Supplementary material

Table of Contents

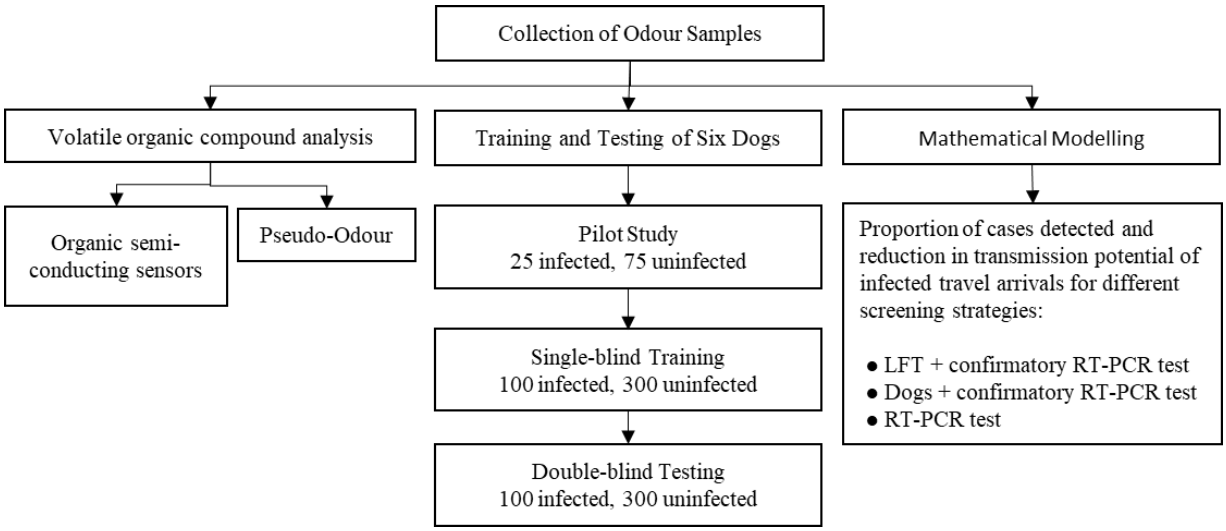
Appendix 1: COVID Dogs Research Team (CDRT) and Hospital Study Abbreviations

Number	Institution	First Name	Last Name
1, 2	Department of Disease Control, Faculty of Infectious and Tropical Diseases, and ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Robert	Jones
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Ana	Assis
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Ewan	Borthwick
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Laura	Caton
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Rachel	Edwards
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Janette	Heal
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	David	Hill
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Nazifa	Jahan
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Cecelia	Johnson
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Angela	Kaye
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Emily	Kirkpatrick
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Sarah	Kisha
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Zaena	Ledeatte Williams
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Robert	Moar
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Tolulope	Owonibi
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Benjamin	Purcell
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Christopher	Rixson
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Freya	Spencer
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Anastasios	Stefanidis
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Sophie	Stewart
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Scott	Tytheridge
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Sian	Wakley
2	ARCTEC, Chariot Innovations Ltd, London School of Hygiene and Tropical Medicine	Shanice	Wildman
3	Medical Detection Dogs	Catherine	Aziz
3	Medical Detection Dogs	Helen	Care
3	Medical Detection Dogs	Emily	Curtis
3	Medical Detection Dogs	Claire	Dowse
3	Medical Detection Dogs	Alan	Makepeace
3	Medical Detection Dogs	Sally-Anne	Oultram
3	Medical Detection Dogs	Jayde	Smith
4	Department of Biosciences, Durham University	Fiona	Shenton
5	Clinical Research Department, London School of Hygiene and Tropical Medicine	Harry	Hutchins
7	School of Chemistry, Cardiff University	Robert	Mart
14	Basildon Hospital (BSDN)	Jo-anne	Cartwright
14	Basildon Hospital (BSDN)	Miranda	Forsey
14	Basildon Hospital (BSDN)	Kerry	Goodsell
14	Basildon Hospital (BSDN)	Lauren	Kittridge
14	Basildon Hospital (BSDN)	Anne	Nicholson
14	Basildon Hospital (BSDN)	Angelo	Ramos
14	Basildon Hospital (BSDN)	Joanne	Ritches
14	Basildon Hospital (BSDN)	Niranjan	Setty
14	Basildon Hospital (BSDN)	Mark	Vertue
15	University College London Hospital (UCLH)	Malin	Bergstrom
15	University College London Hospital (UCLH)	Zain	Chaudhary
15	University College London Hospital (UCLH)	Angus	De Wilton
15	University College London Hospital (UCLH)	Kate	Gaskell
15	University College London Hospital (UCLH)	Catherine	Houlihan
15	University College London Hospital (UCLH)	Imogen	Jones
15	University College London Hospital (UCLH)	Marios	Margaritis
15	University College London Hospital (UCLH)	Patricia	Miralhes
15	University College London Hospital (UCLH)	Leah	Owens
15	University College London Hospital (UCLH)	Tommy	Rampling
15	University College London Hospital (UCLH)	Hannah	Rickman
16	Chelsea & Westminster Hospital Foundation Trust (CAWH)	Marta	Boffito
16	Chelsea & Westminster Hospital Foundation Trust (CAWH)	Candida	Fernandez
17	Kettering General Hospital NHS Foundation Trust (KETG)	Bryony	Cotterell
17	Kettering General Hospital NHS Foundation Trust (KETG)	Anne-Marie	Guerdette

