Canine Scent Detection:

Lung Cancer Target Acquisition

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Psychology

at

The University of Waikato

by

MARGARET A. CRAWFORD

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2023

The general introduction and research problem are provided in the opening section of Chapter 1. The subsequent chapters each begin with a chapter introduction that leads into a manuscript. This thesis includes one unpublished manuscript (Chapter 2) and two published manuscripts (Chapters 3 and 4). Chapters 3 and 4 take the same form as the published papers apart from minor figure and table label changes to maintain continuity. The synthesis of the research presented in Chapters 2 – 4 is provided in the final chapter (Chapter 5), titled General Discussion. Supplementary materials and quantitative analysis model selection criteria are included in the appendices.

Abstract

A growing body of research has developed over the last 30 years exploring disease scent detection using animals. Research project methodologies and results in this field vary significantly, including some which are quite promising. Critiques of aspects of current and past practice in disease scent detection using animals inform recommendations for the development of scientifically robust standard operating procedures in this field. Greater standardisation and transparency of practice has the potential to strengthen and clarify disease scent detection results. My primary aim was to contribute to this standardisation by using a theoretical understanding of concept formation and learning acquisition to inform three experiments in which pet dogs (*Canis lupus familiaris*) were trained to detect the presence of lung cancer in human breath and saliva samples. Operant conditioning processes and an automated apparatus were used to train and test the dogs. Data for all three experiments were collected concurrently. The experiments involved: designing a process for evaluating sample comparisons; evaluating the efficacy of sample re-use for training purposes; and developing a mathematical model to support decision-making about the transition from training to testing of detector organisms. Firstly, we evaluated the comparative utility of human breath and saliva samples to train and test dogs for lung cancer detection. Signal detection measures were used to gauge the dogs' target acquisition and concept formation during the training process. The dogs acquired the lung cancer target concept more quickly from breath samples, but also demonstrated higher-than-chance recognition using saliva samples. Secondly, we systematically evaluated the effect of breath sample re-use on dogs' performance during lung cancer scent detection training. There were no significant changes associated with the detectability of the target across samples re-used up to four times, and observed changes in performance were small. Finally, we explored methods of evaluating when a detection animal is performing at or near the highest accuracy of which they are capable with a view to

identifying the optimal point at which to transition from training to testing. A quantitative model was most informative during our work training dogs to detect lung cancer. Ongoing testing of the dogs' abilities using novel samples occurred during training. However, the final testing (and intermittent training) needed to measure and maintain the dogs' performance against asymptotic predictions was beyond of the scope of the current project. Notwithstanding this, each of the experiments described herein provides both specific data on the performance of our dogs, and procedural information about ways in which different components of scent detection methodology using detector organisms could be strengthened. Both of these resources could be used in concert with other methods of scent detection. Relevant theory on concept formation was reviewed and used to plan and interpret the acquisition of the target scent. Likewise, theoretical explanations of target acquisition and associative learning informed our analysis of data. In this way, each of these experiments contribute to improving practice in the field of disease detection by providing model procedures from which other methods for evaluating sample types and the reuse of samples could evolve. The development of a standardised measure for determining when to stop training and start testing a detecting organism might also be applied to a wide range of learners in myriad contexts to good effect.

Acknowledgements

I would like to thank my chief supervisor, Tim Edwards, for trusting me with such a significant part of his work. His patience, professionalism, and commitment to behavioural research have been outstanding. John Perrone and Clare Browne have added considerably to my understanding of my role as a researcher and have been generous with their time and expertise.

Many people were crucial to the support of our laboratory work; Michael Jameson, Cat Chang and Sandra Hopping each provided advice and practical assistance from the respiratory clinic at Waikato Hospital. Rob Bakker, Steve Newcombe, Merilyn Manley-Harris, Megan Grainger, Judith Burrows and Ivan Bell provided me with help and guidance above and beyond their roles at Waikato University.

The families who leant us their dogs became friends and cheerleaders for this process. Thanks to Jocelyne and Doug (Onyx), Barb and Al (Rylea), Richard and Sue (Bramble), Fiona and the team (Katie), Willem and Vicki (BJ), Philip and Margaret (Rocky) and Tim (Tui).

Thanks to Margaret Crawford Snr, Peter Crawford and Jane Crawford for their curiosity about the world, for being potent role models of life-long learning, and for their support of my education.

Dr David James Willcox provided the first prompt for this work. He has been a stalwart companion, advisor, interference runner and source of encouragement from the beginning. Sweetheart, I am more grateful than you know.

Table of Contents

List of Tables

List of Figures

Publications from this Thesis

Peer-reviewed journal articles reproduced in this thesis with permission from Elsevier Inc. and IOP Publishing.

Crawford, M. A., Perrone, J. A., Browne, C. M., Chang, C. L., Hopping, S., & Edwards, T. L. (2022). Transitioning from training to testing with scent detection animals: Application to lung cancer detection dogs. *Journal of Veterinary Behavior*, *55-56*, 23–34. https://doi.org/10.1016/j.jveb.2022.07.004

Crawford, M. A., Chang, C. L., Hopping, S., Browne, C. M., & Edwards, T. L. (2022). Influences of breath sample re-use on the accuracy of lung cancer detection dogs. *Journal of Breath Research*, *17*(1), 16001–. https://doi.org/10.1088/1752-7163/ac9b7f

Chapter One

General Introduction

Disease detection using animals

A proliferation in the training of animals to detect disease in humans and other animals has occurred in the last thirty years (Cambau & Poljak, 2020; Juge et al., 2022). Early case studies describe spontaneous disease detection by domestic dogs. In 1989 *The Lancet* published an article by Williams and Pembroke on the diagnosis and removal of a cancerous mole from a patient's leg. Prior to diagnosis, the patient's dog persistently licked and even attempted to bite the mole, prompting the owner to have it investigated. Further examples of the same phenomenon have since been documented (Campbell et al., 2013; Church & Williams, 2001; Pickel et al., 2004). Dogs' tendency to lick lesions and wounds extends to a persistent interest in diseased human tissue (Klaus & Joachim, 2017). While dogs have been used extensively as biological scent detectors, the abilities of other animals to discriminate disease in human and non-human animals are also being assessed (Edwards et al., 2017).

A review of recent literature in this field was conducted to provide an overview of the range of methodologies, targets, detection animals and outcomes. Table 1 was generated by completing a review of reviews from 2000 - 2022, searching SCOPUS (59 results, 1 November 2022), Web of Science (301 results, 7 November 2022), and PubMed (150 results, 7 November 2022) using the following search terms: (animal or dog or canine or rat or nematode or bee or moth or wasp) and (cancer or disease) and (odor or odour or olfact* or scent or smell) and detect* and review. A manual evaluation of goodness of fit for the review parameters of animal detection of any disease, including non-human animal diseases, was completed. A total of 73 unique articles were identified in the 18 review articles that met the inclusion criteria for review.

Table 1

Animal and Human Disease Scent Detection via Detector Organism Review

Year of Pub.	Journal Article Authors	Disease	Sample Type	Number of Unique Samples	Species Trained, Number Trained	Sensitivity (as reported)	Specificity (as reported)
2004	Willis	Bladder cancer	Urine	n = 144	Dog, n = 6	Mean success rate: 41%	ND (Not determined)
2004	Pickel	Melanoma	Tumour	Unclear	Dog, n = 2	100%	100%
2006	McCulloch	Lung cancer & breast cancer	Breath	n = 169	Dog, n = 5	*Lung cancer:* 99% *Breast cancer:* 88%	Lung cancer: 88% Breast cancer: 98%
2008	Gordon	Breast cancer, prostate cancer	Urine	n = 493	Dog, *Breast cancer* n = 6, *Prostate cancer* n = 4	"No better than chance"	ND
2008	Horvath	Ovarian cancer	Tumour tissue	Unclear	Dog, n = 1	100%	98%
2008	Richards	Nematode infection (sheep)	Faeces	Unclear	Dog, n = 2	91%	ND
2010	Horvath	Ovarian cancer	Tumour tissue & blood	Unclear	Dog, n = 1	*Tumour:* 100%, *Blood:* 100%	*Tumour: 95%*, *Blood: 98%*
2010	Poling	Tuberculosis	Sputum	n = 10523	Rat, n = 20	82% - 91%	89% - 95%
2011	Mahoney	Tuberculosis	Sputum	n = 12329	Rat, n = 10	96%	74%
2011	Cornu	Prostate cancer	Urine	n = 108	Dog, n = 1	91%	91%
2011	Sonoda	Colorectal cancer	Breath & faecal samples	Unclear	Dog, n = 1	*Breath*: 91%, *Faeces:* 97%	*Breath*: 99%, *Faeces:* 99%
2011	Willis	Bladder cancer	Urine	n = 210	Dog, n = 4	64%	56% - 92%
2011	Suckling	Tuberculosis	Chemicals	n = 3	Honey bee, Unclear	Unclear	Unclear
2012	Ehmann	Lung cancer	Breath	n = 220	Dog, n = 4	71%	93%
2012	Buszewski	Lung cancer	Breath	n = 73	Dog, Unclear	82%	82%
2012	Walczak	Breast cancer, melanoma, lung cancer	Breath	n = 627	Dog, n = 6	79%	78%
2012	Alasaad	Sarcoptic mange (mammals)	Infected carcasses	Unclear	Dog, n = 2	100%	ND

Year of Pub.	Journal Article Authors	Disease	Sample Type	Number of Unique Samples	Species Trained, Number Trained	Sensitivity (as reported)	Specificity (as reported)
2012	Mgode	Tuberculosis	Sputum	n = 289	Rat, Unclear	80%	72%
2012	Bomers	*Clostridium difficile*	Stool samples and patients	n = 300	Dog, n = 1	83%	98%
2013	Dehlinger	Diabetic hypoglycaemia	Skin swab	n = 24	Dog, n = 3	56%	53%
2013	Horvath	Ovarian cancer	Blood Plasma	n = 252	Dog, n = 2	97%	99%
2013	Mahoney	Tuberculosis	Sputum, slides and pots	n = 12	Rat, n = 12	*Slides* = 71%, *Pots* = 82%	Slides = 86%, Pots = 96%
2014	Amundsen	Lung cancer	Trained with tumours. Tested with breath & urine	n = 93	Dog, n = 4	*Breath:* 56%, *Urine: 64%*	*Breath: 33.3%*, *Urine: 29.2%*
2014	Bomers	*Clostridium difficile*	Bacterial culture, faecal sample & patient	n = 371	Dog, n = 1	86%	97%
2014	Elliker	Prostate cancer	Urine	n = 117	Dog, n = 10	17%	73%
2014	Rudnika	Lung Cancer	Breath	n = 253	Dog, n = 2	86%	72%
2015	Hardin	Diabetic hypoglycaemia	Sweat	n = 4	Dog, n = 6	78%	96%
2015	Rodionova	Liver cancer	Faeces & urine	Unclear	Dog, n = 5	90%	99%
2015	Schallschmidt	Lung cancer	Headspace gas of cell culture	Unclear	Dog, n = 2	10%-20%	40%-50%
2015	Reither	Tuberculosis	Sputum	n = 469	Rat, n = 7	59% - 72%	18% - 98%
2015	Hirotsu	Cancer: oesophageal, gastric, colorectal, breast, pancreatic, bile duct, prostate	Cell culture, human tissue, urine, blood serum	Unclear	Nematode, Unclear	96%	95%
2015	Urbanova	Prostate cancer	Urine	Unclear	Dog, n = 1	94%	92%
2016	Angle	Bovine viral diarrhoea virus (cattle)	Viral culture	n = 15	Dog, n = 2	91%	99%

Year of Pub.	Journal Article Authors	Disease	Sample Type	Number of Unique Samples	Species Trained, Number Trained	Sensitivity (as reported)	Specificity (as reported)
2016	Hackner	Lung cancer	Breath	n = 122	Dog, n = 6	79%	34%
2016	Maurer	Bacteriuria	Urine	n = 687	Dog, n = 5	Near 100%	Above 90%
2016	Taverna	Prostate cancer	Urine	n = 114	Dog, n = 2	78%	N/A
2016	Willis	Melanoma	Tissue	n = 741	Dog, n = 2	45%	ND
2016	Neto	Tuberculosis	Chemicals	n = 8	Nematode, Unclear	Unclear	Unclear
2017	Bryce	*Clostridium difficile*	Faecal sample & bacterial culture	Unclear	Dog, n = 1	92%	95%
2017	Dorman	Urinary tract cancer (dog)	Urine	n = 21	Dog, n = 4	Hovered near chance'	ND
2017	Guerrero-Flores	Cervical cancer	Cervical smear & absorbent material	n = 270	Dog, n = 1	*Scrapes:* 93%, *Surgical bandages: 96%*	*Scrapes: 99%, Surgical bandages: 100%*
2017	Guirao Montes	Lung cancer	Breath	n = 113	Dog, n = 1	95%	98%
2017	Kitiyakara	Liver cancer	Breath	Unclear	Dog, n = 1	78%	ND
2017	Koivusalo	*Staphylococcus aureus* (MSRA)	Culture	Unclear	Dog, Unclear	89%	91%
2017	Koskinen	Bacteriuria	Urine	n = 91	Dog, n = 1	24/27 accuracy	ND
2017	Los	Diabetic hypoglycaemia	Body odour	n = 45	Dog, n = 8	36%	ND
2017	Gonder-Frederick	Diabetic hypoglycaemia	Body odour	n = 18	Dog, n = 18	57%	49%
2017	Ellis	Tuberculosis	Sputum	n = 11869	Rat, n = 22	94%	ND
2018	Fischer-Tenhagen	Lung cancer	Breath	n = 60	Dog, n = 2	Correct identification average: 95%	60%
2018	Fischer-Tenhagen	*Staphylococcus aureus* - mastitis (cattle)	Bacterial culture & milk	Unclear	Dog, n = 9	59%	93%
2018	Pacik	Prostate cancer	Urine & Sarcosine	n = 180	Dog, n = 1	90%	95%
2018	Seo	Breast cancer, colorectal cancer	Cell culture liquid	Unclear	Dog, n = 2	Greater than 90%	Greater than 90%

Year of Pub.	Journal Article Authors	Disease	Sample Type	Number of Unique Samples	Species Trained, Number Trained	Sensitivity (as reported)	Specificity (as reported)
2018	Taylor	Toxigenic *Clostridium difficile*	Faecal sample & bacterial culture	Unclear	Dog, n = 2	85%	85%
2018	Mgode	Tuberculosis	Sputum	n = 55148	Rat, Unclear	Mean increase of 43% from DOTS	ND
2019	Catala	Epileptic seizures	Breath & sweat	n = 5	Dog, n = 5	87%	98%
2019	Davies	*Pseudomonas aeruginosa*	Bacterial culture	Unclear	Dog, n = 4	*Mixed Multi-Organism Cultures*: 87%	*Mixed Multi-Organism Cultures*: 84%
2019	Guirao	Lung cancer (SPN)	Breath	n = 109	Dog, n = 1	97%	99%
2019	Junqueira	Lung cancer	Blood serum	Unclear	Dog, n = 4	97%	98%
2019	Murarka	Ovarian cancer	Cell culture & plasma	Unclear	Dog, n = 4	Unclear	Unclear
2019	Thuleau	Breast cancer	Skin secretion	Unclear	Dog, n = 2	90%	N/A
2019	Guest	Malaria	Body odour (Socks)	n = 175	Dog, n = 2	72%	91%
2019	Rooney	Diabetic hypoglycaemia	Breath and sweat	n = 27	Dog, n = 27	Median sensitivity to out-of-range episodes: 70%	ND
2019	Charles	*Clostridium difficile*	Faecal sample & bacterial culture	n = 264	Dog, n = 2	91%	98%
2020	Grandjean	COVID-19	Sweat	n = 177	Dog, n = 1	76% - 100%	ND
2020	Jendrny	COVID-19	Saliva & mucus	Unclear	Dog, n = 8	83%	97%
2020	Mazzola	Lung cancer	Urine	n = 334	Dog, n = 6	45% - 73%	89% - 91%
2020	Reeve	Diabetic hypoglycaemia	Breath	n = 30	Dog, n = 2	0% - 87%	86% - 100%
2020	Yamamoto	Cervical cancer	Urine	n = 195	Dog, n = 1	100%	100%
2020	Schoon	Colon cancer	Faeces	n = 70	Dog, n = 2	84%	88%

Year of Pub.	Journal Article Authors	Disease	Sample Type	Number of Unique Samples	Species Trained, Number Trained	Sensitivity (as reported)	Specificity (as reported)
2021	Eskandara	COVID-19	Saliva & mucus, face masks & clothes	Pharyngeal Secretion: n = 80, fabric: n = 120	Dog, n = 6	*Pharyngeal Secretion:* 65%, *fabric:* 86%	*Pharyngeal Secretion:* 89%, *fabric:* 93%
2021	Guest	Prostate cancer	Urine	n = 50	Dog, n = 2	71%	70 - 76%
2021	Maa	Epileptic seizures	Sweat	n = 60	Dog, n = 13	Probability of diff. ictal versus interictal sweat 93%	Probability of canine detection of seizure scent preceding clinical seizure 82%
2021	Mendel	COVID-19	Breath	Unclear	Dog, n = 4	Average accuracy: 97.5% Positive predictive value: 91.5%	ND

Table content summary

Details of the specific target diseases, sample types, the numbers of unique sample types available, individual animals trained, testing criteria and accuracy measures will be described in the following section.

Target diseases

Twenty-seven diseases including epilepsy (Fisher et al., 2014) were identified as targets for animal scent detection. The most frequently researched human diseases were lung cancer (18%; 12/68), followed by tuberculosis (13%; 9/68) and prostate cancer (12%; 8/68). The range of the remaining human disease detection projects included the following targets: bacteriuria; bile duct cancer; bladder cancer; breast cancer; cervical cancer; *Clostridium difficile*; colorectal cancer; COVID-19; diabetic hypoglycaemia; epileptic seizures; gastric cancer; liver cancer; malaria; melanoma; oesophageal cancer; ovarian cancer; *Pseudomonas aeruginosa*; pancreatic cancer and *Staphylococcus aureus*. Non-human animal disease targets

included: in cattle, bovine viral diarrhoea virus (20%; 1/5), and *Staphylococcus aureus* (20%; 1/5); in sheep, nematode infestation (20%; 1/5); in mammals, sarcoptic mange (20%; 1/5); and in dogs, urinary tract cancer (20%; 1/5).

Sample types

Thirty-two sample types were collected to support the research projects, of which the most frequently occurring were: urine (23%; 17/73); breath (22%; 16/73); faeces (12%; 9/73); bacterial cultures (7/73; 10%) and sputum samples (7/73; 10%). Other sample types included: absorbent material (vaginal discharge); blood; blood plasma; blood serum; body odour; cell culture; cell culture liquid; cervical smears; chemicals; clothes; face masks; human tissue; headspace gas of cell culture; infected carcasses; milk; mucus; plasma; saliva; sarcosine; skin secretions; skin swabs; sputum pots; sputum slides; sweat; the physical presence of the patient; tumour tissue; and viral culture.

Unique samples

A unique sample is the first sample of a particular type provided by a single participant. It remains unique until it has been used for the first time to either train or test a detector organism. For example, a participant could provide several breath samples during a single collection session. However, only one of those – the first sample made available in the first instance to animals for training or testing purposes – could be described as unique. If a range of sample types is collected from a single participant, the first sample used of each type (breath, blood, urine etc.) would be considered unique. The number of unique samples in respective projects ranged from 3 to 55,148, with a median value of 121. Of the unique sample values, 30% (22/73) were unclear. 18% (13/73) of the unique sample values were 50 or lower.

Detector species

Four species were trained to detect disease: dogs (86%; 63/73), rats (*Cricetomys gabianus)* (10%; 7/73), nematodes (*Caenorhabditis elegans*) (3%; 2/73), and honey bees (*Apis mellifera*) (1%; 1/73).

The number of individual animals trained ranged from 1 to 27. Of the projects, 10% (7/73) had unclear numbers of trained animals, 47% (34/73) of projects had either 1 or 2 trained individuals, while 74% (54/73) had trained 1 to 6 individuals.

Measures of performance accuracy

The two measures of performance accuracy reported most frequently in the projects were sensitivity and specificity. Sensitivity and specificity measures are integral to signal detection, which is a general psychophysical approach to measuring decision-making under conditions of uncertainty, based on sensory input (Macmillan & Creelman, 2004). Sensitivity is calculated by dividing the number of correct positive indications of positive samples by the number of known positive samples. Specificity is calculated by dividing the number of correct rejections of negative samples by the number of known negative samples.

Sensitivity reported across the studies ranged between 0% and 100%. Three results were reported as unclear. Other terms used instead of sensitivity included: mean success rate; hovering near chance; no better than chance; 24/27 accuracy; correct identification average; and mean increase from conventional screening for tuberculosis at direct observed treatment, short course (DOTS) clinics. The mean of available sensitivity scores was 80%. Perfect sensitivity scores (100%) were recorded in 7% (5/73) of projects, while a further 10% (7/73) of recorded scores were no better than chance (0% - 50%).

Specificity reported across the studies ranged between 18% and 100%. 26% (19/73) of the projects did not calculate specificity values. Perfect correct rejection scores (100%) were recorded for 4% (3/73) of projects, while 7% (5/73) were no better than chance.

