel coronavirus (SARS-CoV-2) RT-PCR detection kits and strategy to reduce false negative. *PLoS ONE* 2020; 15: e0241469.
- 39 Holborow A, Asad H, Porter L, *et al.* The clinical sensitivity of a single SARS-CoV-2 upper respiratory tract RT-PCR test for diagnosing COVID-19 using convalescent antibody as a comparator. *Clin Med (Lond)* 2020; 20: e209–e211.