17	Kettering General Hospital NHS Foundation Trust (KETG)	George	Tsaknis
17	Kettering General Hospital NHS Foundation Trust (KETG)	Margaret	Turns
17	Kettering General Hospital NHS Foundation Trust (KETG)	Joanne	Walsh
18	Buckinghamshire Healthcare NHS Trust (BUCK)	Lisa	Frankland
18	Buckinghamshire Healthcare NHS Trust (BUCK)	Raha	West
19	Macclesfield District General Hospital (MACH)	Maureen	Holland
19	Macclesfield District General Hospital (MACH)	Natalie	Keenan
19	Macclesfield District General Hospital (MACH)	Helen	Wassall
19	Macclesfield District General Hospital (MACH)	Megan	Young
20	St James's University Hospital (JUHL)	Jade	Rangeley
20	St James's University Hospital (JUHL)	Gwendolyn	Saalmink
21	Kings Mill Hospital (KMSF)	Sanjay	Adlakha
21	Kings Mill Hospital (KMSF)	Philip	Buckley
21	Kings Mill Hospital (KMSF)	Lynne	Allsop
21	Kings Mill Hospital (KMSF)	Susan	Smith
21	Kings Mill Hospital (KMSF)	Donna	Sowter
22	University Hospital Coventry & Warwickshire (UHCW)	Alison	Campbell
22	University Hospital Coventry & Warwickshire (UHCW)	Julie	Jones
22	University Hospital Coventry & Warwickshire (UHCW)	Steve	Laird
22	University Hospital Coventry & Warwickshire (UHCW)	Sarah	O'Toole
22	University Hospital Coventry & Warwickshire (UHCW)	Courteney	Ryan
23, 24	William Harvey Hospital and Queen Elizabeth the Queen Mother Hospital	Jessica	Evans
23	William Harvey Hospital (WHAD)	James	Rand
23	William Harvey Hospital (WHAD)	Natasha	Schumacher
24	Queen Elizabeth the Queen Mother Hospital (QEQM)	Tracey	Hazelton
25	Manchester Royal Infirmary (MCRI)	Andrew	Dodgson
25	Manchester Royal Infirmary (MCRI)	Susannah	Glasgow
25	Manchester Royal Infirmary (MCRI)	Denise	Kadiu
25	Manchester Royal Infirmary (MCRI)	Orianne	Lopuszansky
25	Manchester Royal Infirmary (MCRI)	Anu	Oommen
25	Manchester Royal Infirmary (MCRI)	Joshi	Prabhu
25	Manchester Royal Infirmary (MCRI)	Molly	Pursell
25	Manchester Royal Infirmary (MCRI)	Jane	Turner
25	Manchester Royal Infirmary (MCRI)	Hollie	Walton
26	Musgrove Park Hospital (MGPH)	Robert	Andrews
26	Musgrove Park Hospital (MGPH)	Irena	Cruickshank
26	Musgrove Park Hospital (MGPH)	Catherine	Thompson
26	Musgrove Park Hospital (MGPH)	Tania	Wainwright
27, 28	Pilgrim Hospital and Lincoln County Hospital	Alun	Roebuck
27	Pilgrim Hospital (PGHL)	Tara	Lawrence
27	Pilgrim Hospital (PGHL)	Kimberley	Netherton
28	Lincoln County Hospital (PGHL-LIN)	Claire	Hewitt
28	Lincoln County Hospital (PGHL-LIN)	Sarah	Shephardson
29	George Eliot Hospital (GETH)	Winston Andrew	Crasto
29	George Eliot Hospital (GETH)	Judith	Lake
29	George Eliot Hospital (GETH)	Rosemary	Musanhu
29	George Eliot Hospital (GETH)	Rebecca	Walker
30	University Hospital Morecambe Bay (UHMB)	Karen	Burns
30	University Hospital Morecambe Bay (UHMB)	Andrew	Higham
30	University Hospital Morecambe Bay (UHMB)	Julie	Le Bas
30	University Hospital Morecambe Bay (UHMB)	Nicola	Mackenzie
30	University Hospital Morecambe Bay (UHMB)	Hilary	Thatcher
31	Mid Yorkshire Hospitals NHS Trust (MYSH)	Shannen	Beadle
31	Mid Yorkshire Hospitals NHS Trust (MYSH)	Sarah	Buckley
31	Mid Yorkshire Hospitals NHS Trust (MYSH)	Gail	Castle
31	Mid Yorkshire Hospitals NHS Trust (MYSH)	Aimee	Fletcher
31	Mid Yorkshire Hospitals NHS Trust (MYSH)	Sara	Holbrook
31	Mid Yorkshire Hospitals NHS Trust (MYSH)	Patricia	Kane
31	Mid Yorkshire Hospitals NHS Trust (MYSH)	Kate	Lindley
31	Mid Yorkshire Hospitals NHS Trust (MYSH)	Tracey	Lowry
31	Mid Yorkshire Hospitals NHS Trust (MYSH)	Stephanie	Lupton