Factors influencing the outcomes of disease scent detection

There is evidence that animals can detect disease at above chance levels: the mean reported sensitivity of all projects described above was 80%. However, the extent to which individual animals might be expected to detect disease remains unclear, and is heavily dependent on the methodology applied to their training and testing, as well as the disease itself (Edwards et al., 2017; Pirrone & Albertini, 2017). Research methods likely to skew results include but are not limited to: forced choice line ups that introduce artificial ceiling and base accuracy scores (Edwards et al., 2017); the use of arbitrary accuracy cut-offs (Maughan et al., 2022); maintaining specific ratios of positive to negative samples throughout training (Gadbois & Reeve, 2016); measuring detection accuracy during training with samples familiar to the animals; and testing animals using different sample types than those they were trained with.

Features controlling the behaviour of the individual detector organism are also likely to vary from trial to trial. As is apparent in Table 1, broadly differing detection results have been recorded within and across groups of species, disease, and sample type. The state of the detecting organism itself influences accuracy. Factors influencing individual accuracy include an organism's learning history (rewards or punishment associated with performance) (Staddon & Cerutti, 2003), general health (Troisi et al., 2019), environmental variables like heat or cold; social stressors (Byosiere et al., 2019), levels of rest, hunger, and fatigue (Hayes et al., 2018), preferences and biases (Johnen, 2017); and the rate and frequency with which the task has been performed already (Hall, 2017).

Concurrently, external factors acting upon detecting organisms include: training and testing methods (Keep et al., 2021); sample type (Edwards et al., 2017); sample collection (Guest et al., 2020; Walczak et al., 2012); sample storage and re-use (Crawford et al., 2023;

van der Schee et al., 2012); and training-to-testing terminal criteria (Crawford et al., 2022; Gallistel et al., 2004).

The influence of training conditions on detection outcomes

Training involves the differential reinforcement of an organism's target and non-target behaviours (Cooper et al., 2020). The conditions of training can have a significant impact on an organism's ability to accurately respond to the scent of a target disease. Specifically, differentiation between chemotaxis and conditioned responses; recognising the role of sample presentation variables; training either forced choice or yes/no responding; and managing the presence of other organisms (including trainers and researchers) each has the potential to impact scent detection accuracy.

Nematodes respond to diseased tissue via chemotaxis rather than training (Hirotsu et al., 2015; Neto et al., 2016). Respondent conditioning pairing extension of the proboscis with the presence of a target scent was used to establish scent detection behaviour in bees (Suckling & Sagar, 2011). In all other cases in the reviewed literature, operant conditioning is used to train rats and dogs to systematically respond to human or animal disease targets (Ellis et al., 2019; Hayes et al., 2018).

Training conditions need to reflect test conditions as closely as possible for optimal detection outcomes. Specifically, they need to reflect the random frequency with which target samples occur in test sample arrays (Edwards et al., 2017; Maughan et al., 2022). This involves designing sample presentation to the detector oganism in which the separate evaluation of each sample occurs, irrespective of the status of the other samples. By contrast, many of the reviewed projects developed training session schedules with repeated set ratios of positive to negative samples (forced choice procedures). Forced choice procedures can impact the accuracy values achievable by individual organisms. Forced choice responding to samples during both training and testing places artificial ceilings and floors on the range of

accuracy values generated by organisms (Edwards et al., 2017; Gadbois & Reeve, 2016). By contrast, independent evaluation of each sample using "yes" and "no" or "go/no go" responses allow robust sensitivity and specificity measures to be collected (Macmillan & Creelman, 2004).

Disease scent detection training methods overwhelmingly rely on the presence of a human trainer during sessions. Sometimes the presence of the trainer is maintained into testing. There is potential for the detector organism's responding to be controlled by social cues under these circumstances, as many animals are highly sensitive to non-verbal cues (Krause et al., 2018). While a trainer is present in the training area, it is possible for changes in breath, movement, posture, or pace, to cue the indication of a detection animal (Reid, 2009). These changes are likely to be entirely involuntary on the part of the trainer, but may have a significant effect on detection outcomes. Examples of the impact of involuntary cueing on test accuracy are described in the broadly circulated early twentieth century report of "Clever Hans" (Samhita & Gross, 2013) and in more recent studies (Hayes et al., 2018; Lit et al., 2011). Automated testing, in which the animal is trained to respond to samples in the absence of a trainer, eliminates the possibility of trainer cueing.

Testing

Four reviewed projects could be categorised as self-report surveys rather than tests, as conditions during data collection were not controlled. For example, in one project diabetic owners were asked to report after the fact whether or not their trained hypoglycaemia detection dogs alerted prior to a spike in the owner's blood sugar. In the remaining 69 projects, precautions were taken during testing to blind detection animal handlers to the position of positive samples.

Sample collection and storage

Scent detection requires an organism to differentiate between the environmental "noise" and the accompanying target scent "signal + noise" in a sample (Hautus, 2015). The extent to which the "noise" – the non-target ambient scents - in sample collection environments can be controlled, varies. During testing, scent detection animals are exposed to a range of ambient smells collected at the same time as the substance under review. For example, a novel breath sample is also likely to contain traces of the room in which the sample is collected, as well as the participant's most recent meal; any medication the participant is currently taking; environmental contaminants like smoke; and/or chemical markers of metabolic change associated with the participant's age and gender (Pennazza & Santonico, 2018). In some instances, the detection animal has been trained using samples of tumours or an artificial scent profile made up of high frequency volatile organic chemical (VOC) lung cancer markers, but then tested with breath samples. In these circumstances, the animal's detection task could require a level of target concept generalisation quite different than the task of an animal trained with relatively "noisy" novel breath samples (Schallschmidt et al., 2015).

All tissue types and waste materials intended for presentation to detector organisms need to be collected and stored using methods that maintain the integrity of the sample and avoid cross-contamination or background cueing. Diverse materials and procedures have been used in disease scent detection to date, with different emphases on preservation. For instance, both glass and plastic sample containers appear in different projects. A sample storage material like glass is non-permeable (Goss, 2019). However, glass is also relatively fragile, expensive, and bulky to store and handle. By contrast, commercially available self-sealing food-grade plastic bags are ubiquitous and relatively cheap. Unfortunately, these same bags are also highly permeable and easily punctured (Torri & Piochi, 2016). Therefore,

the function and efficacy of either glass or plastic storage items is likely to be quite different, depending on the ways in which they are used. Another sample preservation factor represented variably in the literature is temperature. The temperature at which a sample is stored will influence its stability over time (Harshman et al., 2016). Unrefrigerated samples with high moisture content and organic matter are likely to be subject to decay (Hudson et al., 2009), precipitated by any bacteria or mold present in the environment at collection. VOCs in samples provide the distinctive disease scent profile target for detection animals (Leitch et al., 2013). VOCs also dissipate at room temperature. On these bases, it could be argued that controlling both sample storage permeability and temperature is critical to maintain the integrity of training and testing samples.

Terminal criteria

Once as many features as possible of a research project have been optimized, and training has been successfully undertaken, the project requires a transition from training to testing the detection animals. Intermittent training to maintain detection behaviors occurs after this transition, but this is only auxiliary to the main focus of final blind testing. Varying accuracy rates and measures were used to determine the transition from training to testing in the reviewed research. The basis for the transition was rarely acknowledged or explored beyond the provision of an arbitrary criterion; often 100% accurate performance during multiple consecutive sessions. Theoretical considerations need to be introduced at this point to more meaningfully describe the extent to which a detector organism has learnt a disease target.

Disease detection theory

Concept formation

Concept formation in human and non-human animals has been extensively researched (Bhatt et al., 1988; Herrnstein, 1990; Keller et al., 1951; Zentall et al., 2002). It is therefore

surprising to discover that there are no universally recognised terminal criteria used to determine when to stop training and start testing the acquisition of new concepts.

In early concept formation experiments, pigeons (*Columba livia domestica)* were trained to perform discriminations between the presence or absence of humans (Herrnstein & Loveland, 1964) or trees (Herrnstein et al., 1976) in photographic images. Through multiple exemplar training, the pigeons' human or tree concepts generalized to unfamiliar images in which the target subjects were extremely varied in appearance, partially obscured, or presented with different colour filters, zoom apertures and/or in different contexts and seasons. 81 novel image exposures were presented during each training session. Using this schedule, the pigeons' accurate responding became apparent within 7 – 10 sessions, and then continued to improve over the following months. As training continued, the pigeons demonstrated increasingly complex discrimination between images that were not even immediately identifiable through casual human observation.

In the context of disease scent detection training, operant conditioning using an antecedent stimulus (in this context, a sample) to elicit a specific response (an indication) from a detector organism in the presence of a target (a disease) is the foundation of concept formation (Herrnstein, 1990). The acquisition of a complex concept like a disease scent target requires the ability to distinguish samples characterized by two sets of features. The first is "category specific" features, which constitute several common elements of the concept that increase its probability of being categorized by the organism as a target. The second set involves "stimulus specific" features, in which only a few of the elements may be present, and the probability of the organism categorizing the sample as a target is low (Wasserman, 2016). Sample types (breath, saliva, sputum, urine etc.) are different representations of relevant stimuli – VOCs in the case of scent detection - that the organism has access to.

Samples used for training are often also re-used. It is reasonable to assume that samples degrade after repeated use (Wang et al., 2018). The pace and extent of the degradation may impact the fidelity of the target concept. Of equal concern is the possibility of contamination during repeated handling and exposure of samples. For example, VOCs exhaled from the bodies of detection organisms as they sniff samples may become part of the scent profile of the sample being examined. The extent to which this differentiates re-used samples from unused samples for the purposes of training has yet to be measured.

Target acquisition

Target acquisition – learning to discriminate accurately between samples that feature or lack properties of a complex concept – involves trial exposure to a range of exemplars. The rate and form of an organism's learning can be represented with a learning curve. The plot of a standard learning curve that shows cumulative performance over time is sigmoid in form (See Figure 1). Its features include: a mark of the moment at which behaviour begins to change (A); an inflection point representing the greatest change in the rate of learning (B); and a value for asymptote, which indicates where further improvement is extremely unlikely to occur (C). Investigation of the acquisition of new, simple information in non-human animal experiments suggests a much more abrupt form of acquisition at the level of individual responses. However, the aggregation of these responses over sessions, days and weeks tends to converge to show the characteristic curve described in early literature (Gallistel et al., 2004). Measuring the rate of acquisition of a complex target concept like the scent of a disease is both valuable and necessary to determine how long training should continue, and when testing should commence.

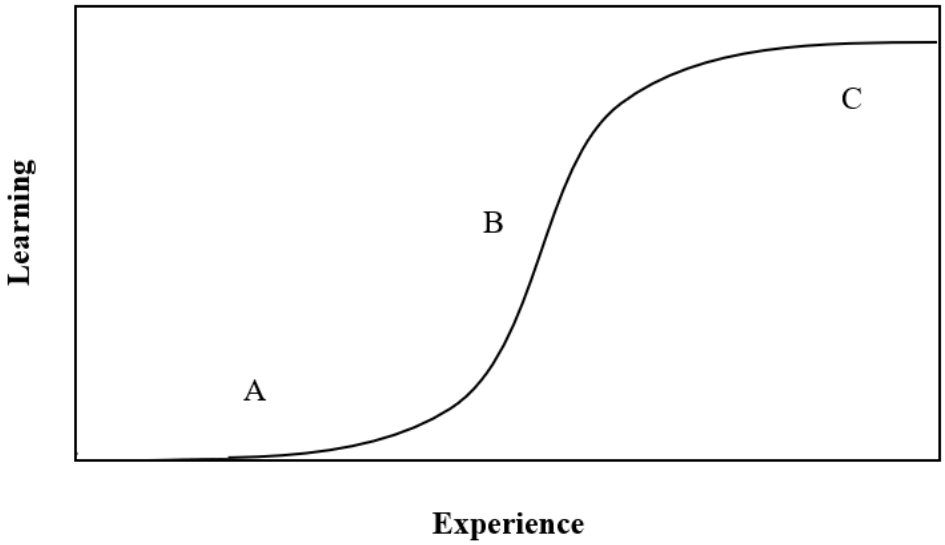
Current research targets/factors

Apparatus

As noted earlier, automation removes the possibility of cueing. It eliminates many

Figure 1

Standard Learning Curve (Based on Description in Murre, 2014)



avenues of human error and allows for very precise measurement of responding. We aimed to optimize disease scent detection in general by identifying procedures that could be applied across a number of contexts. Specifically, we used an automated approach to training and testing that could be standardized (Edwards, 2019). Using an automated training apparatus, we concurrently explored the role of sample re-use; which sample type was more effective; and the process of transitioning from training to testing.

1) Sample type

There has been extensive research on the biological suitability of dogs as scent detectors (Jendrny et al., 2021; Juge et al., 2022). However, there is little research on the comparative strengths and weaknesses of using different sample types during canine scent detection training (Edwards et al., 2017). Dogs have been trained to detect a single disease target across more than one sample type in 9 of the reviewed projects. However, a direct comparison between unique samples of different sample types from the same participants for disease detection purposes has not been completed at the time of writing.

Multiple sample material types are available for training and testing disease scent detection. Routinely eliminated waste like urine, faeces, sputum, breath, and sweat appear frequently in the research. Low-cost, disposable personal items like swabs and face masks are also readily available.

The most frequently used sample types in the reviewed research are urine, breath, and faeces. The collection of these materials is non-invasive. All three waste products are regularly available from participants. The collection of urine and faeces requires privacy, but is otherwise relatively straightforward. By contrast, providing breath samples can become effortful for participants with advanced pulmonary issues. Being coached through the process of providing samples by a technician is a way of mitigating this effort and controlling the uniformity of collection.

Urine can be aliquoted, making more training opportunities available by splitting the initial sample volume. In the same vein, multiple breath samples can be provided in a single session. Urine passes through the kidneys and contains particulates that can indicate several diseases (Ma et al., 2020). Likewise, breath contains particulates from the blood stream as it crosses the alveolar interface while passing through the lungs (Lourenço & Turner, 2014). Both urine and breath samples therefore provide a disease scent profile for training that is not limited to lesion sites. Faeces can include blood and by-products of infection and cell changes in the gut (Shirasu & Touhara, 2011). This is particularly pertinent during testing for bowel cancer (Bosch et al., 2019). Establishing a process for evaluating the relative merits of selecting different sample types for disease scent detection training has the potential to optimize decision-making about which sample types to collect in different contexts.

2) Frequency of sample use

Equally, there is a body of research on the physical properties of different organic sample storage and preservation materials and methods (Guest et al., 2020; Iyengar et al.,

2020). However, there is currently no research targeting the performance of biological detectors exposed to samples that have been used more than once.

In an ideal scenario, disease scent detection samples are so abundant they may be used once for training and then discarded if preferred. Wasserman (2016) notes that repetition of exemplars is not necessary for complex concept formation, and increasing the number of exemplars (samples) "slows acquisition but enhances generalisation" (p. 11). Due to the relative expense and effort associated with sample collection, however, a mixture of novel patient samples (individuals to whom the detection animal has not been exposed before) and re-used samples are likely be used for training purposes (Fischer-Tenhagen et al., 2018). Many research projects have incorporated the re-use of samples for training and testing, although the extent and frequency of sample re-use during training is often not described in detail (Buszewski et al., 2012; Fischer-Tenhagen et al., 2018; Jendrny et al., 2021; Juge et al., 2022; McCulloch et al, 2006; Rudnika et al., 2014; Walczak et al., 2012). Measuring and comparing the viability of used and re-used samples has the potential to support more cost-effective resource management during disease scent detection training.

3) Terminal criteria

Finally, each project described above features animals that have been trained and then tested as detector organisms. As was noted earlier, there are no universally recognised terminal criteria used to determine when to stop training and start testing the acquisition of target disease scents.

There is no consensus amongst the summarized studies about what constitutes the correct point at which a detecting organism is ready to focus primarily on testing unknown samples. Without a theoretical basis for determining when an animal has completed training, practical decisions about when begin testing and then only intermittently offer maintenance training may be founded on received wisdom. This has the potential to truncate the

achievement of the organism if training is completed too soon. Equally, it may also shorten the potential testing time available as a proportion of the animal's life span if training continues too long.

A frequently-used criterion like performance accuracy expressed as a percentage, is a problematic basis on which to transition from training to testing. Individual or aggregated accuracy measures doesn't necessarily take into account acquisition trends in the individual's learning (Baer & Parsonson, 1981). They are also unresponsive to variations in individual ability. Craft knowledge accrued by trainers working over several years may be referenced as a basis for decision-making (Hayes et al., 2018). This is difficult to quantify and inaccessible to less experienced researchers. A more transparent and reliable approach to terminal criteria for training is required for improved decision-making in this area.

Research target: Lung cancer

Our canine scent detection group selected lung cancer as a detection target. A small body of literature exists on this topic, so our methodological decisions could be informed by the experiences of previous researchers. There is evidence that dogs can detect lung cancer in human samples with greater than chance accuracy. Because of difficulties treating advanced lung cancer, progress in the area of early detection has the potential to immediately save lives.

Current lung cancer screening limitations

The World Health Organisation (2021) identifies cancer screening as an important but complex public health intervention, and limited attempts at screening have been offered in several countries. America and Canada have in the past offered targeted lung cancer screening (Jaine et al., 2018). A large-scale lung cancer screening project involving both blood testing and Low Dose Computed Tomography (LDCT) for high-risk participants was begun in London in 2018, but was only exploratory in nature (National Institute of Health

Research, 2018). In America, annual LDCT is recommended for asymptomatic 55 to 80-year-olds who have maintained a heavy smoking habit within the last 15 years (Centre for Disease Control United States of America, 2019). However, risks associated with the LDCT technique in each context include exposing individuals to potentially carcinogenic radiation and false positive indications (Jaine et al., 2018). Currently, lung cancer screening is not available in most countries, including New Zealand (Jaine et al., 2018; National Institute of Health Research, 2018; Ostrin et al., 2020; Singh & Pazdur, 2021). Identifying non-invasive and cost-effective lung cancer screening processes has considerable merit in this context (Blandin et al., 2017; Kauczor et al., 2020).

Thesis structure

The following information outlines three concurrently run experiments designed to investigate the acquisition of a complex concept target lung cancer-positive scent by pet dogs, using and re-using breath and saliva samples from human participants. The timeline of experimental activity is shown in Table 2.

Experiment 1: Breath vs. saliva for lung cancer detection with dogs

Experiment 1 involved a comparison between the use of breath samples and saliva samples to train dogs to detect lung cancer. Seven pet dogs were trained using operant conditioning and an automated apparatus to indicate the scent of lung cancer positive samples collected from a local respiratory clinic. Dogs were selected based on their individual behaviours in the laboratory rather than their breed characteristics. Single subject research design allowed us to compare each dog's developing target acquisition against itself. Training featured no coercive elements; tasks ended at any stage that dogs showed signs of unease (e.g., yawning, panting, pacing). The extent to which the dogs formed a concept of the lung cancer scent target across two sample types could provide insights into the patterns and pace of acquisition not previously noted. Similar dedicated comparisons between sample

Table 2

Canine Scent Detection of Lung Cancer (Breath and Saliva) Project Timeline 2017-2020

Year	Month	Action
2017	November	Funding secured. Ethical approval gained. Liaison with respiratory clinic team underway. Recruitment of pet dogs begins.
	December	Laboratory closes for summer.
2018	January	Dog recruitment continues.
	February	Amyl acetate training of Onyx and Bramble begins (Monday and Tuesday AM)
	March	Amyl acetate training of Rylea and BJ begins (Monday and Tuesday PM).
	April	Amyl acetate training
	May	Tui and Katie (trained in previous project) resume amyl acetate training.
	June	Amyl acetate training
	July	Amyl acetate training
	August	Amyl acetate training
	September	Amyl acetate training
	October	Amyl acetate training
	November	Amyl acetate training
	December	Laboratory closes for summer.
2019	January	Laboratory remains closed.
	February	Laboratory re-opens, amyl acetate training resumes.
	March	Amyl acetate training
	April	Human saliva (Monday) and breath (Tuesday) target training begins (Rylea, Onyx, Bramble AM; BJ, Tui, Katie PM). Data for all three experiments is collected concurrently from April 2019 to November 2020.
	May	Human sample training

Year	Month	Action
	June	Human sample training
	July	Human sample training
	August	Amyl acetate training of Rocky begins.
	September	Human sample training
	October	Order of presentation changed to Breath (Monday), Saliva (Tuesday).
	November	Novel positive sample introduced, unreinforced for first two sessions.
	December	Laboratory closes for renovations.
2020	January	Laboratory remains closed.
	February	Training resumes. Performance falls away – all dogs. Rocky begins breath and saliva target training. Return to early acquisition indication thresholds for all dogs commences 11/02/2020.
	March	COVID-19 Lockdown
	April	COVID-19 Lockdown
	May	25/05/20 Laboratory re-opens.
	June	Saliva data collection ends. Novel positive sample, unreinforced for first two sessions, reintroduced to session arrays.
	July	Human sample training
	August	Human sample training
	September	Final available novel samples consumed.
	October	Human sample training
	November	Final training session 03/11/2020.

types did not exist at the time of writing. While breath samples were frequently used in the training of disease scent detection dogs to detect lung cancer, saliva samples were not. Both sample types seemed promising. Research into the use of saliva for disease detection purposes using scent-detection technology already exists (Ma et al., 2020; Qian et al., 2018). Saliva can be collected by the participant using a non-invasive procedure. The ability to aliquot saliva samples, creating several training samples from one initial donation, was also attractive. A direct comparison between the efficacy of breath samples and saliva samples for lung cancer scent detection training was intended to a provide a model for future comparisons between other sample types and disease categories and directly inform future lung cancer research.