31	Mid Yorkshire Hospitals NHS Trust (MYSH)	Sharon	Oddy
31	Mid Yorkshire Hospitals NHS Trust (MYSH)	Lynda	Slater
31	Mid Yorkshire Hospitals NHS Trust (MYSH)	Martin	Sylvester
32	Doncaster & Bassetlaw Teaching Hospital (DBTH)	Kenneth	Agwuh
32	Doncaster & Bassetlaw Teaching Hospital (DBTH)	Veronica	Maxwell
33	Nottingham University Hospital (NHUH)	Stephen	Ryder
33	Nottingham University Hospital (NHUH)	Kirsty	Topham
34	Central and North West London NHS Foundation Trust (CNWL)	Obi	Egbuniwe
34	Central and North West London NHS Foundation Trust (CNWL)	Rebecca	Matthews
34	Central and North West London NHS Foundation Trust (CNWL)	Alejandro Arenas	Pinto
34	Central and North West London NHS Foundation Trust (CNWL)	Paulina	Prymas
34	Central and North West London NHS Foundation Trust (CNWL)	Abigail	Severn
34	Central and North West London NHS Foundation Trust (CNWL)	Amber	Shaw
35	University Hospitals Birmingham NHS Foundation Trust (BHAM)	Safia	Begum
35	University Hospitals Birmingham NHS Foundation Trust (BHAM)	Daniel	Lenton
35	University Hospitals Birmingham NHS Foundation Trust (BHAM)	Jamie	Scriven
36	Plymouth Hospitals NHS Trust (PLYM)	Lucy	Leeman
36	Plymouth Hospitals NHS Trust (PLYM)	Karen	Rudge
36	Plymouth Hospitals NHS Trust (PLYM)	Emma	Storr
37	Agile Lighthouse	Ana	Alvarez
37	Agile Lighthouse	Kate	Forster
37	Agile Lighthouse	Daniel	Hind

Appendix 2: Flow chart of study design.



Infected = SARS-CoV-2 RT-PCR positive, Infected = SARS-CoV-2 RT-PCR negative, LFT = lateral flow test, RT-PCR = reverse transcription-polymerase chain reaction

Appendix 3: Methods

Method S1: Collection and analysis of odour profiles by OSC (organic semi-conducting) sensors

For each sample, one worn sock was placed in a clean aluminium foil bag (20 cm x 14 cm, WACCOMT Pack INC, CN). The bag was heat sealed. A 18G needle was then inserted into the bag and air pumped into the bag until inflated to 160 ml. The inflated bags were then incubated at 40˚C for 30 min to volatilise the organic compounds into the headspace air. Each bag was then allowed to cool to room temperature (20˚C) and sampled four times, in the fume hood using a Model 307B VOC analyser (RoboScientific Ltd, Cambridgeshire, UK) fitted with a 12-OSC sensor array chosen to be sensitive to the VOCs which were likely to be associated with SARS-CoV-2, based on previous analysis (including ethanol, acetone, methanol, propanol, octanal, heptanal, propanal). Data from the sensor array were automatically recorded by the Roboscientific 307B control software.[1] The VOC analyser was internally sterilised using vapour from 70% ethanol solution to ensure no contaminated air remained in the pipework and sensor array used between each run.

Method S2: Part I of dog training to detect odours from people infected with SARS CoV-2: pilot study

A pilot study was conducted to confirm that three dogs previously trained to discriminate training odour could distinguish between 25 samples taken from participants who tested positive for SARS-CoV-2 and 75 samples from participants who tested negative for SARS-CoV-2 by reverse transcription-polymerase chain reaction (RT-PCR). Face masks and socks were presented to the dogs in order to determine which sample type offered the best discrimination. Socks appeared to give the strongest signature and, therefore, training continued with these samples.

Training sessions required one Bio-Detection trainer to handle and train the dog to discriminate odour, and one assistant to handle samples during the training session. For presentation to the dogs, the lid of the sample was removed, and each glass vial clipped into a stainless-steel arm with a grill covering the vial opening. Each arm was then placed in stainless steel retort stand, placed in a line (Figure S2.1). Cross contamination from one run to the next was prevented by cleaning the plates after every run using a commercial glasswasher at a minimum of 85°C and leaving them to air dry or drying with paper towel before being used again. Glass vials were autoclaved to prevent cross-contamination.

During the pilot study, samples were presented to the dogs in sets of four with no more than one positive sample in each line, however, the dogs were also presented with lines of only negative samples (blank lines) under both unblinded (where the Bio-Detection Trainer knows the positive or negative status of each presented samples in the line) and single blind (where only the assistant placing the samples in the retort stands knows the positive or negative status of each samples in the line) conditions (Figure S2.2).

Figure S2.1: An example of a stand containing a covid sample and dog training using stands.

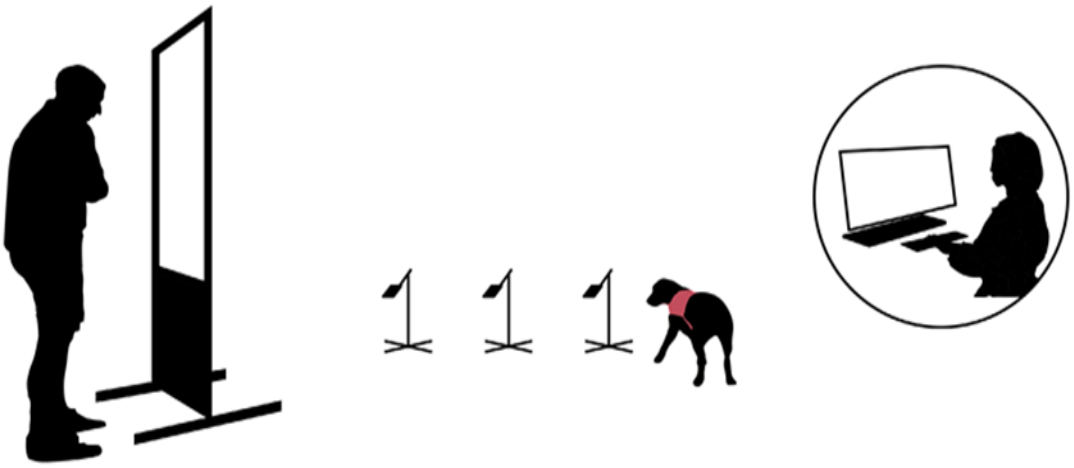


Figure S2.2: Dog training using three stands

The dog trainer tasks the dog to search each stand, off lead with a search command. The trainer then called the dog's decision based on the dog's behaviour i.e. If the dog indicated a sample (either by a sit or stand indication specific to each dog), the trainer called 'indication'. If the dog left the stands without offering an indication, the trainer calls 'negative'. The result was then entered and recorded on a database. If training under single-blind conditions, the decision was verified by the assistant. Under both unblinded and single-blind conditions, a correct response was rewarded with an audible click and food or toy reward. All responses were recorded on specially designed computer software, Medical Detection Dogs – Olfactory Performance Recording Application, (OPRA). Sensitivity and specificity were calculated separately for each trained dog.