Experiment 2: Influences of breath sample re-use on the accuracy of lung cancer detection dogs

Experiment 2 involved an analysis of the use and re-use of breath samples during training to indicate the extent of stimulus (target concept formation) generalisation across a range of sample states. During training, seven pet dogs were exposed to new and used breath samples to learn the lung cancer positive target concept. Samples were supplied by a local respiratory clinic and stored at -60°C or -80°C to retain as many VOCs as possible. Very little reference in the research is made to the re-use of disease scent detection samples (Fischer-Tenhagen et al., 2018). Changes in the quality and quantity of sample stimuli during repeated use might influence generalisation to novel samples, so too requiring investigation. Re-exposure of samples involves a number of co-occurring changes. Identifying processes for the use and re-use of resources for disease detection has broader applicability and implications beyond one disease target. Developing a systematic model for evaluating changes in the sensitivity and specificity of disease scent detection as samples are used and re-used would advance the systematic application of disease scent detection more generally.

Experiment 3: Transitioning from training to testing with scent detection animals: Application to lung cancer detection dogs

Experiment 3 was designed to optimize concept formation (target acquisition). Sensitivity and specificity measures of positive lung cancer scent detection by seven pet dogs during training were used as steps to explore the pace and extent of each dog's learning. We hoped to generate evidence of predictable patterns of acquisition that could be represented using quantitative measures, with a view to standardizing the transition from training to testing. We aimed in this way to contribute to the relatively new field of disease scent detection, with the potential for broader applications. Specifically, we reviewed different constructs of acquisition using learning theory. We drew on ideas about associative learning, informed by behavioural science. While the terms training and testing are used ubiquitously in the literature, target acquisition in the context of disease scent detection has not been explored in-depth. Broader questions about when acquisition is complete were raised. In light of these questions our research team identified a quantitative model that generated asymptotic values for dogs' mean sensitivity and specificity scores. This included the likely number of sessions required to attain stable asymptotic responding if this had not already occurred. Providing a quantitative representation of the process of target acquisition has the potential to strengthen the field of disease scent detection by providing a systematic approach to the crucial decision of when to stop training and start testing.

Chapter Two

Experiment 1

Breath vs. Saliva for Lung Cancer Detection with Dogs[1]

[1] Crawford, M.A., Chang, C.L., Jameson, M.B., Hopping, S., Browne., C.M., & Edwards, T.L. (2023). Breath vs. saliva for lung cancer detection with dogs. [Unpublished manuscript]. School of Psychology, University of Waikato.

Why compare sample types?

The evaluation of human products to determine which could be best for canine scent detection of lung cancer involves direct comparisons of multiple dogs' response sensitivity and specificity across sample types from the same participants. In practice, this involves a participant providing different sample types, dogs being trained to detect the target disease in each sample type, then testing the dogs' response sensitivity and specificity when exposed to the different sample types provided by further participants. Repeating this process with multiple participants provides data showing which sample type is best for target acquisition. No research on this specific topic had been published at the time of writing.

Breath vs. Saliva for Lung Cancer Detection with Dogs

Margaret A. Crawford[1], Catherina L. Chang[2], Clare M. Browne[3], Sandra Hopping[2], Michael B. Jameson[4], & Timothy L. Edwards[1]

[1]School of Psychology, University of Waikato, Hamilton, New Zealand

[2]Department of Respiratory Medicine, Waikato Hospital, Hamilton, New Zealand

[3]School of Science, University of Waikato, Hamilton, New Zealand

[4]Waikato Clinical Campus, University of Auckland, Hamilton, New Zealand

Author Note

Address correspondence to:
Timothy L. Edwards
University of Waikato
Private Bag 3105
Hamilton 3240
New Zealand

Abstract

Background While some research shows that dogs are able to detect lung cancer at above-chance levels using breath samples, the relative utility of other sample types has not been established. We evaluated the comparative utility of human breath and saliva samples for lung cancer detection with dogs.

Methods 7 dogs assessed breath and saliva samples from 490 patients attending a COPD clinic. Dogs were trained using an automated apparatus to identify samples from patients who were later diagnosed with lung cancer. Sensitivity and specificity measures were used to compare the dogs' performance with each sample type.

Results A mixed methods logistic analysis of accurate responding to breath and saliva samples showed significantly higher detection of lung cancer positive breath samples (*mean* = 0.78, 95% CI [0.70, 0.89]) than of lung cancer positive saliva samples (*mean* = 0.41, 95% CI [0.33, 0.49]), p <. 001. There were no significant differences in accurate classification of non-target breath samples (*mean* = 0.68, 95% CI [0.54, 0.82]) versus non-target saliva samples (*mean* = 0.68, 95% CI [0.54, 0.82]), *p* = 0.95.

Interpretation Higher sensitivity to breath samples compared to saliva samples suggests that breath samples have greater utility for canine scent detection of disease. Although these findings are supportive of the continued use of breath samples for VOC-based lung cancer detection, with methodological improvements, saliva samples may also have utility for this purpose.

Keywords: canine, diagnostic accuracy, lung cancer, olfaction, scent detection

Introduction

Lung cancer is the leading cause of cancer-related deaths globally (1). This is due, in part, to lung cancer typically being detected after it has already progressed to later stages. Early detection is an important determinant of treatment success (2-4), but recently implemented screening methods for early detection have been relatively costly, exploratory in nature, or only available to a small subset of the population (5-8). Results from analytical chemistry and scent-detection dog studies suggest that lung cancer is associated with changes in the volatile organic compound (VOC) profile of cancer victims (9-16). VOC analysis, either with machines or animals, may represent a fast, cost-effective, and non-invasive screening method for lung cancer. However, specific comparisons of the relative utility of different sample media (e.g., breath vs. saliva) for VOC-based lung cancer screening have not been conducted.

Studies examining detection of lung cancer using dogs have produced varied results, with sensitivity estimates ranging from 71-91%, and specificity from 40-99% (17-19). Diverse sample collection processes and training and testing methods have been used in the relevant studies (9-12, 20, 21). Most studies to date have used breath samples, which have been collected and stored using a range of media (see S1). At the same time, there is a growing body of research suggesting that saliva has unique biomarkers that can be used to detect the presence of cancer (22-25). Two saliva-based cancer detection research projects have used established microbiological enzyme-linked immunosorbent assay processes (26) and Raman spectroscopy (27) with promising results. The efficacy of saliva as a medium for disease detection with animals has not been explored, but there is some evidence that analysis of VOCs conveying metabolic characteristics of saliva could be used for diagnostic testing of lung cancer (28). Compared to breath samples, saliva has simpler handling and storage requirements and is easier to split into multiple samples, which is beneficial for training. The

objective of the present study was to evaluate the relative effectiveness of breath and saliva samples for canine detection of lung cancer following best-practice guidelines for scent-detection research (29-32).

Methods

Subjects

Pet dogs were recruited from the local community via word-of-mouth and social media in Hamilton, New Zealand. Dogs were selected for training on the basis of temperament, food drive, and independence. The University of Waikato Animal Ethics Committee approved all experimental procedures described herein (Protocol 1029).

Apparatus and Training Procedures

Each dog was trained to work independently in an experiment room, using an automated apparatus in the form of a carousel with 17 enclosed compartments. The front panel of the apparatus (1 m^2) included a central aperture through which the dog put its nose to sample the contents of the available chamber. An automated feeder was positioned on the ground approximately 1.5 m from the apparatus. A light and tone on the apparatus indicated to the dog that a sample was available. The dog would then approach the apparatus, put its nose through the aperture and sniff the sample in the chamber. The duration of sample sniffing, determined by breakage of an infrared beam across the sample aperture, was used to define a positive indication, with the required sniff duration adjusted as required. A positive indication in the presence of a positive sample triggered the release of up to three pieces of kibble from the food dispenser. If a negative sample was present, positive indications had no effect, and the dogs were required to activate an omnidirectional switch attached the edge of the front panel (34).

The dogs were initially trained to detect amyl acetate using the apparatus with the trainer in the room (S2). Once accurate amyl acetate positive and negative indication responses was established, the trainer removed themselves from the experiment room, but was able to view the dogs from the adjacent room via two cameras. The dogs were then further trained to discriminate between lung cancer positive and lung cancer negative human breath and saliva samples. Indication thresholds (the period of time each dog was required to remain with their nose in the port for a response to be treated as a positive indication) were individualised and systematically manipulated between 500 ms and up to 4500 ms, depending on the accuracy of each dog's responses. When a dog was performing with high sensitivity but low specificity, the threshold was systematically increased in 100 ms increments until specificity improved without significantly impacting sensitivity. Faster improvements in sample classification accuracy equated with faster increases in the indication duration threshold, so changes in the threshold value are indicative of the dog's rate of target acquisition.

The dogs trained for two days each week. On the first training day the dogs were exposed to 17 breath samples in randomised order. The following day, saliva samples from the same patients were presented to the dogs, also in randomised order. After 18 weeks of training in this sequence, breath and saliva sample presentation was reversed to evaluate possible order effects. Further methodological details are provided in S3.

Sample Information

Samples were provided by human participants recruited from the Chronic Obstructive Pulmonary Disease clinic of a local hospital. Human ethics approval was obtained from New Zealand's Health and Disability Ethics Committees (17/NTB/178/AM02). Each participant was undiagnosed at the point of recruitment. They were given the opportunity to contribute to

the project as they waited for their consultation appointment to commence. All samples were stored until diagnostic information became available.

Four breath samples were collected from each novel participant and sealed in glass tubes packed with polypropylene fibre and cotton fibre. In the same session, one saliva sample was collected using 10 ml of tap water to facilitate sample generation. This was expectorated into a polypropylene cup for later aliquoting. Samples were immediately stored at -80°C. After selection, individual saliva samples were thawed, aliquoted into 4 samples per participant then used for training or refrozen. Corresponding breath samples were thawed, opened, used for training, then resealed and also refrozen (see S3 for additional details).

Data Analysis

Sensitivity and specificity were calculated for each sample type for each dog. Sensitivity is the number of positive samples correctly indicated divided by the total number of positive samples. Specificity is the number of correct rejections of negative samples divided by the total number of negative samples. Log *d* and Log *B* measures were also calculated for each dog and sample type (see S4 for formulae). Higher log *d* values indicate a greater ability to distinguish between positive and negative samples – this is an index of successful target discrimination. Log *B* is an index of bias or preference in responding. A log *B* value above zero indicates a dog's tendency to classify samples as positive, regardless of their status. A log *B* value below zero indicates a dog's tendency to classify samples as negative, regardless of their status.

These calculations were performed on data from the dogs' first exposure to a novel sample from a novel patient, which is the best indicator of dogs' acquisition of the target concept (35). Data were binary, reflecting the dog's accuracy (sensitivity: 0 = miss, 1 = hit; specificity: 0 = false indication, 1 = correct rejection). Therefore, differences in accurate

responding to the two sample types for each of these measures were evaluated using a mixed methods linear analysis with a logit link function. Individual dog performance was treated as a random factor, with sample type as a fixed factor. A Wilcoxon signed-rank test was applied to the log *d* and log *B* measures derived from the dogs' responses to novel samples.

Additionally, because indication duration thresholds were adjusted according to the dogs' performance with individual samples, the mean indication threshold times for all dogs for saliva and breath samples were also calculated. This provided a useful metric of comparative target acquisition. If one set of sample type response durations increased beyond the other, this suggests that the dogs were acquiring that target more quickly.

A period of unusually accurate target acquisition occurred across all six dogs' responding in the first 15 training days. While this pattern of acquisition may reasonably be expected for simple target acquisition (37), research on complex target acquisition shows that acquisition requires exposure to hundreds of trials with hundreds of unique samples (36, 38). The uncharacteristic pace of lung cancer target acquisition suggested it was likely the dogs were in fact responding to a simple, unintended cue. Consequently, sample handing processes were thoroughly revised and the data generated during this period were eliminated from the analysis (See Crawford et al., 2022 for early acquisition data). The total remaining number of training days completed by the dogs was between 40 and 45 depending on their availability. One dog, Rocky, began training later than the others and after the process revision; his rate of learning was consistent with the anticipated pace of complex target acquisition. This suggests that the revision of sample handling processes had been successful.

Results

Of the 19 dogs recruited for the project, seven dogs were selected to participate in training (Table 1). 490 human participants provided samples for the training portion of the project.

Table 1

Participant Dog Profiles

Name	Sex*	Age at Recruitment	Breed
BJ	Female	10 years	Beagle
Bramble	Female	3 years	English Springer Spaniel
Katie	Female	9 years	Blue heeler X Border Collie
Onyx	Male	2 years	Labrador
Rocky	Male	4 years	Jack Russell X Beagle
Rylea	Female	6 years	Labrador
Tui	Male	3 years	Border Collie X Huntaway X Kelpie

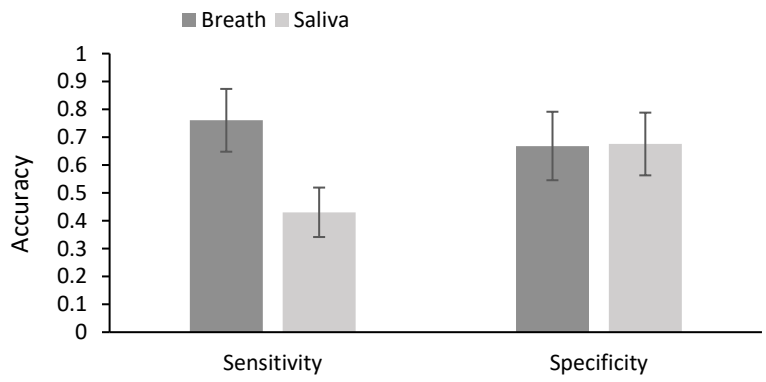
Note: dogs with bold names completed the study

*All dogs were neutered

Sensitivity associated with novel breath samples (*mean* = 0.78, 95% CI [0.70, 0.89]) was significantly higher than sensitivity scores for novel saliva samples (*mean* = 0.41, 95% CI [0.33, 0.49]), $p < .001$ (Figure 2).

Figure 2

Mean Accuracy for Breath vs Saliva Response



Responses to Novel Samples

Similar analysis of specificity data showed no significant difference between novel breath sample responses (*mean* = 0.68, 95% CI [0.54, 0.82]) and novel saliva samples responses (*mean* = 0.68, 95% CI [0.54, 0.82]), $p = 0.95$.

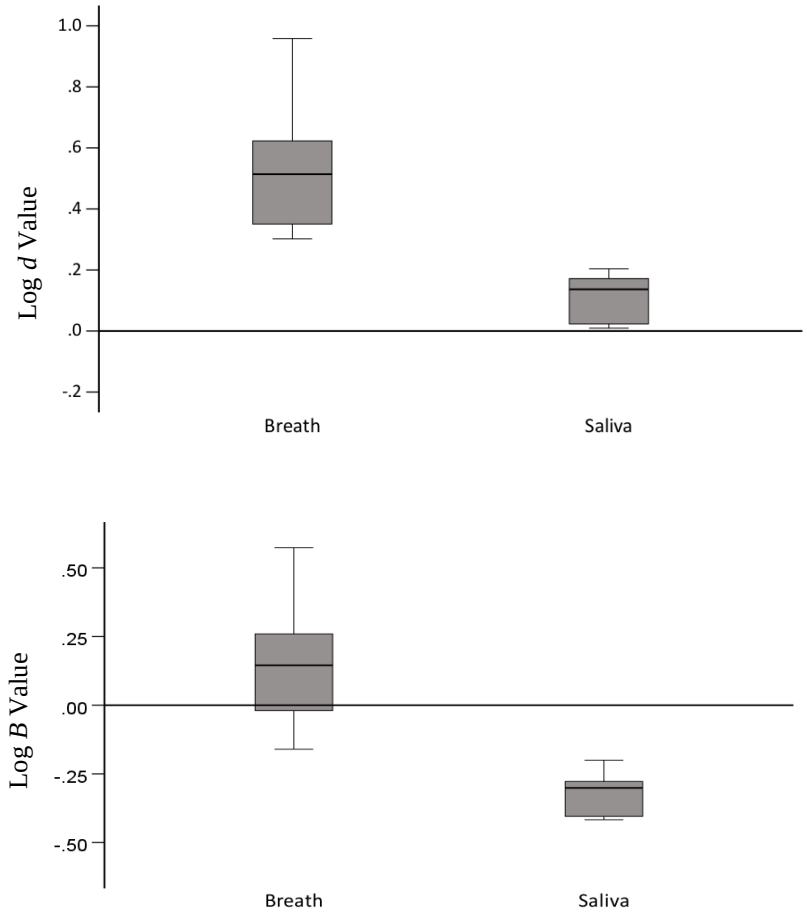
A Wilcoxon signed rank test was used to compare the differences in all dogs' log *d* measures of lung cancer detection using breath and saliva samples. The median value associated with breath sample responses was significantly higher than median saliva sample responses ($Z = -2.028$, $p = .043$). This indicates that the dogs were better able to identify the target scent in breath rather than saliva samples. Using the same analysis, differences in dogs' log *B* measures with breath and saliva samples were found to be statistically significant ($Z = -2.197$, $p = .028$) (Figure 3). The dogs' responses to saliva samples showed a conservative bias, while their responses to breath samples showed a liberal bias.

Indication Thresholds

The mean indication threshold for responses to all saliva samples during training was 2,313 ms. By comparison, the mean indication threshold for responses to all breath samples during training was 2,917 ms. This difference is indicative of the dogs' faster acquisition of the lung cancer target when evaluating breath samples. Figure 4 shows an example of the rate of target acquisition differentiation during 'retraining' after a two-month laboratory hiatus. Saliva sample training was stopped completely after the analysis of these specific data, as the contrast in the dogs' performance was interpreted as a strong indication that it would be most fruitful to work with breath samples going forward. Although the dogs performed better with breath samples in the current study, it is important to note that sensitivity to lung cancer-positive saliva samples increased 21% over the complete training period (see Figure 5).

Figure 3

Signal Detection Measures of Responding Accuracy, Breath vs Saliva



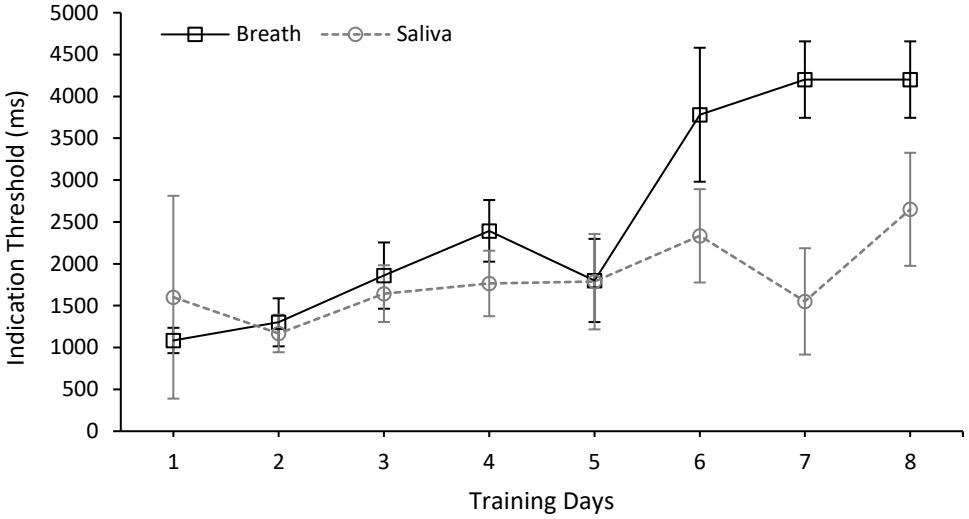
Note: log *d* and log *B* values associated with breath and saliva samples. Error bars indicate 95% confidence intervals.

Discussion

We compared dogs' lung cancer detection performance when using breath and saliva samples. The dogs performed with significantly higher sensitivity when working with breath compared to saliva. There were no differences in specificity between the two sample types. These findings suggest that breath samples may have more utility for VOC-based lung cancer detection than saliva.

Figure 4

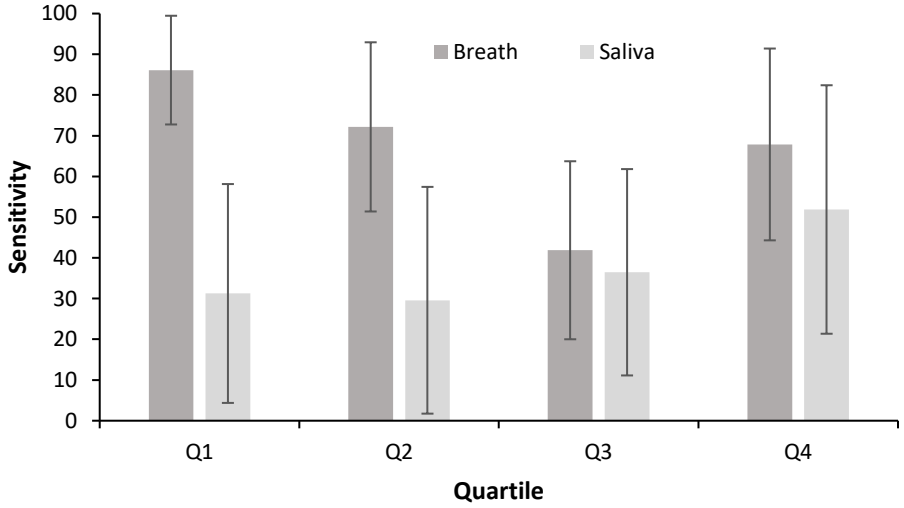
Mean Breath and Saliva Indication Thresholds



Note: Indication thresholds across 8 training days following laboratory closure and reopening. Error bars indicate 95% confidence intervals.