Method S3: Part II of dog training to detect odours from people infected with SARS CoV-2: single-blind training study

This single-blind study was a progression of the pilot study and was designed to estimate an approximate level of sensitivity and specificity of the dogs to detect participants infected with SARS-CoV-2. The dogs that had been pre-trained in the pilot study were trained for a further six to eight weeks, along with four additional dogs. Early recognition of the scent of a SARS-CoV-2-positive sample by a dog was achieved using search and find games, which were gradually replaced by discrimination phases. During training, the reaction of each dog to a positive sample was observed (i.e., standing, sitting or lying down) and this indicating behaviour reinforced by rewarding the dog with and audible clicker followed by food or ball-play. Samples from 105 SARS-CoV-2 – positive participants and 316 negative participants were used in this stage of training.

The experimental set-up used for training and testing consisted of a number of stainless-steel retort stands, each holding an arm with a sealed and vented glass vial containing a sample (Figure S2.1 and S2.2). A grill placed over the mouth of the vials prevents the dog touching the specimen. Each glass vial contained either one positive or one negative sample. Cross contamination from one run to the next was prevented by cleaning the plates after every run using a commercial glasswasher at a minimum of 85°C and leaving them to air dry or drying with paper towel before being used again. Clean arms were used for every sample change. A dog's behaviour at the stand was noted as full alert (indication), heavily investigated (hesitation), weakly investigated (interest) or ignored (no interest). The handler called the final decision as an indicated sample or blank. A dog was rewarded with food or a ball, when it correctly indicated either positive or a negative run.

Progression from unblinded to single blind training and increasing complexity of the discrimination task was dependent on each dog's individual progress. To assist with this, samples were carefully selected, and positive and negative samples matched based on factors including participants' age, gender and ethnicity. The dog's decision on each sample was recorded and rewarded as in the pilot study, however, when training progressed onto multiple target lines, samples where a decision had been 'called' by the dog trainer were then removed and replaced with a new sample to interrogate. Interrogation of samples continued until the dog had given a decision

on each individual sample. All responses were recorded on specially designed computer software, OPRA. Sensitivity and specificity were calculated separately for each trained dog.

Method S4: Part III of dog training to detect odours from people infected with SARS CoV-2: double-blind testing

Six dogs were considered trained and used for testing (Table S4.1). For presentation to the dogs, the samples were defrosted at room temperature for 60 min. The lid was then removed, and each glass vial clipped into a stainless-steel arm with a grill covering the vial opening. Each arm was then placed in a stainless-steel retort stand (Figure S2.1 and S2.2). Cross contamination from one run to the next was prevented by cleaning the plates after every run using a commercial glasswasher at a minimum of 85°C. Disposable gloves were worn throughout the testing and changed them between samples. Tests were performed in an air-conditioned room at 14·2-20·6ºC and 38-53% relative humidity.

A preparator (SM, SA) not involved in dog handling, prepared the randomisation schedule for the trial. Samples were loaded onto the stands by a blinded assistant (CD). A blinded handler, positioned behind a one-way screen (so that the dog when working could not receive visual prompts), tasked the dog to search the stands off lead. Once a decision had been made by a dog, the blinded handler called the sample as infected or uninfected, and the result recorded by a blinded monitor on the database, which then revealed the answer to be correct or incorrect. The result was then double verified by an unblinded monitor. If the correct answer was given, the handler then rewarded the dog with an audible click and food or toy reward. The sample was then removed from the test line and replaced with a random but known filler sample (positive or negative status) so that the dog could be tasked on to search any remaining test samples. Data was not collected from filler samples. This process was repeated until all test samples in the line had been searched and a decision made on all samples. Once the line was complete, a new, full test line was presented, and the process repeated. All dogs were allowed a maximum of three explorations per sample to allow the handler to confirm a behaviour offered by the dog before making a final decision. Dogs were not permitted to take multiple passes of the line which may have resulted in more than three explorations of any test sample.

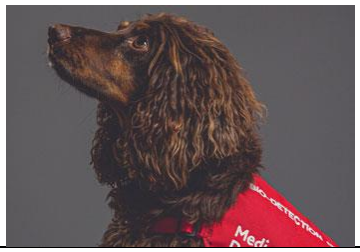
	Photo	Breed	Sex	Age
Asher		Cocker spaniel	M	8
Kyp		Labrador cross	M	4
Lexi		Labrador	F	5
Marlow		Labrador	M	4
Millie		Golden Retriever	F	4
Tala		Labrador	M	3

Table S4.1: Characteristics of dogs trained in the study

Methods S5: Sample Size calculations

For the double-blind testing, the number of independent samples, encountered by each dog during testing determined the precision with which its sensitivity and specificity could be estimated. With 200 positive samples i.e. from individuals positive for SARS-CoV-2 RNA by real-time RT-PCR, and an expected sensitivity of 85%, observed estimate would have a 95% Confidence Interval (95% CI) of 79% to 90%. With 200 true negative samples, and an expected specificity of 90%, observed specificity would have a 95% CI of 85% to 94%.