Figure 5

Mean Sensitivity to Breath and Saliva Samples (Quartile Analysis)



Note: Dogs' mean sensitivity associated with novel breath and saliva samples across quartiles of the research period. Error bars indicate 95% confidence intervals.

A clear differentiation between the dogs' performance with breath and saliva samples was also apparent in the lower mean indication threshold (the minimum response duration required for a response to be classified as a positive indication) associated with the dogs' training with saliva. The lower mean threshold value was a result of the dogs acquiring the lung cancer target more slowly when working with saliva samples because the indication threshold was only raised when dogs met performance criteria. Because the mean indication threshold was approximately 25% lower with saliva samples compared to breath samples, it was less effortful for the dogs to indicate saliva samples as positive. Despite the increased effort required to indicate breath samples as positive, dogs were still more likely to do so compared to saliva samples.

The discriminability index, log *d*, indicated that dogs could more readily distinguish between individuals with and without lung cancer using breath compared to saliva. The bias index, log *B*, indicated a liberal bias towards indicating breath samples as positive, and a conservative bias away from indicating saliva samples as positive. A likely explanation for the liberal bias with breath samples (and conversely, the conservative bias with saliva samples) is that dogs' positive indications were more often correct, and therefore reinforced, when working with breath samples. This, in turn, led to an increased tendency to indicate breath samples as positive more generally, including when evaluating samples with an unusual VOC profile (i.e., under uncertain conditions). Practically speaking, a liberal bias can be favourable under conditions in which sensitivity is of greater importance than specificity, which is often the case for technology positioned near the beginning of the diagnostic cascade. However, low specificity can be costly, as it necessitates additional diagnostics, and can cause unnecessary distress. If dogs were to be used in an operational screening scenario, one approach to potentially improving sensitivity and specificity and also allowing for adjustment of the sensitivity/specificity trade-off would be to use a group of dogs with group

decision criteria, rather than making decisions based off of any individual dog's data. Such an approach has been used successfully for tuberculosis screening using giant African pouched rats (39). Nevertheless, additional research is required to determine dogs' suitability for lung cancer screening under operational conditions before they can be seriously considered as a diagnostic tool for this purpose (40).

Regarding the differences in sample characteristics responsible for the discrepancy in performance associated with breath and saliva samples, it is reasonable to conclude that there were fewer relevant VOCs in the saliva (as analysed), or there was more "noise" in saliva samples (i.e., more irrelevant VOCs) which obscured the relevant VOCs, or both. The current consensus is that differences between the VOC profiles associated with lung-cancer-positive and negative samples, to the extent that they exist, are likely to be in the form of pattern differences involving a large number of VOCs (7). Therefore, comparisons of VOCs across different sample types via chemical analysis, while potentially useful, may not be straightforward to interpret. For example, without a clear indication of which VOCs are relevant and which are irrelevant, it would not be possible to determine the relative availability of such VOCs. Collaborations between scent-detection animal researchers and chemists working to determine the feasibility of VOC-based disease detection may be fruitful. For example, samples that are consistently correctly or incorrectly classified by animals could be analysed for clues about which compounds may be relied upon for accurate classification.

With some modifications to the sample collection process, saliva samples may still have a role to play in VOC-based lung cancer screening. Methods of increasing sample material (for example, by sampling multiple times) rather than diluting the sample to increase volume, could enhance performance. Increases in the surface area of the sample could also

facilitate performance in future research. With machine-based approaches to saliva analysis, the saliva itself can also be analysed, rather than the headspace VOCs alone (26, 27). Our findings are primarily applicable to headspace analysis of saliva samples and are not reflective of the utility of saliva for cancer screening more generally.

In conclusion, the present findings suggest that dogs respond more accurately to breath samples than saliva samples but that saliva samples can provide sufficient scent to produce better-than-chance responding. Based on these results, breath samples may be a better choice for training dogs to detect lung cancer, and potentially for VOC analysis using machines. Prolonged training with saliva increased sensitivity by 21%, suggesting that, with further refinement of the saliva sample collection process, saliva may be a viable substrate for volatile-based lung cancer detection.

References

1. World Health Organisation. Cancer 2021. [cited 2024, Apr 8]. Available from: https://www.who.int/news-room/fact-sheets/detail/cancer.

2. Vansteenkiste J, De Ruysscher D, Eberhardt WEE, Lim E, Senan S, Felip E, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. Ann. of Oncol. 2013;24(suppl_6):vi89-vi98.

3. Nardi-Agmon I, Peled N. Exhaled breath analysis for the early detection of lung cancer: recent developments and future prospects. Lung Cancer (Auckland, NZ). 2017;8:31-8.

4. Elisa D, Tommaso C, Emanuela F, Marco C, Marinos K, Giulia V, et al. Biomarkers and lung cancer early detection: state of the art. Cancers. 2021;13(15):3919.

5. National Institute of Health Research. Largest ever UK lung cancer screening study launches in London 2018 [cited 2024, Apr 8]. Available from: https://www.ucl.ac.uk/news/2018/dec/largest-ever-uk-lung-cancer-screening-study-aims-improve-early-detection

6. Centre for Disease Control, The United States of America. Lung Cancer [cited 2024, Apr 8]. Available from: https://www.cdc.gov/cancer/lung/basic_info/screening.htm

7. Ostrin EJ, Sidransky D, Spira A, Hanash SM. Biomarkers for lung cancer screening and detection. Cancer Epidemiol Biomarkers Prev. 2020;29(12):2411.

8. Singh H, Pazdur R. Contribution of early clinical benefit end points to decreased lung cancer mortality rates. JAMA Oncol. 2021;7(6):829-30.

9. McCulloch M, Jezierski T, Broffman M, Hubbard A, Turner K, Janecki T. Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. Integr Cancer Ther. 2006;5(1):30-9.

10. Buszewski B, Ligor T, Jezierski T, Wenda-Piesik A, Walczak M, Rudnicka J. Identification of volatile lung cancer markers by gas chromatography–mass spectrometry: comparison with discrimination by canines. Anal Bioanal Chem. 2012;404(1):141-6.

11. Walczak M, Jezierski T, Górecka-Bruzda A, Sobczyńska M, Ensminger J. Impact of individual training parameters and manner of taking breath odor samples on the reliability of canines as cancer screeners. J Vet Behav. 2012;7(5):283-94.

12. Rudnicka J, Walczak M, Kowalkowski T, Jezierski T, Buszewski B. Determination of volatile organic compounds as potential markers of lung cancer by gas chromatography–mass spectrometry versus trained dogs. Sens Actuators B Chem. 2014;202:615-21.

13. Fischer-Tenhagen C, Johnen D, Nehls I, Becker R. A proof of concept: are detection dogs a useful tool to verify potential biomarkers for lung cancer? Front. Vet. Sci. 2018;5:52.

14. Pennazza G, Santonico M. Breath analysis. San Diego: Elsevier Science & Technology; 2018.

Milanowski M, Pomastowski P, Ligor T, Buszewski B. Saliva - volatile biomarkers and profiles. Crit Rev Anal Chem. 2017;47(3):251-66.

15. Jendrny P, Twele F, Meller S, Osterhaus A, Shalke E, Volk HA. Canine olfactory detection and its relevance to medical detection. BMC Infect. Dis. 2021;21(1):838.

16. Juge AE, Foster MF, Daigle CL. Canine olfaction as a disease detection technology: A systematic review. Appl. Anim. Behav. Sci. 2022;253:105664.

17. Pirrone F, Albertini M. Olfactory detection of cancer by trained sniffer dogs: a systematic review of the literature. J Vet Behav. 2017;19:105-17.

18. Lazarowski L, Krichbaum S, DeGreeff LE, Simon A, Singletary M, Angle C, et al. Methodological considerations in canine olfactory detection research. Front. Vet. Sci. 2020;7:408.

19. Bauer P, Leemans M, Audureau E, Fromantin I. Remote medical scent detection of cancer and infectious diseases with dogs and rats: a systematic review. 2021;15347354221140516-15347354221140516.

20. Jezierski T, Walczak M, Ligor T, Rudnicka J, Buszewski B. Study of the art: canine olfaction used for cancer detection on the basis of breath odour. Perspectives and limitations. J Breath Res. 2015;9:027001.

21. Ehmann R, Boedeker E, Friedrich U, Sagert J, Dippon J, Friedel G, et al. Canine scent detection in the diagnosis of lung cancer: revisiting a puzzling phenomenon. Eur. Respir J. 2012;39(3):669.

22. Printz C. Saliva yields clues to early cancer detection. Cancer. 2012;118(13):3224-.

23. del Nogal Sánchez M, Hernández García E, Pérez Pavón JL, Moreno Cordero B. Fast analytical methodology based on mass spectrometry for the determination of volatile biomarkers in saliva. Anal Chem. 2012;84(1):379-85.

24. Zhang L, Xiao H, Zhou H, Santiago S, Lee J, Garon E, et al. Development of transcriptomic biomarker signature in human saliva to detect lung cancer. Cell. Mol. Life Sci. 2012;69(19):3341-50.

25. Katakura A, Yamamoto N, Sakuma T, Sugahara K, Onda T, Noguchi S, et al. A screening test for oral cancer using saliva samples: proteomic analysis of biomarkers in whole saliva. J. Oral Maxillofac Surg Med Pathol. 2015;27(1):1-5.

26. Chakraborty D, Viveka TS, Arvind K, Shyamsundar V, Kanchan M, Alex SA, et al. A facile gold nanoparticle–based ELISA system for detection of osteopontin in saliva: towards oral cancer diagnostics. Clin. Chim. Acta. 2018;477:166-72.

27. Qian K, Wang Y, Hua L, Chen A, Zhang Y. New method of lung cancer detection by saliva test using surface-enhanced Raman spectroscopy. Thorac. Cancer. 2018;9(11):1556-61.

28. Bel'skaya LV, Sarf EA, Kosenok VK, Gundyrev IA. Biochemical markers of saliva in lung cancer: diagnostic and prognostic perspectives. Diagnostics. 2020;10(4).

29. Schoon GAA, De Jonge D, Hilverink P. How dogs learn to detect colon cancer – optimizing the use of training aids. J. Vet. Behav. 2020;(35):38-44.

30. Edwards TL, Browne CM, Schoon A, Cox C, Poling A. Animal olfactory detection of human diseases: guidelines and systematic review. J Vet Behav. 2017;(20):59-73.

31. Ellis H, Mulder C, Vlverde E, Poling A, Edwards, T. Reproducibility of African giant pouched rats detecting Mycobacterium tuberculosis. BMC Infect. Dis. 2017;17(1):298.

32. Goss KU. The physical chemistry of odors – consequences for the work with detection dogs. Forensic Sci. Int. 2019;(296),110-114.

33. Edwards TL, Giezen C., & Browne CM. Influences of indication response requirement and target prevalence on dogs' performance in a scent-detection task. Appl. Anim. Behav. Sci. 2022;105657.

34. Edwards TL. Automated canine scent-detection apparatus: technical description and training outcomes. Chem. Senses. 2019;44(7):449-55.

35. Crawford MA, Perrone JA, Browne CM, Chang CL, Hopping S, Edwards TL. Transitioning from training to testing with scent detection animals: application to lung cancer detection dogs. J. Vet. Behav. 2022; 55-56:23-34.

36. Wasserman EA. Conceptualization in pigeons: the evolution of a paradigm. Behav. Processes. 2016;(123):4-14.

37. Gallistel CR, Fairhurst S, Balsam P. The learning curve: implications of a quantitative analysis. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(36):13124-13131.

38. Herrnstein RJ, Loveland DH. Complex visual concept in the pigeon. Science. 1964;146 (3643):549-550.

39. Edwards TL, Valverde E, Mulder C, Cox C, Poling A. Pouched rats as detectors of tuberculosis: comparison to concentrated smear microscopy. Eur. Respir. J.. 2016;48(2):579-82.

40. Hackner K, Errhald P, Mueller MR, Speiser M, Marzluf BA, Schulheim A, Schenk P, Bileck J, Doll T. Canine scent detection for the diagnosis of lung cancer in a screening-like situation. J. Breath Res. 2016;10(4):046003.

Chapter Three

Experiment 2

Influences of breath sample re-use on the accuracy of lung cancer detection dogs[2]

[2] Crawford, M. A., Chang, C. L., Hopping, S., Browne, C. M., & Edwards, T. L. (2022). Influences of breath sample re-use on the accuracy of lung cancer detection dogs. *Journal of Breath Research*, *17*(1), 16001–. https://doi.org/10.1088/1752-7163/ac9b7f

Learning via chemistry – concept formation and stimulus generalisation

The VOCs available in a breath sample have a high vapour pressure at room temperature (Williams & Koppmann, 2007). Therefore, the chemical profile of a breath sample used to train dogs disease scent detection could potentially change each time a sample is opened, each time a dog sniffs it, and each time it is stored at room temperature. The extent of the changes in chemical composition through re-use could be measured by organic chemists. At the same time, the significance of any changes in chemical profile can be evaluated during the training of scent detection dogs. Generalisation of the dogs' lung cancer-positive target concept is particularly pertinent to the re-use of breath samples during lung cancer detection training. Measuring the extent to which individual dogs are able to retain the trained target with fidelity when the scent profile changes over time gives us insight into the utility of re-using samples for this purpose. No research on this specific topic had been published at the time of writing.

Influences of breath sample re-use on the accuracy of lung cancer detection dogs

Margaret A. Crawford[1], Catherina L. Chang[2], Sandra Hopping[2], Clare M. Browne[3], and Timothy L. Edwards[1]

[1] School of Psychology, University of Waikato, Hamilton, New Zealand

[2] Department of Respiratory Medicine, Waikato Hospital, Hamilton, New Zealand

[3] School of Science, University of Waikato, Hamilton, New Zealand

Author Note

Address correspondence to:
Timothy L. Edwards
University of Waikato
Private Bag 3105
Hamilton 3240
New Zealand
edwards@waikato.ac.nz

Abstract

Evaluations of dogs as lung cancer detectors using breath samples have produced a variety of results, some quite promising. Breath samples are typically collected onto a substrate and stored in a sealed container when not in use, but volatile compounds dissipate when the substrate is exposed during training and evaluation sessions. Collection of appropriate samples for training and testing dogs requires significant resources and strict control of recruitment and sample collection processes. Therefore, some researchers re-use samples while training dogs. No systematic evaluation of the effect of sample re-use on dogs' training performance has been conducted, so the influence of this potentially important training factor is not known. We trained seven dogs to indicate the presence of lung cancer positive breath samples using an automated apparatus. The samples were stored at -60°C or -80°C. Samples from 460 individuals who were classified as positive or negative for lung cancer were used for training samples. Individual samples were presented to dogs up to four times over a period of two years. As sample re-use increased, sensitivity declined (-6.65, p = <.001, 95% CI [-10.56, -2.76]), specificity increased (2.87, p = .036, 95% CI [.19, 5.55]), and the dogs' bias shifted in the direction of a negative indication bias (-.094, p = <.001, 95% CI [-.149, -.39]). However, there were no significant changes in the measure associated with the detectability of the target (-0.30, p = .285, 95% CI [-.087, .26]). All observed changes in performance across sample re-use were small. Therefore, these findings suggest that sample re-use may be appropriate for training, but additional research is required to determine which factors underly changes in performance as breath samples are re-used.

Keywords: canine, diagnostic accuracy, olfaction, scent detection

Introduction

Worldwide, lung cancer is the number one cause of cancer-related mortality (1). Treatment outcomes improve significantly if lung cancer is detected in its early stages (2). However, there are few screening programs for lung cancer detection internationally, with most countries having no such program in place (3, 4). One of the barriers to implementation of lung cancer screening programs is the absence of inexpensive and non-invasive technology for lung cancer screening (5). Analysis of volatile organic compounds (VOCs[3]) in breath samples may represent one such alternative (7).

There is some evidence that dogs can be trained to detect lung cancer using human breath samples (8-13). It does not appear, however, that dogs can rely upon one or even several biomarkers to discriminate between cases and controls (14). A variety of chemical analyses have been conducted on the constituents of breath for the purposes of identifying lung cancer, but no reliable individual markers have yet been identified (2, 15, 16). Therefore, to the extent that dogs can accurately classify breath samples, it is likely that they achieve this by learning to identify complex and variable patterns that are indicative of lung cancer status (17). This type of learning has been studied under the "concept formation" moniker in other species, such as the pigeon (18, 19) and, more recently, has been studied and successfully applied with machine learning applications for classification purposes (20). There are many varieties of lung cancer and many other variables that can influence the VOC profile present in each breath sample, including smoking status, other underlying chronic diseases like Chronic Obstructive Pulmonary Disease, treatments such as the use of inhalers, and individual ventilation-related factors including low minute ventilation causing poor air mixing with individual breaths. Potential disease related variables include individual cancer cell lineage specific

[3] VOCs are organic compounds with high vapour pressure at 25°C (6).

characteristics, variables related to the cancer's physical characteristics, and variables associated with the impact of cancer on surrounding normal tissue. Consequently, the lung cancer VOC pattern (i.e., concept) is an exceedingly complex target. Many unique samples are typically required for concept formation training with dogs to succeed, with success being determined by accurate classification of novel exemplars.

Acquiring sufficient samples for training of dogs can be costly and challenging. Samples from cases and controls must be collected in the same location and in the same manner to avoid introduction of olfactory cues, and sample status must be known before samples can be used for training. Therefore, some researchers in this area rely upon a combination of novel samples and re-used samples when training dogs to identify the lung cancer concept (10). Although samples from multiple unique sources must be used for concept formation training to be successful (17), sample repetition is likely to increase learning effects to some extent (13). Re-use, however, may change a sample's chemical composition, depending on the environmental conditions in which it has been opened, stored, and reopened (21). The efficacy and limits of sample re-use have not been systematically measured.

1.1 Sample Re-use in Dog Training

Canine olfaction is highly sensitive, with dogs able to detect the presence of VOCs in volumes of 1 part per trillion (ppt) in some cases (22). The extent to which VOCs remain present in human breath samples at detectable levels for dog training after re-use and long-term storage is unclear. VOCs from breath sampling materials can dissipate quickly at room temperature (23), so it is important to determine how often samples can be re-used and still remain fit for purpose (10, 14). Samples in a training session may be composed of any combination of the following: (i) novel samples from novel sources (i.e., new samples from individuals to which the dogs have never been exposed); (ii) novel samples from old sources (i.e., new samples from individuals to which the dogs have previously been exposed); and (iii)

re-used samples from old sources (i.e., samples that have already been used in previous training sessions). Samples in this last category could include those that have been used repeatedly within the same day without being returned to storage and those which have been used, stored, and then presented for training again. In the present study, sample "use" is defined as removing a sample container from storage, opening the sample container, exposing the sample to dogs for training, closing the container, and returning it to storage. "Re-use" involves completing this sequence more than once.

Descriptions of training protocols, including practices related to sample re-use, are often missing in the relevant literature (see Table 4). For example, we were not able to determine whether samples were re-used during training by McCulloch et al. (11), Buszewski et al. (8), Walczak et al. (13), or Rudnicka et al. (12). McCulloch et al. (11) described the replacement of samples during training, while the extensive repetition of samples from old sources is implied. However, it is unclear whether any breath samples specifically from lung cancer cases or controls were re-used. Walczak et al. (13) measured and tabulated changes in the responses of dogs to samples encountered repeatedly over the course of a single day. However, it is unclear whether samples were re-used during training. Rudnicka et al. (12) used dogs that had been previously trained during research by Walczak et al. (13). The re-use of samples is explicitly addressed by Ehmann et al. (9) and Fischer-Tenhagen et al. (10). Ehmann et al. (9) did not re-use any samples for training or testing purposes. Fischer-Tenhagen et al. (10) noted that the number of samples was limited in their proof-of-concept work, so they re-used the same samples for multiple, but not a specified number of, training sessions.

Table 3

Sample Use and Accuracy Information from Canine Lung Cancer Detection Studies

Study	Re-used Samples	Sensitivity	Specificity
McCulloch et al. (2006)	Unclear	99%	99%
Buszewski et al. (2012)	Unclear	82%	82%
Ehmann et al. (2012)	No	71%	93%
Walczak et al. (2012)	Unclear	79%	78%
Rudnicka et al. (2014)	Unclear	86%	72%
Fischer-Tenhagen et al. (2018)	Yes	9/9, 8/9	8/10, 4/10

1.2 Breath Sample Collection and Storage

Exhaled breath is made up of vapor and condensates (24). When we breathe, endogenous breath-borne VOCs are released as gas by the lung alveoli (25). The VOCs cross the blood/breath barrier, reach equilibrium in the lungs, and are exhaled in the last 350 mL of a single breath (24). Other VOCs present in a breath sample can have a range of origins from prior exposure to exogenous VOCs; products from food; biota and metabolites in the bronchial tubes, mouth and nose; and the dead space containing unmixed air inhaled from the immediate environment (26-28). Breath samples containing all facets of an exhalation described above can be referred to as “mixed expiratory samples,” while “late expiratory” and “end-tidal” samples attempt to isolate alveolar breath (29).

Sample preservation is an important factor in breathomics research (27). Specifically, researchers in this field note that VOCs are only present in very small quantities and dissipate at room temperature; therefore, it is important to trap and store them on enclosed substrates or in a sealed container as part of any diagnostic process with high fidelity (21, 25, 30). The fidelity of breath sample storage is dictated by the sampling materials, storage temperature, storage duration, and the procedures used for collection (31). Low temperatures slow the

dissipation of VOCs (32-34). Hudson et al. (35) demonstrated that after an initial reduction in potency, breath samples stored in glass vials at low temperatures change less than samples stored at room temperature as time passes (in their study, up to 7 weeks), and are likely to retain "primary odor compounds" (36).

Breath samples for dog training purposes have been collected and stored using a range of components and materials. Components include the breath storage container, the container's closures, and the substrate stored inside the container. Ehmann et al. (9) and Fischer-Tenhagen et al. (10) used glass tubing. Glass tubing does not absorb VOCs and is impermeable at room temperature (35, 37). Polypropylene sampling tubes were used by McCulloch et al. (11), Buszewski et al. (8), Walczak et al. (13), and Rudnicka et al. (12). Polypropylene is a "rigid and crystalline thermoplastic" (38). Goss (31) asserts that plastics make problematic breath sample receptacles as they are scent-permeable and present a high risk of cross-contamination if samples are stored in close proximity. The materials used to close the breath tubes were rubber (9), polybutylene-terephthalate with silicone septa (10), or unspecified (8, 11, 12). Like plastics, rubber is scent-permeable (39). However, in Ehmann et al. (9), the thickness of the container plugs might contribute to a delay in permeation. Silicone rubber has varied degrees of permeability depending on its manufacturing process (40). By itself, silicone is a scent-absorbent polymer (31). Silicone is sufficiently scent-absorbent that four of the canine lung detection projects used polypropylene wool steeped in silicone oil to create both hydrophilic and hydrophobic qualities in the sample collection substrate (9-12). The substrate used by Buszewski et al. (8) was not specified. Walczak et al. (13) used a cotton substrate for their sample collection. Cotton fibers are naturally hydrophilic (41). A significant proportion of human breath is made up of water vapor, which is co-mingled in the exhaled breath vapor containing VOCs that is formed by cooling breath samples (24). Canine scent detection research projects have used breath samples under a variety of storage conditions for laboratory-

based training days, weeks, or even months, after initial collection (Fischer-Tenhagen et al., 2018; Schoon, 2005). For example, Schoon's (2005) forensic research results suggest that with sample repetition (in which several samples have been taken from a single subject, each for a single use) but no sample re-use, dogs can accurately indicate the presence of a target scent on crime scene objects up to 6 months old. The method of storing these samples was not recorded.

1.3 Sample Presentation

In all of the relevant research, to date, samples have been presented in "open" environments; no attempt was made to retain the sample odors as they equalized with the air in the training room (and other samples). Specifically, McCulloch et al. (11) placed opened sample tubes inside half-pint polypropylene containers with holes punched in the lids. Ehmann et al. (9) used sample training stations in which sample tubes were open and appeared to be mounted on metal pegs (the specific materials were not described). Walczak et al. (13) and Buszewski et al. (8) removed the substrates from tubes and placed them in sterile polypropylene boxes with pierced lids. Rudnicka et al. (12) followed the same process as Walczak et al. (13), but used glass rather than polypropylene boxes. Fischer-Tenhagen et al. (10) set each tube in a polyethylene cone with a perforated polyethylene plate placed over the tube mouth. Because Fischer-Tenhagen et al. (10) explicitly identified the re-use of samples, it is important to note the sample decontamination procedures used. If a dog's nose was seen to contact the perforated plate covering the tube mouth, the plate was changed. The plates were cleaned with a 90-s ultra-sonic bath, while the cones were wiped with a damp cloth (cleaning fluid unspecified).

1.4 The Present Study

Seven dogs were trained to use an automated apparatus to detect lung cancer VOCs using breath samples from human lung cancer cases and controls that were collected in a respiratory clinic. When not in use, samples were stored at -80 or -60°C to aid with preservation, as we

aimed to evaluate the effects of sample re-use under optimal conditions (i.e., when using ideal preservation methods between uses). Samples were re-used up to three times, allowing for analysis of the effects of re-use on the dogs' performance. We did not re-use samples more frequently than this because we did not know the effects of sample re-use prior to the completion of this study. We hypothesized that sample re-use would negatively impact the dogs' performance. We aimed to test this hypothesis and, if confirmed, to discover the extent and the specific nature of the impact of re-use on the dogs' scent detection performance with a view to informing future research involving breath sample re-use.

Method

2.1 Subjects

Pet dogs were recruited for this project via word-of-mouth and other channels, including social media. The dogs were evaluated with respect to their general suitability for laboratory research (e.g., levels of arousal in the laboratory in the presence of the researcher and other dogs); their motivation to work for food; and their responsiveness to initial training to use the automated scent-detection apparatus. Nineteen dogs were considered for participation, seven of which met the participation requirements (Table 5). The dogs worked for either two mornings or two afternoons each week.

2.2 Apparatus and Setting

The automated apparatus used to train and evaluate the dogs for lung cancer detection consisted of a carousel with 17 aluminum chambers sandwiched between two stainless steel circular plates, with a 1-m^2 acrylic front panel (see Figure 6). Each 3.57-L chamber had a hinged flap on the outer face with a stainless-steel L-bracket attached to the interior of the flap.

Table 4

Participant Dog Profiles

Name	Sex*	Age at Recruitment	Breed
BJ	Female	10 years	Beagle
Bramble	Female	3 years	English springer spaniel
Katie	Female	9 years	Blue heeler X border collie
Onyx	Male	2 years	Labrador retriever
Rocky	Male	4 years	Jack Russell X beagle
Rylea	Female	6 years	Labrador retriever
Tui	Male	3 years	Border collie X huntaway X Australian kelpie

*All dogs were neutered

This bracket limited the opening action on contact with the apparatus lid and added weight to the flap. Consequently, the flap only opened when pressure was applied by the dogs and closed when pressure was removed. One chamber at a time could be accessed by the dogs through a 10-cm-diameter sample port in the center of the front panel. A grid of infrared sensors was positioned behind the sample port to measure the duration of the dogs' access to each chamber. An omnidirectional switch, which was used by the dogs to advance the carousel to the next sample, was positioned 50 cm from the bottom of the apparatus on the right side of the front panel. The apparatus remotely controlled a kibble dispenser (Treat and Train Remote Dog Trainer manufactured by Pet Safe) situated 1.5 m away. The room housing the apparatus was 3.2 m by 4.3 m and was maintained at an ambient temperature of approximately 20°C. The apparatus was controlled by a computer in an adjacent room. After preliminary training, each dog worked in the room alone and was monitored by a researcher in the adjacent room watching

Figure 6

Apparatus Photographs

Note: Left: front panel of the apparatus with odor port and rotation switch (on right-hand side of front panel). Right: 17-segment carousel with lid removed. Note: dogs were only able to access the front panel of the apparatus.

a live feed from two cameras in the experimental room. See (43, 44) for additional information about the apparatus, standard operating procedures, and performance obtained using the apparatus with standardized samples.The apparatus was cleaned at the end of every training day. After the samples had been returned to the freezer, all exposed surfaces of the apparatus were wiped with a 1:1 isopropanol/water solution and paper towels. The chambers were washed using a dishwasher powder containing enzymes, rinsed in tap water, then rinsed again using 1:1 isopropanol/water solution and left to air dry overnight.

2.3 Sample Materials and Handling

All breath sample tubes were assembled by a single researcher at the dog laboratory. Latex gloves were used when handling all sample material. Gloves were changed as soon as any new sample was handled. Annealed glass tubes (Duran tubing, 120 mm long, 26 mm OD, 2-mm wall thickness) were packed with 1 g of L10Y4 polypropylene fiber (manufactured by IFG Asota) sandwiched between two halves of a split single cotton wool ball. In instances where glass tubes were re-used, they were soaked for 24 hours in a 60% nitric acid solution, then oven dried after being rinsed with deionized water. Tube caps (tapered plug/cap LDPE

red 23.62 mm D MX, manufactured by Hi-Q Components) were dipped in boiling water, rinsed in a 1:1 isopropanol/water solution, then air-dried prior to use. The tubes were transported and stored in sealed plastic containers cleaned using 1:1 isopropanol/water solution and air-dried for 24 hours before use.

During the course of the three-year project the same two nursing staff gathered breath samples in one designated consulting room from 900 pre-cancer diagnosis participants attending a local hospital respiratory clinic. Four tubes were prepared for each participant. Two tubes were used by the scent-detection laboratory, while the other two were sent to the university's chemistry department for chemical analysis (for a separate study). Specific participant instructions for sample collection were as follows: "Take a deep breath and breathe slowly through the tube. Exhale completely, until there is no air left in your lungs. Breathe three times through each of the tubes." Illustrations were provided on the instruction sheet, and the process was modelled by the nurses. Immediately after sampling was completed, each tube was sealed, placed in an individual plastic bag which was also sealed, and all four bagged and sealed breath samples were then placed in a larger bag, which was also sealed and transferred to a -80°C freezer at the hospital. Samples were packed in dry ice when they were transferred from the hospital to the scent detection laboratory's -80°C or -60°C freezers, subject to available space. The disease-status of the first 460 samples was prospectively adjudicated for training purposes by a specialist respiratory physician with access to up-to-date clinical records including results from multi-disciplinary meetings following review of cancer diagnosis. The adjudication results (355 lung cancer negative, 99 lung cancer positive, 3 unknown, 3 consent withdrawn) were then shared with the scent-detection researchers.

Given that current smokers were more prevalent in the group of individuals who had been diagnosed with lung cancer than those who were not, the proportion of positive and negative samples from current smokers were adjusted to be approximately equivalent each day so that

dogs would not develop a smoking-related bias. Samples were positioned in the apparatus in a randomized order. Prior to use, samples were removed from freezer storage. Each sample was placed in its own metal tray and remained sealed for at least 30 minutes, which was the time required for the samples to reach room temperature. Latex gloves were changed every time a new sample was handled, and care was taken not to touch the outside of the outer bag after having handled interior bags and samples, to prevent cross-contamination. After 30 minutes, the researcher transferred the first sample tube to the experiment room, removed the tube caps, placed the caps face up in the first chamber of the apparatus, then positioned the first tube on its side on the base of the chamber with both ends exposed. The researcher then returned to the work room and repeated the process 16 more times. This required approximately 15 to 20 minutes. Once all samples were in position, the apparatus lid was lowered on top of the chambers, sealing them. The headspace in the chambers was allowed to approach equilibrium for 10 minutes prior to working with the dogs. After training, the samples were sealed and returned to the freezers following the same process in reverse.

2.4 Procedure

One array of 17 samples was assembled and placed in the apparatus for presentation to the dogs in the experiment room each training day. Patients from each required category (according to lung cancer and smoking status) were randomly selected for inclusion in the array. After the first sample from one patient had been presented for the first time, in a subsequent array, the second sample from the same patient would be presented. Subsequently, the samples from that patient were re-used in the same alternating fashion (see Table 6). Two samples from the same patient were never presented in the same array. On each training day, each dog had up to 10 opportunities to evaluate the same breath sample array.

Table 5

Order of Use for 2 Samples from 1 Patient

Training Day[a]	Sample Number	Re-use Frequency
1	1	0
2	2	0
3	1	1
4	2	1
5	1	2
6	2	2
7	1	3
8	2	3

[a]Training days with an individual patient's samples were not necessarily consecutive

During training, indication duration requirements for a sample to be treated as "indicated" by the dogs began at 500 ms. An indication duration below 500 ms, followed by activation of the switch to advance the carousel to the next sample, was treated as a "miss" response to a lung-cancer positive sample or a "correct rejection" response to a lung cancer negative sample. Positive indication duration was increased by 100 ms after each accurately-completed session at one duration value. These values were increased to an upper indication threshold of 3500 - 4500 ms, based on the threshold that produced optimal performance for each individual dog. In instances where a dog missed (failed to indicate) a positive sample, the indication threshold associated with that positive sample for the individual dog was lowered and then gradually raised to facilitate acquisition of the lung cancer target. These adjustments to individual sample indication thresholds were carried out across sessions on individual days; a single indication threshold was used for all samples in the first session of each day.

2.5 Data Analysis

The dogs' performance was evaluated according to the frequency of sample re-use. Data related to sample tubes that had been opened and used once for training, then re-used up to three more times on later training days, were used for the present analysis. The dogs evaluated the same samples repeatedly on training days. Only data from the dogs' first session of the relevant training day were used when evaluating their tendency to indicate samples according to their re-use status. At the beginning of the training phase, some of the dogs demonstrated extremely rapid target acquisition, which suggested that there may have been a contamination issue with the preparation procedures for training sessions (e.g., systematic introduction of olfactory cues). After re-evaluating and refining the preparation procedures, accuracy returned to the levels expected during early training. Consequently, data from the first 15 training days were eliminated from this analysis.

A linear mixed model was used to evaluate the relationship between sample re-use frequency and accuracy measures. Individual dog performance was identified as a random factor in the equation, while sample use and re-use frequency and sample number (S1 or S2) were fixed factors. The data from the training and testing phases of the experiment were combined for this analysis. IBM SPSS Statistics 27 was used for all statistical analyses. Statistical significance was determined with $\alpha = .05$.

Log *d* and log *B* measures were calculated for each sample re-use frequency level to investigate potential changes in the dogs' ability to identify positive samples or changes in response bias. Log *d* is a signal detection measure that reflects the detector's ability to identify the target, with higher values indicating greater target detectability.

$$\log d = \frac{1}{2}\log\left[\left(\frac{\text{hit}}{\text{miss}}\right)\left(\frac{\text{correct rejection}}{\text{false alarm}}\right)\right]$$

Log *B* is a signal detection measure that can determine the extent to which a detector shows a positive or negative indication bias. Numbers greater than zero indicate a positive indication bias (liberal response bias), while numbers less than zero indicate a negative indication bias (conservative response bias).

$$\log B = \frac{1}{2}\log\left[\left(\frac{\text{hit}}{\text{miss}}\right)\left(\frac{\text{false alarm}}{\text{correct rejection}}\right)\right]$$

Results

Seven dogs participated in the study from the outset, but attrition reduced the number of participants to four. Rylea and BJ became ill and withdrew, while Bramble developed an aversion to the food dispenser which could not be remediated. The data produced by these dogs prior to their departure were included in this analysis.

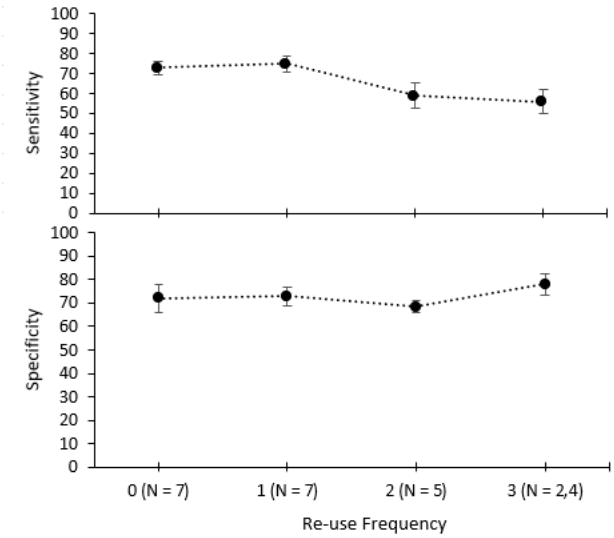
Figure 7 shows mean dog accuracy with breath samples that were opened once, twice, three times and four times during training. The mean sensitivity shows a drop in accuracy as the number of exposures increases, while the mean specificity shows a very slight increase. Individual dog performance data are available in the supplementary file, S1.

Figure 8 displays log *d* and log *B* values across re-use frequency. Log *d* values, indicative of the dogs' ability to detect the target, did not differ in the first two re-use frequency levels, but dropped slightly across the last two levels (Figure 8, top). The first two log *B* values were near zero, while the last two were below zero, indicating that the dogs were more inclined to treat samples as negative as the sample re-use frequency increased (Figure 8, bottom).

The linear mixed model analysis of sensitivity data showed significant effects for the coefficient for re-use frequency (-6.65, p = <.001, 95% CI [-10.56, -2.76]). Significant effects were also identified for the coefficient for re-use frequency in the specificity data (2.87, p = .036, 95% CI [.19, 5.55]). Analysis of log *B* data also identified the coefficient of re-use

Figure 7

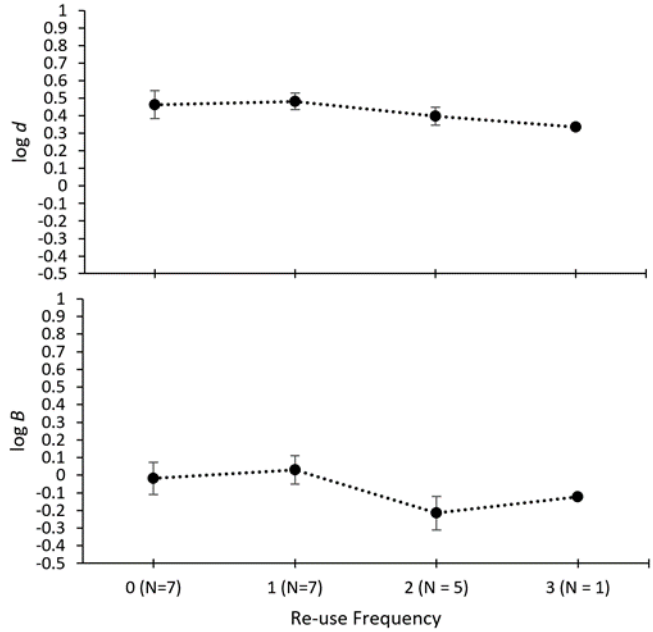
Sensitivity and Specificity Across Responses to Reused Samples



Note: Mean sensitivity (top) and specificity (bottom) across sample re-use frequency. Error bars represent standard error of the mean. For use frequency 3, N = 2: sensitivity value, 4: specificity value.

Figure 8

Log d *and* B *Measures Across Responses to Reused Samples*



Note: Mean log *d* (top) and log *B* (bottom) measures across sample re-use frequency. Error bars represent standard error of the mean. The N value beside each re-use frequency level indicates the number of dogs contributing data to each level.

frequency as a significant factor (-.094, p = <.001, 95% CI [-.149, -.39]), but this coefficient was not significant in the analysis applied to the log *d* results (-0.30, p = .285, 95% CI [-.087, .26]). The results from the linear mixed model analyses did not reveal any significant effects of accuracy rates between S1 and S2 responses (sample number factor) or individual dog performance.

Discussion

We evaluated the effect of sample re-use on the accuracy of dogs working to identify breath samples from lung cancer positive individuals. Sample re-use frequency was found to have a significant effect on the dogs' sensitivity, which decreased as sample use frequency increased. Re-use frequency also significantly influenced specificity, which increased as sample use frequency increased. As sample use frequency increased, a measure of bias (log *B*) shifted from neutral to a bias away from indicating samples as positive. Interestingly, there was no significant change in the measure of detectability of the lung cancer target (i.e., the dogs' ability to distinguish between positive and negative samples; log *d*), although there was a slight downward trend in this measure as use frequency increased.

The finding that the shift in bias was statistically significant but not the decrease in detectability is somewhat puzzling. If sample degradation is responsible for the change in sensitivity and specificity, it seems that this change would be the result of reduced availability of VOCs that are used to classify the samples and allow dogs to develop a target concept. Reduced availability of sample information should reduce the detectability of the target (log *d*) rather than the dogs' response bias (log *B*). However, it may be that dogs are simply less inclined to indicate a sample as positive when there is less information present. Our finding that sensitivity decreased despite the dogs being exposed to the samples previously suggests that the effects of sample degradation (which is expected to decrease sensitivity) outweigh the

potential influence of the dogs' previous exposure and feedback for responding to specific samples (which is expected to increase sensitivity).

To the extent that dogs can "recognise" previously encountered samples, this indicates that samples are preserved sufficiently so that these individual sample characteristics are still available. Accurate performance with re-used samples, even if it is partially due to experience with those specific samples, suggests that the samples are still useful for training. It is difficult to disentangle the effects of sample re-use on dogs' learning to respond to individual samples and dogs' learning to respond to a general cancer-related profile, something that this study was not designed to do. One approach that could be used to help to provide a clear picture of the effects of sample re-use on dogs' learning of a general concept may be to present samples that have been previously used with other dogs to dogs that have never been exposed to those specific samples.

A potential shortcoming of the method described above is that each dog inhales and exhales rapidly and repeatedly into each sample chamber during their encounter with the sample (45). Although dogs could not come into direct contact with the samples, it is possible that dogs could change the VOC profile of samples during this process. Given that positive samples are more likely to be indicated as positive, these samples may be changed more than negative samples, in general. Analysis of the residual VOCs that dogs are breathing into the sample chambers during presentation might therefore have merit. Questions remain about whether or not dogs could learn to indicate on other dogs' cues. No clear evidence that this occurred was apparent in the current project as dog order, which was changed regularly, was not correlated with accuracy. One testing procedure to evaluate the effects of sample re-use without the potential for dog-related VOC contamination is to open samples and present them as they are typically used for training but without exposing them to dogs, then testing the dogs' performance with them subsequently.

Importantly, the two primary limitations of the present study (the possible influence of dog-related VOCs from previous sample evaluations and the entanglement of learning effects and sample degradation effects) are representative of the conditions that are in effect when samples are re-used. Therefore, the findings are informative with respect to outcomes of sample re-use that can be expected. Nevertheless, it would be helpful to clarify the contribution of these factors to these outcomes.

Although the coefficients associated with sample re-use trends were found to be statistically significant for three out of the four dependent measures, the number of dogs contributing data to each re-use frequency was reduced at the higher re-use levels. For log *d* and *B* calculations, only one dog encountered sufficient samples that had been re-used thrice to calculate these measures. Therefore, the data associated with the highest re-use frequency level should be interpreted with caution.