With regards to VOC analysis, based on previous studies,[2] at least 25 sock samples from infected and uninfected participants were required to build a disease-specific model with electronic sensors.

Methods S6: PCA and Discriminant Analyses

A principal component analyses (PCA) was performed for each of the days separately using the responses from the 12 sensors based on a centred and standardized responses (i.e., matrix of correlations). Following these analyses, scree-plots were examined to determine the number of relevant dimensions and a biplot generated to discriminate which sensor is more associated with the groups under evaluation (negative or positive groups).

In addition, a discriminant analysis (DA) was performed to each day separately, with the aim of obtain and assess if a model will allow to discriminate between the two groups. The dataset used corresponded to the same 12 sensors recorded. In order to assess the 'predictability' of the model fitted, a cross-validation procedure was followed, where 20% of the observations were selected at random as validation set, and the other 80% was considered as training set. After fitting the DA model, the predictions of the testing set were evaluated and a sensitivity and specific was obtained. The previous process was repeated 20 times for each of the day datasets.

Methods S7: Bayesian latent class analysis of Dog Covid Diagnostics

We fitted five different models with varying assumptions. The first. called *gold.* assumes PCR to be a perfect reference standard with 100% sensitivity and specificity and assumes that there is no correlation between the results from different dogs, that is if one dog indicates other dogs do not become more likely to indicate as well given the disease status. The second, called *nocorr*, does not assume that PCR is a perfect reference standard but instead puts semi-informative priors on the diagnostic accuracy of PCR (s~beta(5,1), with 95% of the prior probability mass above s=0.55, c~beta(10,1) with 95% of the mass above c=0.74). The third, called *gold corr*, assumes again PCR to be a perfect reference standard while allowing for correlation between the dogs for both Covid positive and negative participants using a random effect model. The fourth, called *noninf*, allows for conditional dependence and does not assume PCR to be a perfect reference standard but instead uses the same vague priors as above. The fifth, called *info*, is the same model as *noninf* but with more highly informative priors on PCR (s~beta(16,2) giving a 95% range from 0.71 to 0.99, c~beta(93,2) giving a 95% prior range to be above 94%). The sensitivity of each dog was modelled as $logit(s_i^{Dog-x}) = logit(s_0^{Dog-x}) + \epsilon_i$ where i is the individual, $\epsilon_i \sim N(0, \sigma)$. s_0^{Dog-x} has a beta(0.75,0.75) prior and σ a Gamma(1,1) which results in a non-informative prior for the mean of s_i^{Dog-x}.

Methods S8: Mathematical Modelling

Briefly, the cycle threshold (Ct) of infected individuals was simulated from trajectories defined by a starting Ct, a peak Ct and a total duration of infection, assuming a random time since exposure. We assumed that 31% of individuals were asymptomatic on average[3] (with 40% shorter duration of infection)[4] and that 70% of individuals with symptoms prior to departure would not travel[5-6] (thus increasing the proportion of asymptomatic cases among those who travel). Dogs were assumed to be able to detect infection with a sensitivity range between 80-90% informed from the results of the double-blind testing. The sensitivity of PCR was assumed to be either 100% up to a Ct of 35 and 0% thereafter or 100% up to a Ct of 40. This permitted exploration of the impact of uncertainty in the sensitivity of PCR for detecting low viral loads (with Ct between 35 and 40) when used in practice. The Ct-dependent sensitivity of the LFT[7] was estimated by fitting a logistic function to the data presented in Peto.[8]

Appendix 4: Results

Results S1: Adverse events

A total of 343 adverse events were recorded from human participants in the study. Only two were possibly related to the study. One reported an itchy body from wearing the shirt and one reported a spot on the nose from wearing a face mask. Ten serious adverse events were recorded, none of which were deemed related to the study.

Twelve adverse events were reported in the dogs, with one dog (Asher) experiencing a severe adverse event and being withdrawn from the study. All adverse events were unrelated to the study.

Results S2: Sample characteristics used for OSC Sensors

	Infected group (RT-PCR +ve, n=26)	Uninfected group (RT-PCR -ve, n=27)
Source of sample		
NHS hospitals	4 (15·4%)	22 (81·5%)
ARCTEC/LSHTM call centre & Agile Lighthouse	22 (84·6%)	5 (18·5%)
Gender		
Women	17 (65·4%)	18 (66·7%)
Men	9 (34·6%)	9 (33·3%)
Age, years		
Age: 16-50	16 (61·5%)	14 (51·8%)
Age: 50+	10 (38·5%)	13 (48·1%)
Ethnicity		
White	25 (96·2%)	26 (96·3%)
Black	1 (3·8%)	1 (3·7%)
Symptoms at enrolment		
Classic SAR-CoV-2	13 (50·0%)	10 (37·0%)
Non-Classic SAR-CoV-2	13 (50·0%)	10 (63·0%)
Hospital patients	0 (0%)	0 (0%)
Symptoms at sample receipt at site		
Classic SAR-CoV-2	8 (30·7%)	0 (0%)
Non-Classic SAR-CoV-2	13 (50·0%)	24 (88·9%)
Unknown	5 (19·2%)	3 (11·1%)
Symptoms after 14 days		
Classic SAR-CoV-2	6 (23·1%)	1 (3·7%)
Non-Classic SAR-CoV-2	18 (69·2%)	24 (88·9%)
Unknown	2 (7·7%)	2 (7·4%)

Symptoms at enrolment, at sample receipt at site and 14-day follow-up were categorised as “classic SARS-CoV-2” if fever, cough, or loss or change of smell or taste were reported, and “non-classic SARS-CoV-2” for those who reported no symptoms or where other symptoms were reported, including, shortness of breath, abdominal pain, muscle and joint pain, conjunctivitis or nausea. NHS hospitals: CAWH (2 uninfected), DBTH (3 uninfected), JUHL (5 uninfected), KMSF (1 infected, 3 uninfected), MACH (1 infected, 3 uninfected), MCRI (1 uninfected), MGPH (3 uninfected), PGHL (1 uninfected), UHCW (1 infected, 2 uninfected). All swabs were processed through routine NHS channels.

Results S3: Part I of dog training to detect odours from people infected with SARS CoV-2: pilot study

The pilot study comprised 100 samples (75 negative, 25 positive), to determine if SARS-CoV-2 had an odour that could be detected by dogs (Table S3.1). The highest performing dog achieved 88.3% sensitivity and 90.0% specificity. Overall, the three dogs achieved a sensitivity range of 75.9-88.3% and a specificity range of 90.0-95.1%. These results represent all encounters with the samples under single blind conditions, with the result from each dog decision taken from the number of times the sample was correctly classified in the final pass. The samples included in this part of the study had previously been used in training, so these results do not account for the effects of novel exposure to new samples. Despite this, single-blind training conditions provide a more accurate representation of the dogs' ability to detect the odour in testing, as the trainer does not know the status of the samples in the line. This means that the trainer cannot cue the dog onto target (positive) samples and must rely on reading the indication behaviour offered by the dog. Following these encouraging initial results, we progressed to a larger number of samples and more dogs.

	Infected group (RT-PCR +ve, n=25)	Uninfected group (RT-PCR -ve, n=75)
Source of sample		
NHS hospitals	11 (44·0%)	44 (58·7%)
ARCTEC/LSHTM call centre & Agile Lighthouse	14 (56·0%)	31 (41·3%)
Gender		
Women	15 (60·0%)	45 (60·0%)
Men	10 (40·0%)	30 (40·0%)
Age, years		
Age: 16-50	18 (72·0%)	54 (72·0%)
Age: 50+	7 (28·0%)	21 (28·0%)
Ethnicity		
White	19 (76·0%)	57 (76·0%)
Asian	3 (12·0%)	9 (12·0%)
Other	1 (4·0%)	3 (4·0%)
Unknown	2 (8·0%)	6 (8·0%)
Symptoms at enrolment		
Classic SAR-CoV-2	21 (84·0%)	24 (32·0%)
Non-Classic SAR-CoV-2	4 (16·0%)	51 (68·0%)
Hospital patients	0 (0%)	0 (0%)
Symptoms at sample receipt at site		
Classic SAR-CoV-2	8 (32·0%)	2 (2·7%)
Non-Classic SAR-CoV-2	16 (64·0%)	73 (97·3%)
Unknown	1 (4·0%)	0 (0%)
Symptoms after 14 days		
Classic SAR-CoV-2	6 (24·0%)	1 (1·3%)
Non-Classic SAR-CoV-2	19 (76·0%)	72 (96·0%)
Unknown	0 (0%)	2 (2·7%)

Table S3.1. Sample characteristics used for pilot study. Symptoms at enrolment, at sample receipt at site and 14-day follow-up were categorised as "classic SARS-CoV-2" if fever, cough, or loss or change of smell or taste were reported, and "non-classic SARS-CoV-2" for those who reported no symptoms or where other symptoms were reported, including, shortness of breath, abdominal pain, muscle and joint pain, conjunctivitis or nausea. NHS hospitals: BHAM (1 uninfected), BUCK (3 uninfected), CAWH (1 uninfected), DBTH (3 infected, 3 uninfected), GETH (2 uninfected), JUHL (2 uninfected), KETG (2 infected, 6 uninfected), KMSF (1 infected, 3 uninfected), MACH (2 uninfected), MCRI (1 uninfected), MGPH (2 uninfected), MYSH (1 infected, 3 uninfected), PGHL (2 uninfected), PGHL-LIN (1 uninfected), UHCW (3 infected, 6 uninfected), UHMB (3 uninfected), WHAD (1 infected, 3 uninfected). All swabs were processed through routine NHS channels.

Results S4: Part II of dog training to detect odours from people infected with SARS CoV-2: single-blind training study

Characteristics of the study samples are summarised in Tables S4.1. After 6 weeks of training with 105 positive and 316 negative sock samples, six dogs achieved a sensitivity range of 75.1%-83.9% and a specificity range 90.8-95.4% under single-blind conditions (Table S4.2). Lexi was the best performing dog, achieving 83.9% sensitivity and 95.4% specificity, whilst the lowest performing dog, Tala, achieved a 75.1% sensitivity and 90.8% specificity.

Samples presented in training may have been presented to the dogs during unblinded training (where the trainer knows the status and position of each samples in the line to promote odour recognition) prior to presentation under single blind conditions, however, the rate of novel exposure in single blind increased closer to the start of double-blind testing. Through training, seven samples (three positives, five negatives) were presented to the

dogs which proved to be highly challenging under both unblinded and single blind conditions. Samples were considered challenging when the sample was correctly classified <66.7%, with agreement from three or more dogs. This may suggest that these samples were false positive and false negative by PCR testing. When these samples are removed from the results, increases in sensitivity and specificity are seen, as shown in Table S4.2.