Standardizing procedures for training and testing medical detection animals, including the selection of research materials (31), storage and handling of samples (42), and training and testing methodology and measurement (14, 44, 46-47) is likely to increase the quality of research and practice in this area. Practical considerations must also be taken into account when exploring the influence of relevant factors. For example, in the present study, samples were stored at -80°C or -60°C; some researchers or practitioners may not have access to freezers with these capabilities. Similar research examining sample re-use after storage at -18°C (standard domestic freezer temperature) or 4°C (standard refrigeration temperature) would clarify whether sample re-use is advisable when samples are stored under less ideal conditions.

Conclusion

There is some evidence that dogs can be trained to detect lung cancer using human breath samples. Although it appears that breath samples are commonly re-used for lung cancer research with dogs, there is limited information about the precise effects of sample re-use on

training performance. Sourcing breath samples for such research can be challenging and resource-intensive. Therefore it is important to determine the influence of this factor so that researchers or practitioners can make informed decisions about sample re-use for dog training. The present findings suggest that samples can be re-used at least three times under the conditions described herein. Based on these results, we can expect dogs' sensitivity to decline, specificity to improve, and bias to shift away from a positive indication bias across successive re-uses. The specific influences of previous exposure to specific samples and potential odor cues introduced by previous exposure of the samples to dogs remain to be explored.

Acknowledgements

Funding for the experimental procedures associated with this project was provided by the Health Research Council of New Zealand (grant 18/632), the University of Waikato, and the Waikato Medical Research Foundation (grant 281).

Ethical statement

Human ethics approval for this research was obtained from New Zealand's Health and Disability Ethics Committees (17/NTB/178/AM02). Animal ethics approval was obtained from the University of Waikato Animal Ethics Committee (Protocol 1029).

References

[1] World Health Organization. Cancer 2021 [7 September 2021]. Available from: https://www.who.int/news-room/fact-sheets/detail/cancer.

[2] Elisa D, Tommaso C, Emanuela F, Marco C, Marinos K, Giulia V, et al. Biomarkers and lung cancer early detection: state of the art. Cancers. 2021;13(15):3919.

[3] Kauczor HU, Baird AM, Blum TG, Bonomo L, Bostantzoglou C, Burghuber O, et al. ESR/ERS statement paper on lung cancer screening. Eur Radiol. 2020;30(6):3277-94.

[4] Ostrin EJ, Sidransky D, Spira A, Hanash SM. Biomarkers for lung cancer screening and detection. Cancer Epidemiol Biomarkers Prev. 2020;29(12):2411.

[5] Blandin SK, Crosbie PA, Balata H, Chudziak J, Hussell T, Dive C. Progress and prospects of early detection in lung cancer. Open Biology. 2017;7(170070).

[6] Williams J, Koppmann R. Volatile organic compounds in the atmosphere: an overview. Volatile organic compounds in the atmosphere 2007. p. 1-32. Available from: https://doi.org/10.1002/9780470988657.ch1

[7] Binson VA, Subramoniam M, Mathew L. Discrimination of COPD and lung cancer from controls through breath analysis using a self-developed e-nose. J Breath Res. 2021;15(4):46003.

[8] Buszewski B, Ligor T, Jezierski T, Wenda-Piesik A, Walczak M, Rudnicka J. Identification of volatile lung cancer markers by gas chromatography–mass spectrometry: comparison with discrimination by canines. Anal Bioanal Chem. 2012;404(1):141-6.

[9] Ehmann R, Boedeker E, Friedrich U, Sagert J, Dippon J, Friedel G, et al. Canine scent detection in the diagnosis of lung cancer: revisiting a puzzling phenomenon. Eur Respir J. 2012;39(3):669.

[10] Fischer-Tenhagen C, Johnen D, Nehls I, Becker R. A proof of concept: are detection dogs a useful tool to verify potential biomarkers for lung cancer? Front Vet Sci. 2018;5:52.

[11] McCulloch M, Jezierski T, Broffman M, Hubbard A, Turner K, Janecki T. Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. Integr Cancer Ther. 2006;5(1):30-9.

[12] Rudnicka J, Walczak M, Kowalkowski T, Jezierski T, Buszewski B. Determination of volatile organic compounds as potential markers of lung cancer by gas chromatography–mass spectrometry versus trained dogs. Sens Actuators B Chem. 2014;202:615-21.

[13] Walczak M, Jezierski T, Górecka-Bruzda A, Sobczyńska M, Ensminger J. Impact of individual training parameters and manner of taking breath odor samples on the reliability of canines as cancer screeners. J Vet Behav. 2012;7(5):283-94.

[14] Pirrone F, Albertini M. Olfactory detection of cancer by trained sniffer dogs: a systematic review of the literature. J Vet Behav. 2017;19:105-17.

[15] Nardi-Agmon I, Peled N. Exhaled breath analysis for the early detection of lung cancer: recent developments and future prospects. Lung Cancer (Auckland, NZ). 2017;8:31-8.

[16] Ratiu IA, Ligor T, Bocos-Bintintan V, Mayhew CA, Buszewski B. Volatile organic compounds in exhaled breath as fingerprints of lung cancer, asthma and COPD. J Clin Med. 2020;10(1):32.

[17] Edwards TL, Browne CM, Schoon A, Cox C, Poling A. Animal olfactory detection of human diseases: guidelines and systematic review. J Vet Behav. 2017;20:59-73.

[18] Herrnstein RJ, Loveland DH. Complex visual concept in the pigeon. Science. 1964;146(3643):549-50.

[19] Herrnstein RJ, Loveland DH, Cable C. Natural concepts in pigeons. J Exp Psychol, Anim Behav Process. 1976;2(4):285-302.

[20] Zhan X, Wang Z, Yang M, Luo Z, Wang Y, Li G. An electronic nose-based assistive diagnostic prototype for lung cancer detection with conformal prediction. Measurement. 2020;158:107588.

[21] Wang XR, Cassells J, Berna AZ. Stability control for breath analysis using GC-MS. J Chromatogr B Analyt Technol Biomed Life Sci. 2018;1097-1098:27-34.

[22] Angle C, Waggoner LP, Ferrando A, Haney P, Passler T. Canine detection of the volatilome: a review of implications for pathogen and disease detection. Front Vet Sci. 2016;3:47.

[23] Shirasu M, Touhara K. The scent of disease: volatile organic compounds of the human body related to disease and disorder. J Biochem. 2011;150(3):257-66.

[24] Alonso M, Sanchez JM. Analytical challenges in breath analysis and its application to exposure monitoring. Trends Anal Chem. 2013;44:78-89.

[25] Tang Z, Liu Y, Duan Y. Breath analysis: technical developments and challenges in the monitoring of human exposure to volatile organic compounds. J Chromatogr B Analyt Technol Biomed Life Sci. 2015;1002:285-99.

[26] Amann A, Costello BdL, Miekisch W, Schubert J, Buszewski Ba, Pleil J, et al. The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. J Breath Res. 2014;8(3):034001.

[27] Pennazza G, Santonico M. Breath Analysis. San Diego (US): Elsevier Science & Technology; 2018. 202 p.

[28] Reeve C, Wentzell P, Wielens B, Jones C, Stehouwer K, Gadbois S. Assessing individual performance and maintaining breath sample integrity in biomedical detection dogs. Behav Process. 2018;155:8-18.

[29] White IR, Fowler SJ. Capturing and storing exhaled breath for offline analysis. In: Pennazza G, Santonico M, editors. Breath Analysis. San Diego (US): Academic Press; 2019. p. 13-31.

[30] Herbig J, Beauchamp J. Towards standardization in the analysis of breath gas volatiles. J Breath Res. 2014;8(3):037101.

[31] Goss K-U. The physical chemistry of odors — consequences for the work with detection dogs. Forensic Sci Int. 2019;296:110-4.

[32] Harshman SW, Mani N, Geier BA, Kwak J, Shepard P, Fan M, et al. Storage stability of exhaled breath on Tenax TA. J Breath Res. 2016;10(4):046008.

[33] Ho SSH, Wang L, Chow JC, Watson JG, Xue Y, Huang Y, et al. Optimization and evaluation of multi-bed adsorbent tube method in collection of volatile organic compounds. Atmos Res. 2018;202:187-95.

[34] van der Schee MP, Fens N, Brinkman P, Bos LDJ, Angelo MD, Nijsen TME, et al. Effect of transportation and storage using sorbent tubes of exhaled breath samples on diagnostic accuracy of electronic nose analysis. J Breath Res. 2012;7(1):016002.

[35] Hudson DT, Curran AM, Furton KG. The stability of collected human scent under various environmental conditions. J Forensic Sci. 2009;54(6):1270-7.

[36] Curran AM, Prada PA, Furton KG. The differentiation of the volatile organic signatures of individuals through SPME-GC/MS of characteristic human scent compounds. J Forensic Sci. 2010;55(1):50-7.

[37] Holloway DG. The physical properties of glass. London (England): Wykeham Publications; 1973. 220 p.

[38] SpecialChem. The definitive guide to polypropylene (PP): Omnexus; 2021 [updated 18 September 2021]. Available from: https://omnexus.specialchem.com/selection-guide/polypropylene-pp-plastic.

[39] Hentz RV, Traina GC, Cadossi R, Zucchini P, Muglia MA, Giordani M. The protective efficacy of surgical latex gloves against the risk of skin contamination: how well are the operators protected? J Mater Sci Mater Med. 2000;11(12):825-32.

[40] Zhang H, Cloud A. The permeability characteristics of silicone rubber. Society for the Advancement of Material and Process Engineering; 2006 [cited 2021 18 September]. Available from: https://imageserv5.team-logic.com/mediaLibrary/99/D116_20Haibing_20Zhang_20et_20al.pdf.

[41] Carter ME. Essential fiber chemistry. New York (US): M. Dekker; 1971. 216 p.

[42] Schoon GAA. The effect of the ageing of crime scene objects on the results of scent identification line-ups using trained dogs. Forensic Sci Int. 2005;147(1):43-7.

[43] Edwards TL. Automated canine scent-detection apparatus: technical description and training outcomes. Chem Senses. 2019;44(2):449-55.

[44] Edwards TL, Giezen, C, Browne, CM. Influences of indication response requirement and target prevalence on dogs' performance in a scent-detection task. Appl Anim Beha. Sci. 2022;253:105657.

[45] Kokocińska-Kusiak A, Woszczyło M, Zybala M, Maciocha J, Barłowska K, Dzięcioł M. Canine olfaction: physiology, behavior, and possibilities for practical applications. Animals (Basel). 2021;11(8):2463.

[46] Ellis H, Mulder C, Valverde E, Poling A, Edwards TL. Reproducibility of African giant pouched rats detecting *Mycobacterium tuberculosis*. BMC Infect Dis. 2017;17(1).

[47] Crawford MA, Perrone, JA, Browne CM, Chang, CL, Hopping, S, Edwards TL. Transitioning from training to testing with scent detection animals: application to lung cancer detection dogs. J Vet Behav. 2022;55-56:23-24

Chapter Four

Experiment 3

Transitioning from Training to Testing with Scent Detection Animals: Application to Lung Cancer Detection Dogs[4]

[4] Crawford, M. A., Perrone, J. A., Browne, C. M., Chang, C. L., Hopping, S., & Edwards, T. L. (2022). Transitioning from training to testing with scent detection animals: Application to lung cancer detection dogs. *Journal of Veterinary Behavior*, *55-56*, 23–34. https://doi.org/10.1016/j.jveb.2022.07.004

Model fitting and analysis

If scent detection target acquisition can be represented as a sigmoid curve, the upper asymptote on that curve shows the instance where the task of identifying a scent target cannot be improved upon in any meaningful way by an individual organism. The detector organism in this example needs no more training apart from the intermittent behavioural maintenance work required during testing. The value of asymptote will vary significantly between individual animals as a result of environmental and idiosyncratic differences. However, the performance of a team of scent detection animals treated as a single detector could provide a solid basis for deciding on the reliability of specific samples for specific targets. Having a quantitative basis – a mathematical model - for choosing when to transition from training to testing makes disease scent detection projects more easily replicable, and more accessible to researchers who may not necessarily have extensive experience training dogs. No research on this specific topic had been published at the time of writing.

Transitioning from Training to Testing with Scent Detection Animals: Application to Lung Cancer Detection Dogs

Margaret A. Crawford[1], John A. Perrone[1], Clare M. Browne[2], Catherina L. Chang[3], Sandra Hopping[3], & Timothy L. Edwards[1]

[1] School of Psychology, University of Waikato, Hamilton, New Zealand

[2] School of Science, University of Waikato, Hamilton, New Zealand

[3] Department of Respiratory Medicine, Waikato Hospital, Hamilton, New Zealand

Author Note

Address correspondence to:
Timothy L. Edwards
University of Waikato
Private Bag 3105
Hamilton 3240
New Zealand
edwards@waikato.ac.nz

Abstract

Scent-detection animals are routinely evaluated for detection accuracy for a wide range of targets. The obtained estimates of accuracy vary widely, even with the same species of detector (e.g., dogs) and target type (e.g., lung cancer). One factor that may contribute to these varied results is variability in the point at which the detection animals are transitioned to the accuracy test. If training is incomplete, a nonrepresentative estimate of the animal's potential performance will likely be obtained. Descriptions of evaluation research with scent-detection animals commonly include little information about the factors influencing the researchers' decision to transition from training to testing. When this information is reported, the decision criteria are often problematic. Herein, we explore methods of evaluating when a detection animal is performing at or near the highest accuracy of which they are capable and, therefore, when it may be suitable to transition to an accuracy test. We then apply these methods to an applied scent-detection project in which dogs were being trained to classify breath samples according to the status of the sample donor as lung-cancer-negative or lung-cancer-positive. In this specific case, a quantitative model of acquisition was most informative. We recommend that researchers in this field apply and report on evaluations of acquisition when conducting and describing evaluation research with scent-detection animals.

Keywords: acquisition, diagnostic accuracy, medical detection, modelling, olfaction.

Dogs are often employed to detect chemically complex targets such as cash, food items, cryptic animals and material decay because of their acute olfaction and trainability (Browne et al., 2015, Kokocińska-Kusiak et al., 2021). Recently, dogs and other animals have been evaluated for their potential to serve as detectors of diseases in humans and other species (Pleil and Giese, 2017, Ellis et al., 2017, Edwards et al., 2017, Pirrone and Albertini, 2017). Researchers working with disease-detection and other detection animals often undertake to produce accurate estimates of the animals' ability to detect the target. Some of this research is aimed at determining the suitability of the detection animal for operational use. For example, African pouched rats have been evaluated for and are currently used for detection of tuberculosis (Ellis et al., 2017). In other cases, researchers produce accuracy estimates that can be used as comparative estimates for the development of machine-based detection technology (Fischer-Tenhagen et al., 2018). In either case, an accurate representation of the detector's ability to correctly classify potential targets as positive or negative is crucial.

For any given detector and target combination, estimates of diagnostic accuracy vary widely. For example, Edwards et al. (2017) found that studies evaluating dogs as detectors of lung cancer produced estimates of sensitivity (correct positive indications/total positive samples) between 56% and 99% and estimates of specificity (correct rejection of negative samples/total negative samples) between 8% and 99%. They also found that estimates associated with other diseases were highly variable. The anatomic and ontogenetic diversity of the animal participants across studies may contribute to this variance (Stevens, 2013; Teodoro-Morrison et al., 2014; Lazarowski et al., 2019).

Edwards et al. (2017) and other researchers (Johnen et al., 2013; Elliker et al., 2014, Jezierski et al., 2015; Jamieson et al., 2017; Johnen, 2017; Pirrone and Albertini, 2017) have also identified specific methodological differences that may be responsible for these various outcomes, and have highlighted the need for a more uniform approach and agreed standards

of practice in this field of research. These methodological differences include variability in sample collection, handling and storage; determination of the number of unique sample sources used in training and testing, which should be but is usually not supported by a power analysis; sample source (i.e., patient and control) characteristics; and blind-testing protocols, including the method of sample presentation and determination of the detector's response to each individual sample.

One factor that may also contribute to variable outcomes in this evaluation research is the criteria employed to determine when an animal is finished training. Similar decision-making processes are involved if phase changes are required during complex target acquisition. Specifically, if an animal being trained to detect a disease or another target has not attained or approached the maximum accuracy of which it is capable, a blind test (or change in training conditions) will provide a misleading representation of the animal's ability to classify potential targets. Additionally, the blind test conditions may result in significant disruptions in the animal's performance (Jakovcevic et al., 2013; Elliker et al., 2014; Hall, 2017). If the researcher changes the training phase or transitions to testing too late, they consume scarce resources – including the animal's output as a proportion of their overall lifespan – without generating reliable evidence of performance. Training-to-testing transition criteria have received little attention in the field of olfaction, yet they represent a major source of potential variability in estimates of diagnostic accuracy.

Herein, we explore key considerations when making the decision to transition from training to testing, including production, selection and analysis of relevant data. We then apply this analysis in the context of a project in which dogs were being trained to detect lung cancer using an automated apparatus. Our aims with this work are to identify best practice (or approximations thereof) for transitioning from training to testing, evaluate the analytical tools

that are available for this purpose, and provide a demonstration of the application and interpretation of these analyses in an applied research project.

Canine Detection of Lung Cancer

Lung cancer is responsible for more cancer-related deaths than any other type of cancer, worldwide (Barta et al., 2019). Long-term patient survival rates improve if the cancer is detected early, but in approximately 75% of cases, the disease is not detected until it is well advanced (stage III or IV) (Blandin et al., 2017; World Health Organization, 2019). There appear to be some reliable changes in volatile organic compounds (VOCs)[5] in the breath of individuals with lung cancer (Saalberg and Wolff, 2016). To the extent that differences in VOC patterns between people with and without lung cancer can be reliably detected, VOC analysis may represent a potential non-invasive lung cancer screening method. Dogs trained to identify breath samples collected from individuals with lung cancer have produced a wide range of estimates of diagnostic accuracy (see Table 7).

The methodology across these studies differed considerably, and their variable outcomes could have been the result of any combination of factors from sampling through to testing conditions. Criteria for transitioning from training to testing also varied widely. The criteria in all of these studies, where specified, were applied to evaluation of samples that the dog may have been exposed to previously. These criteria are problematic for several reasons, which we outline below.

Consideration for Transitioning from Training to Testing

Novel Sample Sources

Re-use of samples is common when training detection animals, particularly when samples are difficult to obtain (McCulloch et al., 2006; Walczak et al., 2012; Rudnicka et al.,

[5] VOCs are low-chemical weight substances that are released into the atmosphere at room temperature. See Hung, R., Lee, S. & Bennett, J. 2015b. Fungal volatile organic compounds and their role in ecosystems. Appl. Microbiol. Biotechnol., 99**,** 3395-3405.

Table 6

Research on Canine Scent Detection of Lung Cancer using Breath Samples

Study	Sensitivity	Specificity	Training to Testing Criteria
McCulloch et al. (2006)	99%	99%	30 consecutive correct trials
Buszewski et al. (2012)	82%	82%	40 consecutive correct trials
Ehmann et al. (2012)	71%	93%	Not available
Rudnicka et al. (2014)	86%	72%	40 consecutive correct trials
Amundsen et al. (2014)	56-76%	8-33%	"99% test sensitivity"
Fischer-Tenhagen et al. (2018)	9/9, 8/9	8/10, 4/10	"80% correct indications"

2014; Jezierski et al., 2015; Fischer-Tenhagen et al., 2018). Novel samples from the same source (e.g., the same patient) are also commonly used in training. Neither practice is necessarily problematic, provided that the animal is exposed to enough unique sample sources that they can learn to generalize accurate classification responses to novel sample sources (Herrnstein and Loveland, 1964; Herrnstein et al., 1976). Generalisation to novel sources is precisely the ability that we aim to evaluate in a testing scenario and, therefore, it should also be the primary consideration when deciding when the animal is ready to be tested. Consequently, the only data that are relevant to testing are obtained from the animal's first exposure to samples from novel sources.

For example, if a dog has been trained with repeated exposure to samples from 10 individuals with breast cancer and 30 without, the only new information that will inform us of the dog's ability to identify the breast cancer profile will come from its first response to

samples from individuals that were not in this original training set (e.g., an 11th person with breast cancer and a 31st person without). Classification accuracy of samples from the original training set will tell us nothing about the dog's ability to do what the test aims to evaluate unless we retrospectively look at their response when they first encountered a sample from each individual. The animal's responses to samples to which they have already been exposed can be evoked by idiosyncratic features of the individuals or samples rather than the general target profile. Edwards et al. (2017) suggest that an efficient approach to training a detection animal involves regular introduction of samples from novel sources while attending to the animal's performance with these samples (as opposed to introducing samples from all available sources at once, for example). Under no circumstances should the decision to transition from training to testing be based on performance with sample sources to which the animal has already been exposed.

Accuracy Measures

Diagnostic Accuracy Measures. Sensitivity and specificity are commonly reported indicators of diagnostic accuracy (Zhou, 2011). Sensitivity, often recorded as a percentage, is generated by dividing the number of correct positive indications by the number of known positive samples. This measure represents the detector's accuracy with positive samples. Specificity is calculated by dividing the number of correct rejection responses by the total number of negative samples available in the array. This measure indicates the detector's accuracy with negative samples; in practice, this is the proportion of non-target samples correctly ignored (Lazarowski et al., 2020). Both measures are important when evaluating the usefulness of potential diagnostic technology. Low sensitivity indicates that the test is missing many positive cases, whereas low specificity indicates that the detector is producing many false alarms.

Sometimes, a single measure of diagnostic accuracy can be useful. For example, when analyzing data for stability, as we aim to do herein, it is more straightforward to apply an analytic technique to a single measure (Gallistel et al., 2004). One simple approach to producing a single measure involves summing sensitivity and specificity measures (i.e., sensitivity + specificity). Interpretation of this measure is clear and intuitive; values lower than 1.0 indicate lower-than-chance performance, while those above 1.0 show higher-than-chance performance. The maximum value, perfect accuracy, is 2.0. A similar approach involves calculating the mean of the summed specificity and sensitivity data, which restricts the measure to a maximum of one and a minimum of zero instead. Values near 0.5 represent chance performance. We have used this latter measure in the present study.

Signal detection theory offers some alternative approaches to generating meaningful accuracy measures (Macmillan and Creelman, 2004). One such measure, *d'*, is a parametric index which provides an estimate of accuracy with both positive and negative samples, independent of bias (Hautus, 2015). Bias represents the degree to which a detector tends to classify samples as positive or negative and is independent of the detector's ability to detect the target. For example, providing a large reward for hits but arranging no consequences (such as delayed access to the next sample) for false alarms may result in a strong bias toward indicating that the target is present. However, it will not influence the participant's ability to detect the target (Macmillan and Creelman, 1990). In practice, a dog may tend to indicate samples as positive or negative regardless of sample status; *d'* provides an estimate of the dog's true ability to identify the target without the influence of this underlying bias.

Unit of Analysis. The selection of relatively small units of analysis, including individual responses to samples or session-by-session data, will produce more variable data compared to the data produced by larger units of analysis, such as daily or weekly performance (Johnston et al., 2019). For the purpose of evaluating stability, selecting too

small of a unit of analysis (e.g., individual responses) may produce too much "noise" for some analytic techniques to be usefully applied. However, if too large of a unit of analysis (e.g., monthly data) is applied, then the ability to detect changes within the period is lost. Herein, we primarily rely upon session data, a session being comprised of 17 sample evaluations, with between three and nine positive samples.

Stability, Not Accuracy

The goal of testing is to produce useful estimates of accuracy. Therefore, it is inappropriate to require a specific degree of accuracy before transitioning to testing. Accuracy tends to improve throughout training up to a level at which it stabilizes and no longer improves significantly. The point at which a function such as the acquisition curve approaches a theoretical maximum is referred to as its "asymptote". It is not possible to predict the level at which an individual's accuracy will stabilize (i.e., reach asymptote), so training and testing criteria must be based on stability rather than accuracy criteria. Researchers should endeavor to move from training to testing only when they have evidence that dogs' accuracy has reached or approached asymptote. Asymptote can be identified using a variety of methods, including visual analysis and quantitative analysis. Herein, we explore some of the approaches to identifying asymptote that seem suited to the task of making decisions in a scent-detection scenario.

Approaches to Analysis

Visual Analysis

Visual analysis is commonly applied to data from single-subject (i.e., repeated-measures, within subjects) designs (Lane and Gast, 2014; Johnston et al., 2019). This approach to analysis involves identification of stability, which is defined as the absence of trends and excessive variability in the data. When applied to acquisition data, identification of an increasing trend would lead the researcher to conclude that acquisition is not yet complete.

An experienced visual analyst can incorporate many features of the data at once when analyzing single-subject data. Therefore, visual analysis can be superior to quantitative approaches that rely only on one or a few features of the data (Johnston et al., 2019). Nevertheless, concrete, quantitative rules for determining when data are stable (or when asymptote has been reached) can be beneficial for a variety of reasons. Quantitative rules can be reported when describing the decision to transition between training phases or from training to testing in the research methods. Researchers who are not experienced with visual analysis can apply a quantitative rule without the need to acquire visual analysis skills. Additionally, visual analysts, even highly trained ones, often disagree on conclusions that can be drawn from the data (Jones et al., 1978; Nourbakhsh and Ottenbacher, 1994).

Simple Quantitative Analysis

Simple quantitative analysis approaches to analyzing single-subject data have been developed as alternatives to visual analysis (Baer and Parsonson, 1981; Nourbakhsh and Ottenbacher, 1994). Most relevant to the current assessment, these approaches to analysis include methods of determining whether trends are present within phases. For example, the "split-middle method" involves identifying horizontal and vertical median points and using these to construct a trendline (Zhan and Ottenbacher, 2001). Applying a simple linear regression model to the data is another approach. The main issue with these approaches to analysis for the purpose of determining when acquisition is complete is that all data in a phase are used to calculate a trend and, therefore, a trend would nearly always be detected when applied to acquisition data. This is because accuracy measures are expected to start at a low value and progress to a higher value, where they should eventually stabilize.

One potential solution to this issue is to apply the trend analysis only to the most recently obtained data. For example, a trend analysis might be conducted on the data from the most recent 10 sessions. The main weaknesses of this approach are as follows. First, the

number of sessions to which the data analysis will be applied is arbitrarily determined, and the outcome of the analysis is likely to differ depending on the time period chosen. Second, the slope criterion that the researcher uses to determine that a trend is present is also arbitrarily determined. Therefore, an analysis of trend alone does not take variability into account. Probability values (*p*-values) that are produced by statistical analyses such a linear regression, which are partially a function of variability, are not suitable for making decisions about transitioning from training to testing. If the data are highly variable, the *p*-value is likely to be non-significant, even if a trend is present, so the researcher may decide to transition to testing too early. Nevertheless, applying a trend analysis to data from the most recent "X" sessions may represent a useful quantitative approach to determining when acquisition is complete, taking these limitations into account.

Models of Acquisition

The data associated with acquisition are often characterized by initial chance performance, followed by an increase in accuracy, then a gradual levelling off (i.e., approaching asymptote), so appropriate models will accommodate this general shape (Murre, 2014). A variety of mathematical models have been applied to acquisition data (Mazur and Hastie, 1978; Gallistel et al., 2004; Deliano et al., 2016). One sigmoidal model that appears to be appropriate for acquisition applied to accuracy data is the Weibull cumulative function.

The Weibull cumulative function is commonly used in psychophysics (Gallistel et al., 2004; Mortensen, 2002) and has several features that are useful in the context of identifying asymptote in scent-detection animals using the measure that we have chosen (the mean of sensitivity and specificity): the model assumes a starting value of 0.5, only two free parameters (slope and inflection point) are in the model, the model has a sigmoidal form, the model has predictive value (i.e., future accuracy values can be forecast), and the model can

predict asymptotic values lower than one. Therefore, we applied this model in the present study, but other models may also be appropriate for this purpose.

The Present Study

In the present study, we evaluated the utility of the analytical techniques described above for determining when to transition dogs from training to testing for lung cancer detection. Seven dogs were trained to identify breath samples from individuals with lung cancer. An automated apparatus that presented samples, recorded responses, and provided feedback to the dogs enabled the dogs to work in the absence of a researcher (see Edwards, 2019). We monitored the dogs' performance with samples from novel patients and analysed the data visually and quantitatively to determine when the dogs should be transitioned between training phases and ultimately to the testing phase of a larger diagnostic study.

Method

Subjects

Nineteen dogs were screened for inclusion in the study. Out of the seven dogs that met the inclusion criteria (see Table 8), four dogs completed the study. Initial screening included the assessment of the effectiveness of dry kibble as a reinforcer for each dog and the dog's suitability for intraday housing and work in a laboratory environment. Five dogs were eliminated at the induction stage for showing signs of sustained anxiety in the laboratory, while four dogs were eliminated for signs of aggression in the presence of other dogs. A further three dogs were not sufficiently food motivated to complete the training. Three dogs were inducted into the training program but did not continue to the testing phase, two of them because of ill health and one because it acquired an aversion to the feeding apparatus.

Four of the seven dogs completed the training and transitioned to the testing phase. Two dogs, Tui and Katie, had previously participated in another scent detection project (target: amyl acetate) using the same apparatus that was used in this research (Edwards et al.,

Table 7

Participant Dog Profiles

Name	Sex*	Age at Recruitment	Breed
BJ	Female	10 years	Beagle
Bramble	Female	3 years	English springer spaniel
Katie	Female	9 years	Blue heeler X border collie
Onyx	Male	2 years	Labrador retriever
Rocky	Male	4 years	Jack Russell X beagle
Rylea	Female	6 years	Labrador retriever
Tui	Male	3 years	Border collie X huntaway X Australian kelpie

Note: dogs with bold names completed the study

*All dogs were neutered

2022). The other two dogs, Onyx and Rocky, learned to use the apparatus to indicate the presence of amyl acetate prior to training with breath samples for lung cancer detection. Dogs were only housed in the research facility during the day, typically for a morning or afternoon twice each week. Handling and care of the dogs was in accordance with a standard operating procedure (SOP 20). This SOP and the use of dogs for this project was approved by the University of Waikato Animal Ethics Committee (Protocol 1029).

Apparatus and Setting

The apparatus used for this research was a fully automated 1-m^3 device housing a carousel with 17 enclosed segments that was situated behind a panel with a 10 cm port through which segments were accessed (see Figure 9). An infrared beam running across the port aperture was used to record dogs' observations (described below) and indications of

samples. An omnidirectional switch used by the dogs to rotate the carousel to the next segment was positioned on the right edge of the panel, 50 cm from the floor. An automated food dispenser (Treat & Train Remote Dog Trainer manufactured by Pet Safe) was placed approximately 1.5 m away from the port. See Edwards (2019) and Edwards et al. (2022) for further apparatus specifications and results obtained using the apparatus with standard chemicals.

The apparatus was situated in a 3.2-m by 4.3-m room where, after initial training, dogs operated the apparatus alone. This activity was monitored and recorded via two cameras connected to a computer in an adjacent room. Another computer ran the software that was used to control the apparatus and record the dogs' responses. Ambient temperature in the experiment room was maintained at approximately 20°C.

Samples were accessed by the dog via the port on the panel at the front of the carousel (see Figure 9). The infrared sample port beam had to be broken by the dog for at least 500 ms

Figure 9

Apparatus Photographs

Note: Left: front panel of the apparatus with odor port and rotation switch (on right-hand side of front panel). Right: 17-segment carousel with lid removed. Note: dogs were only able to access the front panel of the apparatus.

for the sample to be treated as “observed”. Activation of the omnidirectional switch would only rotate the carousel to the next sample after an observation response had occurred. The beam-break duration required for a sample to be treated as “indicated” was set based on an optimal value for each individual dog, typically between 3500 ms and 4500 ms. Indication responses to a negative sample or a “non-reinforced” positive sample (when intermittent reinforcement was arranged) had no consequences; in this circumstance the dog was required to activate the omnidirectional switch in order to progress to the next sample. When the dogs were fully trained, a single session with one carousel rotation through all 17 samples took approximately three to four minutes.

Samples

Breath samples were collected from patients visiting a respiratory clinic, prior to diagnosis. Four breath samples were collected from each patient using glass tubes (26 mm x 2 mm Duran tubing, 120 mm long, cute and flamed, 26 mm OD, 2 mm wall thickness) packed with 1 gram of L10Y4 polypropylene fiber (manufactured by IFG Asota) between two halves of a single cotton wool ball. Prior to use, the tubes were either annealed at 540°C or soaked for 24 hours in a 60% concentrated nitric acid solution, then rinsed with deionized water and oven dried. The tubes were sealed with plastic caps (tapered plug/cap LDPE red 23.62 mm D MX, manufactured by Hi-Q Components) which were taped into place after three full exhalations by the patient into each tube. Each tube was individually packing into its own plastic resealable bag, which in turn was placed in a bag containing all samples from each respective patient. From the outset to the end of the project, sample collection was carried out by the same nursing staff in a single designated consulting room of the respiratory clinic. Over a period of three years, 900 individual samples were collected.

The samples were stored in -80°C freezers on site at the hospital, then in equivalent freezers at the dog laboratory. For training purposes, the lung cancer status and cancer type of

the first 400 samples was prospectively adjudicated by a specialist respiratory physician, who relayed this information to the scent detection researchers. Analysis of the frequency of gender, smoking status and cancer type associated with patients who provided the samples in the lung-cancer-positive and lung-cancer-negative groups was completed. Samples were randomly selected for training except that additional randomly selected control samples from smokers were sometimes used to replace samples from non-smokers to correct for a higher proportion of smokers in the pool of lung-cancer-positive samples. This was done to avoid the development of smoking-related biases in the dogs. No significant gender differences between cancer patients and controls were observed, so no correction for gender was required.

To prevent cross-contamination, all materials associated with the experiment were handled with latex gloves, which were changed following contact with any patient's samples or other materials. Individual samples were prepared for use on individual metal trays stored on non-absorbent, sterilized surfaces. The trays were washed with detergent then rinsed with a 1:1 water/isopropanol solution and air-dried after each use. All samples were returned to the freezers each day before any cleaning was undertaken.

Procedure

After the dogs completed training using the apparatus to identify amyl acetate (see Edwards, 2019 for a description of this training protocol), the target was changed to breath samples from lung-cancer-positive individuals. The initial schedule of positive to negative breath samples was 9:8 with an indication threshold of 500 ms. The indication threshold (i.e., the duration of a sample-port beam break that was treated as a positive indication) increased by 100 ms each time the dog accurately identified all of the positive samples in the session. The sample ratio changed by one unit each week, ultimately reducing to 4:13, positive to negative. The number of kibble pellets dispensed for correct positive indication increased

from 1 (with the 9:8 positive to negative ratio) to 2 (7:10), then 3 (4:13). Finally, intermittent reinforcement of one of the four samples in each array was introduced. Exposure to a sample from a novel source was always used for this phase of training. Correct positive indication of this sample was not reinforced for the first two sessions, mimicking the lack of reinforcement for correct positive sample indication in a blind test. At this phase of the training, the number of positive samples per session was varied between 3:14 and 5:12 on a random schedule. The order in which the dogs accessed the experiment room each training day was also randomized to control for dogs' potential use of scent cues from other dogs when doing the detection task.

When dogs missed a positive sample in the first session on a training day, the trainer initially began to manually reinforce subthreshold observations of these missed samples in subsequent sessions. This process of reducing the threshold to shape indications to previously missed positive samples was subsequently achieved automatically through alteration of the software controlling the apparatus. The training conditions altered to some extent over time. Richer reinforcement ratios were reintroduced and then phased out again, while indication thresholds were sometimes dropped then increased to previous levels. These alterations were made to facilitate the dogs' performance, as required, following interruptions associated with equipment malfunctions and other disruptions, including a nine-week hiatus during New Zealand's COVID-19 lockdown period.

Data Analysis

Sensitivity and specificity values were calculated following each session. Tui, Katie, and Onyx all attained consecutive 100% mean sensitivity and specificity scores for at least five sessions from the fifth training session where exposure to novel samples occurred. This raised questions about the possibility of some form of unintentional cue associated with positive samples. Although the experimental sessions were automated, the researcher was not blind to sample status when preparing the samples and testing the apparatus prior to daily

sessions. Additional measures were taken to ensure that all samples were treated the same way. These initial target acquisition data are not included in this analysis but are available as supplementary material (see S1). Accuracy declined with the introduction of intermittent reinforcement, so intermittent reinforcement was temporarily discontinued until accuracy improved again. The second introduction of intermittent reinforcement continued until the dogs transitioned to testing. Training with intermittent reinforcement corresponded most closely with blind testing conditions, in which blind samples would be interspersed among samples of known status, thereby creating a situation in which the dogs would be working under a variable, intermittent schedule of reinforcement. In the planned blind test conditions, indications of positive blind samples would not be reinforced but indications of positive unblinded samples would be reinforced. Therefore, it was important to establish asymptotic performance under similar conditions.

For the purpose of determining when the dogs were ready to transition to testing, analysis was conducted with the data gathered from Tui, Katie, and Onyx during the first and second phases of intermittent reinforcement (session 19 – 23, 29 – 33+). Rocky started training approximately 150 sessions after the other dogs. His results did not show the same rapid, problematic initial increase in accuracy that the other dogs displayed. Consequently, the target acquisition analysis was applied from the beginning of Rocky's training.

Results

Sensitivity and specificity data from the introduction of intermittent reinforcement training for Tui, Katie, Onyx, and Rocky are shown in Figure 10. The sparser, black data paths display the dogs' accuracy when first exposed to samples from novel patients whereas the grey data paths display the dogs' accuracy with all samples, including those from patients to which the dogs had already been exposed. Trends in the data were not immediately apparent because of the variability in the data. This was particularly pronounced in the data

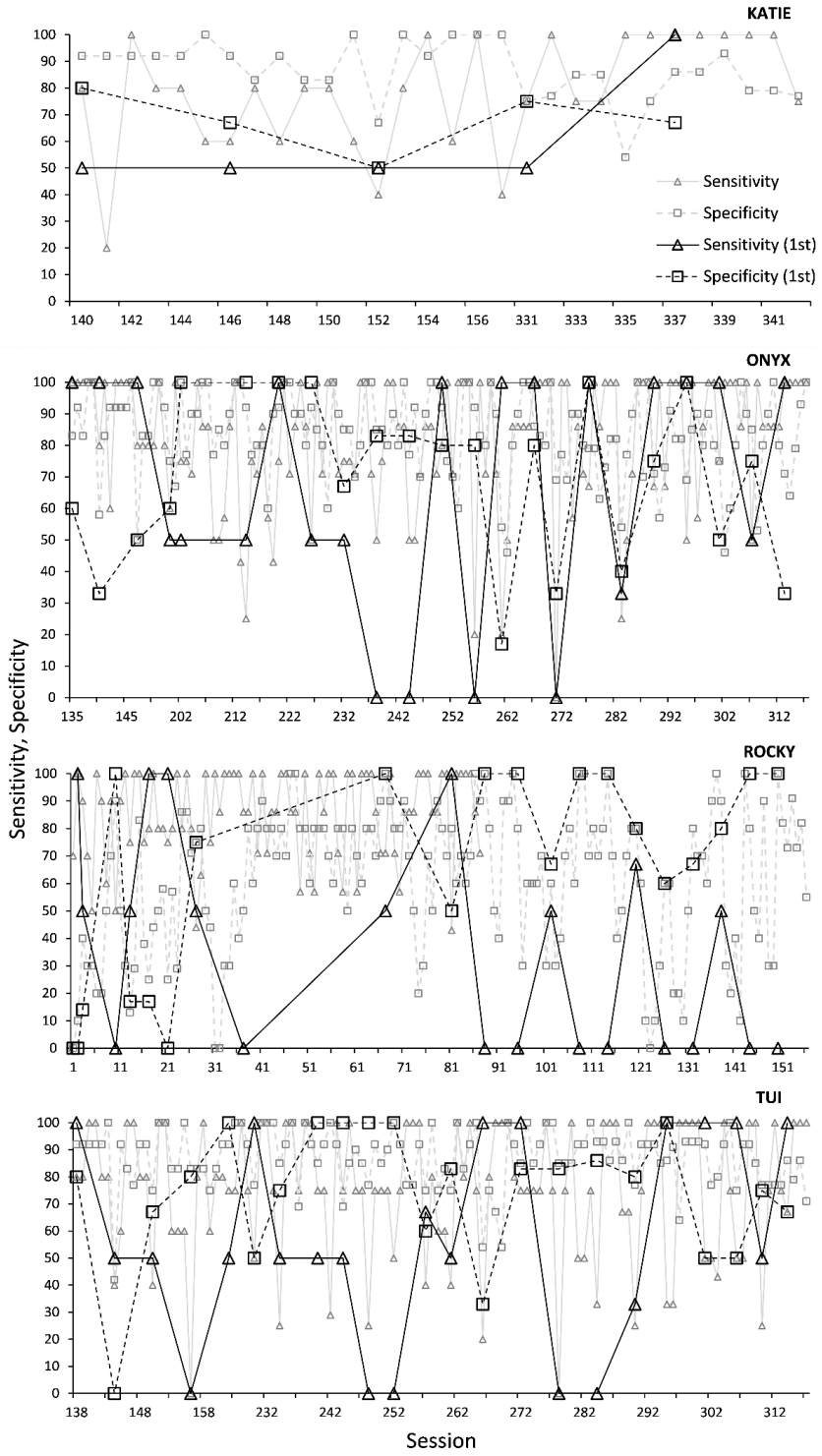
generated by exposure to novel samples as a result of accuracy measures being based on very few samples in each session. Positive samples from novel sources (i.e., from patients that the dogs had not yet encountered) made up, on average, 1.85 (range 0 to 9) out of the 17 samples presented in each session across the first 52 days of training. A mean of 3.12 negative samples from novel sources (range 0 to 8) were included in each sample array, while the remaining positive and negative samples were re-used up to four times or new samples from patients that the dogs had already been exposed to.

Figure 11 displays the mean of sensitivity + specificity generated by Tui, Katie, Onyx, and Rocky's first exposure to samples from novel sources across the same range of sessions displayed in Figure 10. With this summary analysis, it appears that Katie's accuracy was trending upward in later sessions, Tui's data appeared to show a gradual upward trend with later performance generally better than earlier performance, but no trend was apparent in Onyx's data. Rocky's data appeared to trend upward; variability in later sessions made it difficult to determine if his accuracy was reaching asymptote.

Figure 12 displays the Weibull cumulative function applied to the same mean sensitivity + specificity data for Tui, Katie, Onyx, and Rocky. Analysis of the dogs' performances using the Weibull cumulative function allowed for identification of specific slope and inflection values associated with any x-axis value. These were then used to calculate likely asymptote values at a specified gradient (.005) on the upper region of the curve, and the number of sessions required to complete training. Based on a visual analysis of the cumulative function and the underlying data, we used .005 as a working definition of the gradient associated with asymptote. Using this gradient value, estimations of the number of sessions required to reach asymptote (or whether asymptote had already been reached) were possible.