	Infected group (RT-PCR +ve, n=105)	Uninfected group (RT-PCR -ve, n=316)
Source of sample		
NHS hospitals	18 (17·1%)	296 (93·7%)
ARCTEC/LSHTM call centre & Agile Lighthouse	87 (82.9%)	20 (6·3%)
Gender		
Women	80 (76·2%)	239 (75·6%)
Men	23 (21·9%)	71 (22·5%)
Intersex	2 (1·9%)	6 (1·9%)
Age, years		
Age: 16-50	63 (60·0%)	187 (59·2%)
Age: 50+	42 (40·0%)	129 (40·8%)
Ethnicity		
White	100 (95·2%)	301 (95·3%)
Asian	2 (1·9%)	6 (1·9%)
Other	1 (1·0%)	3 (0·9%)
Unknown	2 (1·9%)	6 (1·9%)
Symptoms at enrolment		
Classic SAR-CoV-2	88 (83·8%)	39 (12·3%)
Non-Classic SAR-CoV-2	17 (16·2%)	277 (87·7%)
Hospital patients	3 (2.9%)	0 (0%)
Symptoms at sample receipt at site		
Classic SAR-CoV-2	38 (36·2%)	3 (0·9%)
Non-Classic SAR-CoV-2	57 (54·3%)	293 (92·7%)
Unknown	10 (9·5%)	20 (6·3%)
Symptoms after 14 days		
Classic SAR-CoV-2	18 (17·1%)	3 (0·9%)
Non-Classic SAR-CoV-2	84 (80·0%)	303 (95·9%)
Unknown	3 (2·9%)	10 (3·2%)

Table S4.1: Sample characteristics used for training and single-blind testing. Symptoms at enrolment, at sample receipt at site and 14-day follow-up were categorised as "classic SARS-CoV-2" if fever, cough, or loss or change of smell or taste were reported, and "non-classic SARS-CoV-2" for those who reported no symptoms or where other symptoms were reported, including, shortness of breath, abdominal pain, muscle and joint pain, conjunctivitis or nausea. NHS hospitals: BHAM (3 uninfected), BSDN (5 uninfected), BUCK (35 uninfected), CAWH (3 uninfected), DBTH (1 infected, 4 uninfected), GETH (5 uninfected), JUHL (2 infected, 13 uninfected), KETG (50 uninfected), KMSF (2 infected, 30 uninfected), MACH (5 infected, 16 uninfected), MCRI (1 infected, 5 uninfected), MGPH (33 uninfected), MYSH (1 infected, 31 uninfected), PGHL (2 infected, 8 uninfected), PGHL-LIN (1 uninfected), QEQM (1 uninfected), UHCW (2 infected, 10 uninfected), UHMB (1 infected, 3 uninfected), WHAD (1 infected, 40 uninfected). All swabs were processed through routine NHS channels.

	Average %		Average % (challenging samples removed)	
	Sensitivity	Specificity	Sensitivity	Specificity
Asher	79·2	94·2	82·9	94·7
Kyp	77·4	93·9	79·4	94·1
Lexi	83·9	95·4	85·6	96·0
Marlow	80·0	93·5	80·9	93·5
Millie	76·0	93·4	78·4	93·4
Tala	75·1	90·8	78·1	91·2

Table S4.2: Single-blind results. Average percentage sensitivity and specificity of each dog recorded over a six-week single-blind training phase (n = 105 positive and 316 negatives per dog) and with challenging samples removed (n = 102 positive and 311 negatives per dog).

Results S5: Bayesian latent class analysis of dog diagnostics

Results and corresponding 95% BCI are presented in the Table S5.1. There is a clear indication for correlation between the dogs and therefore, the models *noninf* and *info* are most relevant for interpretation. There is little difference between diagnostic accuracy of dogs between the model with informative and vague priors for the reference standard. Mean estimates of the sensitivity and specificity of dogs range from 82.1% to 94.3%, and 76.4% to 92.0% in the case of vague priors. For informative priors, the mean estimates of the sensitivity and specificity of dogs range from 81.4% to 93.1%, and 77.0% to 92.3%. Estimates of PCR sensitivity are similar for both models at 95.5% and 95.0%. Estimates of PCR specificity are influenced by the prior that assumed it to be above 95% with 95% certainty with 93.8% for the vague priors and 95.9% for the informative priors. The model which is most similar to a separate analysis for each dog with PCR as the gold standard is the *gold corr* model.

Estimates for the diagnostic accuracy of dogs for SARS-CoV-2 from the full model taking into account the imperfect diagnostic accuracy of PCR and conditional dependence between the dogs are very similar to the model neglecting both. This indicates that the estimates are not sensitive to the reference standard used in this case due to the high sensitivity and specificity estimated for PCR. There is evidence for an imperfect specificity of PCR with the upper end of the 95% BCI at 98.6%. The sensitivity of PCR is estimated to be high at 95.9% which is at the upper end of the prior evidence included in the study.

	gold		nocorr		gold corr		noninf		info	
	Sensitivity % (95% CI)	Specificity % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)
Asher	91·5 (86·3-95·7)	85·6 (79·0-91·0)	90·1 (84·1-94·8)	84·7 (77·8-90·5)	88·1 (82·4-92·8)	83·3 (76·8-88·8)	90·9 (85·3-95·4)	84·8 (77·9-91·1)	90·8 (85·0-95·2)	85·6 (78·8-91·3)
Kyp	88·4 (83·6-92·6)	77·4 (71·4-83.1)	88·9 (84·0-93·0)	77·0 (70·6-82·6)	85·2 (80·3-89·5)	74·9 (69-80·5)	88·5 (83·6-92.8)	76·4 (70·3-82·1)	88·1 (83·4-92·3)	77·0 (71·2-82·7)
Lexi	89.0 (84·2-93·0)	84·8 (79·6-89·6)	92·1 (87·8-95·5)	86.7 (81.5-91·2)	85·3 (80·5-89·6)	81·9 (76·8-86·7)	90·8 (86·0-94·9)	85·3 (79·9-90·2)	89·9 (84·8-94·2)	85·6 (80·1-90·3)
Marlow	81·1 (75·4-86·3)	90·4 (85·9-94·1)	82·6 (76·7-87·8)	90·7 (86·2-94·4)	77·7 (72·0-82.9)	88·0 (83·5-91·9)	82·1 (76·3-87·3)	90·1 (85·4-93·9)	81·4 (75·3-86·8)	90·6 (85·9-94·4)
Millie	84·2 (78·9-88·8)	82·5 (77·0-87·5)	86·4 (81·3-91·0)	83·8 (78·4-88·6)	80·7 (75·4-85·4)	79·9 (74·4-84·7)	85·5 (80·1-90·5)	82·6 (76·9-87·6)	84·6 (78·6-89·7)	82·9 (77·4-87·7)
Tala	92·0 (87·9-95·5)	91·2 (86·8-94·8)	95·4 (91·6-98·3)	93·2 (89·4-96·4)	88·3 (83·9-92·3)	88·5 (83·9-92·5)	94.3 (89.4-98)	92·0 (87·6-95·8)	93·1 (87·8-97·2)	92·3 (87·7-96·0)
PCR	99·9 (99·9-100)	99·9 (99·9-100)	94·2 (90·3-97·2)	92·4 (88·4-95·8)	100 (100-100)	100 (100-100)	95·9 (92·2-98·8)	93·8 (89·2-97·5)	95·5 (91·7-98·3)	95·9 (92·6-98·6)

Table S5.1 Model results from the Bayesian analysis. Where gold assumes RT-PCR is a perfect test and there is no correlation in the way that dogs respond to an odour sample, nocorr assumes the RT-PCR is imperfect and there is no correlation between dogs, gold corr assumes that RT-PCR is imperfect and there is correlation in the way dogs respond to the samples, noninf assumes that RT-PCR is imperfect with a higher sensitivity than specificity (since RT-PCR will identify people several weeks after stopping being infectious, info assumes that RT-PCR is imperfect).

Results S6: Sensitivity and specificity and sample characteristics

	RT-PCR +ve				RT-PCR -ve			
	n/N	Sensitivity (%)	Odds Ratio (95% CI)	p-value	n/N	Specificity (%)	Odds Ratio (95% CI)	p-value
Sex								
Female	712/831	85.7	1		736/873	84.3	1	
Male	245/298	82.2	0.77 (0.54, 1.10)	0.155	206/259	79.5	0.72 [0.50, 1.03]	0.07
Age Group								
16-50	614/724	84.8	1		541/656	82.5	1	
50+	343/405	84.7	0.99 (0.70, 1.38)	0.933	401/476	84.2	1.14 [0.83, 1.57]	0.427
Ethnicity								
Other	42/58	72.4	1		31/46	67.4	1	
White	915/1071	85.4	2.28 (1.25, 4.19)	0.008	825/986	83.7	2.51 [1.32, 4.79]	0.005
Symptoms at enrolment								
No Symptoms	58/81	71.6	1		714/839	85.1	1.37 [0.94, 2.00]	
Classic SARS-CoV-2	717/838	85.6	2.34 (1.39, 3.96)		189/234	80.8	1	
Non-classic SARS-CoV-2	182/210	86.7	2.57 (1.37, 4.83)	0.004	39/59	66.1	0.46 [0.24, 0.86]	0.001

Appendix 6: References

[1] Ruszkiewicz, Dorota M et al. "Diagnosis of COVID-19 by Analysis of Breath with Gas Chromatography-Ion Mobility Spectrometry - a Feasibility Study." *EClinicalMedicine* 29-30 (2020): 100609–. Web.

[2] Wintjens AGWE, Hintzen KFH, Engelen SME, Lubbers T, Savelkoul PHM et al. Applying the electronic nose for pre-operative SARS-CoV-2 screening. *Surg Endosc* 2020 Dec. https://doi.org/10.1007/s00464-020-08169-0

[3] Buitrago-Garcia DC, Egli-Gany D, Counotte MJ, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: a living systematic review and meta-analysis. *PLoS Med* 2020; 17**:** e1003346.

[4] Kissler SM, Fauver JR, Mack C, et al. SARS-CoV-2 viral dynamics in acute infections. *medRxiv* 2020; published online Dec 1. https://doi.org/10.1101/2020.10.21.20217042 (preprint).

[5] Gostic K, Gomez AC, Mummah RO, Kucharski AJ, Lloyd-Smith JO. Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19. *eLife* 2020; **9**. DOI:10.7554/eLife.55570.

[6] Clifford S, Quilty BJ, Russell TW, *et al.* Strategies to reduce the risk of SARS-CoV-2 re-introduction from international travellers. *medRxiv*, 2020 DOI:10.1101/2020.07.24.20161281.

[7] Wise J. Covid-19: Lateral flow tests miss over half of cases, Liverpool pilot data show. (Published 15th December 2020) *BMJ* 2020;371:m4848

[8] COVID-19: Rapid Antigen detection for SARS-CoV-2 by lateral flow assay: a national systematic evaluation for mass-testing. Tim Peto, UK COVID-19 Lateral Flow Oversight Team *medRxiv* 2021.01.13.21249563