Figure 10

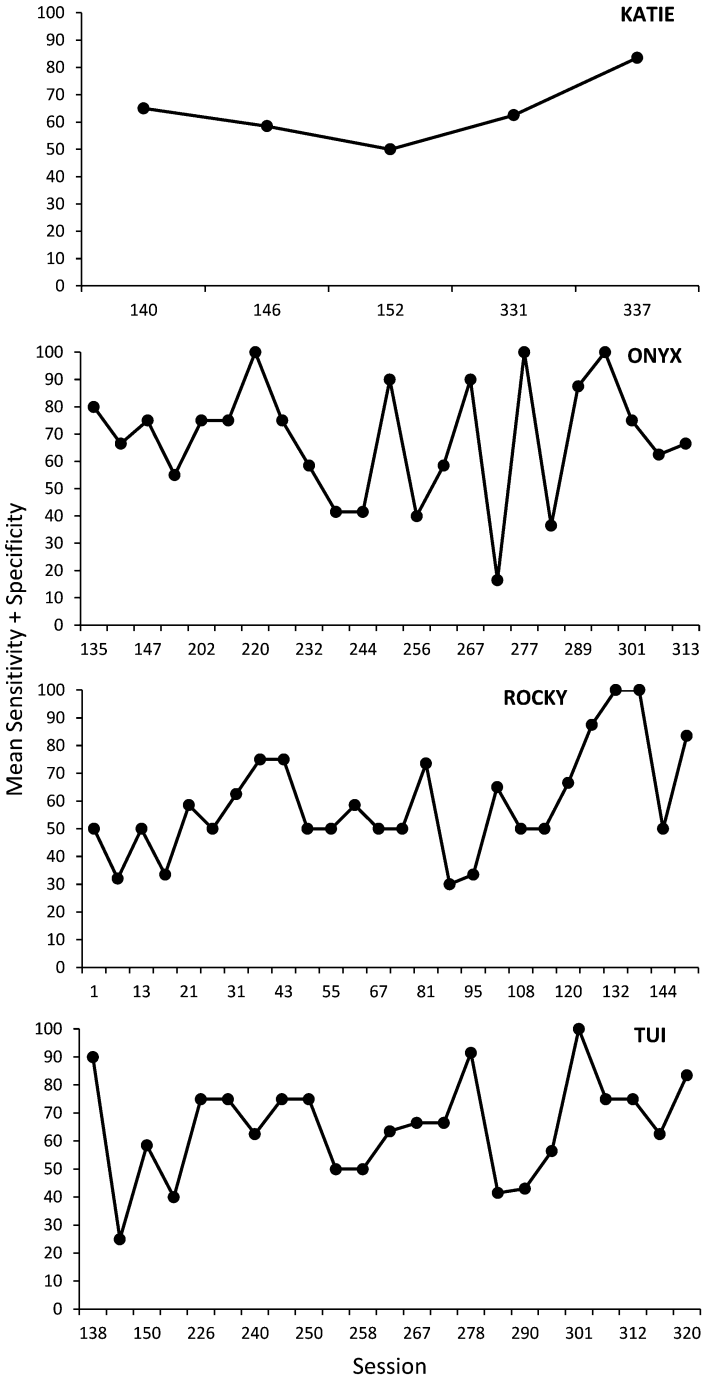
Session Sensitivity and Specificity Percentage Data



Note: Data associated with all samples from each training session are displayed in grey. Dogs' accuracy when they were first (1st) exposed to samples from novel patients is displayed in black.

Figure 11

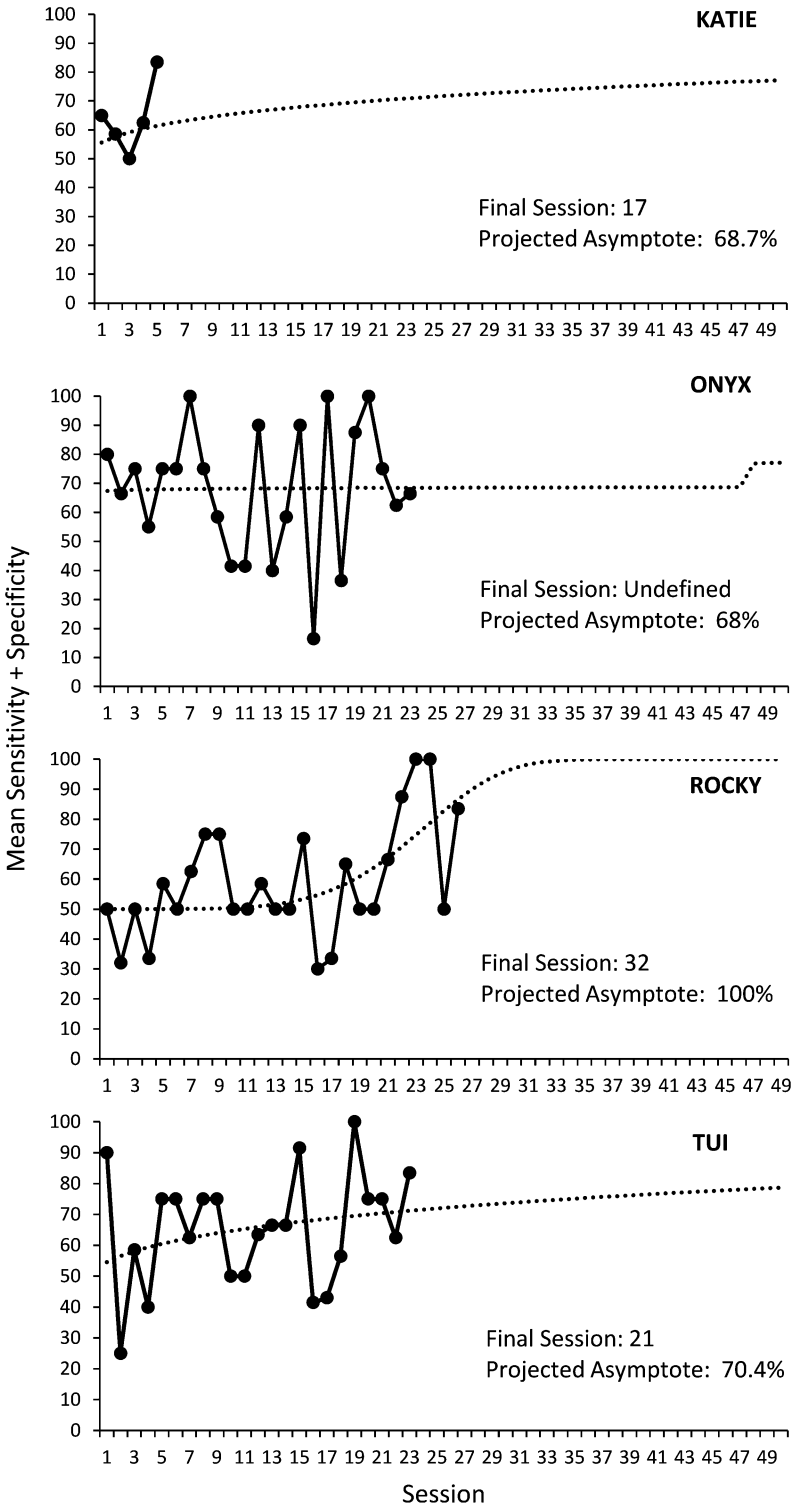
Mean of Novel Sample Response Sensitivity and Specificity ([Sensitivity + Specificity]/2)



Note: The x-axis labels show the session number in which the relevant data were obtained (not all sessions included samples from novel sources).

Figure 12

Weibull Function Applied to Novel Sample Response Mean Sensitivity + Specificity Values



Note: Calculated from exposure to novel source responses.

The data terminate at the point in time when we decided to transition to testing based on these analyses and also due to constraints on sample availability for additional training. Both Tui and Onyx reached asymptote according to the Weibull cumulative function measure, while Rocky would have required approximately six more training sessions, and Katie, twelve. Four dogs completed target acquisition training. See Table 9 for specific details regarding the number and frequency of training sessions for each dog.

Discussion

We evaluated methods of determining when scent-detection dogs are prepared to transition from training to testing for complex target detection. In this case, dogs evaluated breath samples from lung-cancer-positive individuals. Using an automated apparatus, the dogs were trained to identify the presence of lung cancer by evaluating a combination of breath samples from novel sources and previously encountered sources and samples. Evaluation of their ability to identify the target, samples from lung-cancer-positive individuals, was based only on their initial performance with samples from novel sources (i.e., patients). Visual analysis of sensitivity and specificity data separately proved challenging because of the variability of the data. Visual analysis of a summary accuracy measure, the mean sensitivity + specificity, aided with identification of the dogs' progress with target acquisition. The mean sensitivity + specificity values were also analysed by fitting the Weibull cumulative function to these data. The fitted line served as a valuable visual aid. The model was used to predict asymptote values, and predict the number of sessions required to complete training when asymptote was defined as beginning with a gradient of .005. Based on this information and logistical constraints on additional training, the dogs were transitioned to testing at the point where two dogs were already performing at asymptotic levels (according to the Weibull model) and two were predicted to reach asymptote within six to twelve training sessions.

Table 8

Complete Training Session Frequency (Lung Cancer Detection)

Dog	Number of Training Sessions	Mean Number of Training Sessions/Day	Number of Training Half Days Completed
Katie	365	5.60	65
Onyx	345	5.24	65
Rocky	167	5.75	44
Tui	346	5.21	67

The present analysis suggests that these methods, or variations of them, can be fruitfully applied when determining when to transition scent-detection animals from training to testing. We do not suggest that researchers use these data to draw conclusions about the number of sessions that will be required to train dogs to acquire a lung cancer target in other research projects. Idiosyncratic factors such as the number of samples per array, the positive:negative sample ratio, sample characteristics, dog characteristics, and training methods may vary between projects. These factors are likely to influence the amount of training that is required before performance stabilizes. The aim of this work is to demonstrate a method rather than to provide information regarding the number of sessions that are required for dogs to reach stable performance for lung cancer detection.

The data used for this exercise were influenced by several challenges specific to this project. Interruptions to training with work on the laboratory building from December 2019 through February 2020 and the advent of the COVID-19 lockdown in New Zealand from late

March to mid-May 2020 disrupted training and impacted the performance of the dogs on their return to training. These interruptions prompted a reduction in indication thresholds, an increase in the magnitude of reinforcement, and the return of continuous reinforcement until accuracy returned to levels that were observed prior to these disruptions. Other limitations impacting the project included the intermittent supply of novel samples (also related to COVID-19 disruptions), and changes in the health status and availability of individual dogs. Under better circumstances, it is unlikely that the dogs would require training of this duration to reach asymptotic accuracy levels. The data from Rocky provide some evidence of this effect. He started much later than the other dogs but was performing at similar levels by the end of the training phase. Nevertheless, conducting this exercise with these less-than-ideal data may benefit others, as challenges with training are common in applied scent-detection research.

It is highly unlikely that any organism learning a complex target would reach near-perfect performance within as few as 10 sessions, as observed with Katie, Onyx, and Tui. These results suggested that some systematic difference, other than disease status, may have been present at the outset of our study. We were unable to determine what was specifically responsible for this initial high accuracy, but after taking additional measures to eliminate differences between handling of positive and negative samples, results did appear to be more in line with expectations, and we did not observe this pattern in the results of Rocky, who started training later.

After a review of the available models, the Weibull cumulative function appeared to be a suitable model for analyzing mean sensitivity and specificity data for the reasons specified in the Methods section. The model does, however, assume that the data should follow a sigmoidal pattern and, therefore, it did not fit the data well when they were not sigmoidal. For example, when applied to Onyx's data, the sigmoidal element of the model had no predictive

value. Because of the challenges faced when collecting these data, including the initial near-perfect performance, not all data were included when conducting the analyses with three of the four dogs. It is likely that applying the model to data obtained later in training will not be as effective as application from the beginning of training because the model assumes a starting point of 0.5. Nevertheless, this model does appear to provide some additional value when determining when asymptote has been or will be reached, even under suboptimal conditions such as the ones encountered in the process of conducting the present study. Other models may also be useful; researchers considering using a model for this purpose may wish to explore other alternatives. Our use of the summary measure (sensitivity + specificity)/2 is also not without limitations. An advantage of this measure is that sensitivity and specificity are weighted equally (rather than according to the positive:negative sample ratio), but a disadvantage is that it obscures the contribution of sensitivity and specificity to the summary measure. This would be most problematic if one measure was decreasing and the other increasing; these changes would cancel each other out in the summary measure and produce the illusion of stability. Therefore, examining the underlying measures separately before summarizing them would be prudent. The individual measures could both be analysed separately with a quantitative approach as well.

Individual dogs showed different rates of target acquisition as the research progressed, but, for logistical reasons, all dogs needed to move to the testing phase at the same time. This presented the challenge of making group decisions based on disparate individual performance. This challenge was also encountered when making decisions about specific phases of training. While not all dogs may have reached asymptote, it may still be appropriate to begin a new phase or begin to test the entire group. The logistics of training and the relative scarcity of samples from novel sources meant that all of the dogs could only be exposed to 17 training samples per day. The indication thresholds and reinforcement

schedules could be manipulated to meet each dog's individual training needs, but there were significant limitations on the degree to which the sample arrangement could be customized for individual dogs. One strategy that helped to integrate a new dog (Rocky) into the training, although other dogs had progressed to a more advanced stage with fewer positive samples, was to skip some negative samples. This increased the likelihood of the new dog encountering a positive sample and presented more opportunities for reinforcement of positive indications.

Conclusion

The dogs' position on the acquisition curve when they are transitioned to testing appears to be a critical determinant of the accuracy estimates that will be obtained in the blind tests. Therefore, we recommend that researchers who are conducting evaluation research with detection animals use methods such as we have applied herein when deciding to transition from training to testing. We also recommend that a description of the criteria that have been applied to determine when animals are deemed to be ready for testing should be described in reports describing such research.

Acknowledgements

Funding for the experimental procedures associated with this project was provided by the Health Research Council of New Zealand (grant 18/632), the University of Waikato, and the Waikato Medical Research Foundation (grant 281).

We are grateful to the owners of the dogs that participated in this project for supporting the project by making their dogs available to us and reliably transporting them to the laboratory.

Conflict of interest statement: The authors have no conflicts of interest to disclose.

Authorship statement: The experiment was designed by Browne, Crawford, Edwards, and Perrone; the experiment was performed by Hopping and Crawford; the data were analysed by Crawford, Edwards and Perrone; the paper was written by Browne, Chang, Crawford, Edwards, and Perrone.

Animal ethics statement: The experimental procedures described herein comply with international regulations on the ethical treatment of animals and were approved by the University of Waikato Animal Ethics Committee.

References

Amundsen, T., Sundstrøm, S., Buvik, T., Gederaas, O.A., Haaverstad, R., 2014. Can dogs smell lung cancer? First study using exhaled breath and urine screening in unselected patients with suspected lung cancer. Acta Oncol. 53**,** 307-315.

Baer, D.M., Parsonson, B.S., 1981. Applied changes from steady state: Still a problem in the visual analysis of data. In: Bradshaw, C.M., Szabadi, E. & Lowe, C.F., (Eds.) Symposium on recent developments in the quantification of steady-state operant behaviour. Elsevier/North Holland Biomedical Press, Owens Park, Manchester. 273-285.

Barta, J.A., Powell, C.A., Wisnivesky, J.P., 2019. Global epidemiology of lung cancer. Ann. Glob. Health. 85**,** 8.

Blandin, S.K., Crosbie, P.A., Balata, H., Chudziak, J., Hussell, T., Dive, C., 2017. Progress and prospects of early detection in lung cancer. Open Biol. 7**,** 170070.

Browne, C.M., Stafford, K.J., Fordham, R.A., 2015. The detection and identification of tuatara and gecko scents by dogs. J. Vet. Behav. 10**,** 496-503.

Buszewski, B., Ligor, T., Jezierski, T., Wenda-Piesik, A., Walczak, M., Rudnicka, J., 2012. Identification of volatile lung cancer markers by gas chromatography–mass spectrometry: Comparison with discrimination by canines. Anal. Bioanal. Chem. 404**,** 141-146.

Deliano, M., Tabelow, K., König, R., Polzehl, J., 2016. Improving accuracy and temporal resolution of learning curve estimation for within- and across-session analysis. PLoS ONE, 11**,** 0157355.

Edwards, T. L., 2019. Automated canine scent-detection apparatus: Technical description and training outcomes. Chem. Senses, 44**,** 449-455.

Edwards, T.L., Browne, C.M., Schoon, A., Cox, C., Poling, A. 2017. Animal olfactory detection of human diseases: Guidelines and systematic review. J. Vet. Behav. 20**,** 59–73.

Edwards, T.L., Giezen, C., & Browne, C.M. 2022. Influences of indication response requirement and target prevalence on dogs' performance in a scent-detection task. Appl. Anim. Behav. Sci., 105657.

Ehmann, R., Boedeker, E., Friedrich, U., Sagert, J., Dippon, J., Friedel, G., Walles, T., 2012. Canine scent detection in the diagnosis of lung cancer: Revisiting a puzzling phenomenon. Eur. Respir. J. 39**,** 669-676.

Elliker, K., Sommerville, B., Broom, D., Neal, D., Armstrong, S., Williams, H., 2014. Key considerations for the experimental training and evaluation of cancer odour detection dogs: Lessons learnt from a double-blind, controlled trial of prostate cancer detection. BMC Urol. 14**,** 22.

Ellis, H., Mulder, C., Valverde, E., Poling, A., Edwards, T.L., 2017. Reproducibility of African giant pouched rats detecting *Mycobacterium tuberculosis*. BMC Infect. Dis., 17, 298.

Fischer-Tenhagen, C., Johnen, D., Nehls, I., Becker, R., 2018. A proof of concept: Are detection dogs a useful tool to verify potential biomarkers for lung cancer? Front. Vet. Sci. 5**,** 52.

Gallistel, Fairhurst, S., Balsam, P., 2004. The learning curve: Implications of a quantitative analysis. Proceedings of the National Academy of Sciences of the United States of America, 101**,** 13124-13131.

Grier, J.B., 1971. Nonparametric indexes for sensitivity and bias: Computing formulas. Psychol. Bull. 75**,** 424-429.

Hall, N.J., 2017. Persistence and resistance to extinction in the domestic dog: Basic research and applications to canine training. Behav. Processes, 141**,** 67-74.

Hautus, M., 2015. Signal detection theory. In: Wright, J.D. (Ed.) International Encyclopedia of the Social & Behavioral Sciences (2nd ed.). Elsevier.

Herrnstein, R.J., Loveland, D.H., 1964. Complex visual concept in the pigeon. Science, 146**,** 549-550.

Herrnstein, R.J., Loveland, D.H., Cable, C. 1976. Natural concepts in pigeons. J. Exp. Psychol. Anim. Behav. Process. 2**,** 285-302.

Hung, R., Lee, S., Bennett, J., 2015. Fungal volatile organic compounds and their role in ecosystems. Appl. Microbiol. Biotechnol. 99**,** 3395-3405.

Jakovcevic, A., Elgier, A., Mustaca, A., Bentosela, M., 2013. Frustration behaviors in domestic dogs. J. Appl. Anim. Welf. Sci. 16**,** 19-34.

Jamieson, L.T.J., Baxter, G.S., Murray, P.J., 2017. Identifying suitable detection dogs. Appl. Anim. Behav. Sci. 195**,** 1-7.

Jezierski, T., Walczak, M., Ligor, T., Rudnicka, J., Buszewski, B., 2015. Study of the art: canine olfaction used for cancer detection on the basis of breath odour. Perspectives and limitations. J. Breath. Res. 9**,** 027001.

Johnen, D., Heuwieser, W., Fischer-Tenhagen, C., 2013. Canine scent detection—fact or fiction? Appl. Anim. Behav. Sci. 148**,** 201-208.

Johnen, D., Heuwieser, W., Fischer-Tenhagen, C., 2017. An approach to identify bias in scent detection dog testing. Appl. Anim. Behav. Sci. 189**,** 201-208.

Johnston, J., Pennypacker, H., Green, G., (Eds). 2019. Strategies and Tactics of Behavioral Research and Practice. Routledge.

Jones, R.R., Weinrott, M.R., Vaught, R.S., 1978. Effects of serial dependency on the agreement between visual and statistical inference. J. Appl. Behav. Anal. 11**,** 277-283.

Kokocińska-Kusiak, A., Woszczyło, M., Zybala, M., Maciocha, J., Barłowska, K. & Dzięcioł, M. 2021. Canine olfaction: physiology, behavior, and possibilities for practical applications. Animals (Basel), 11**,** 2463.

Lane, J.D., Gast, D.L., 2014. Visual analysis in single case experimental design studies: brief review and guidelines. Neuropsychol. Rehabil. 24**,** 445-463.

Lazarowski, L., Rogers, B., Waggoner, L.P., Katz, J.S., 2019. When the nose knows: ontogenetic changes in detection dogs' (*Canis familiaris*) responsiveness to social and olfactory cues. Anim. Behav. 153**,** 61-68.

Lazarowski, L., Krichbaum, S., DeGreeff, L.E., Simon, A., Singletary, M., Angle, C. & Waggoner, L.P. 2020. Methodological considerations in canine olfactory detection research. Front. Vet. Sci. 7**,** 408-408.

Macmillan, N., Creelman, C., 1990. Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. Psychol. Bull. 107**,** 401-413.

Macmillan, N.A., Creelman, C.D., 2004. Detection theory: a user's guide. Taylor & Francis Group.

Mazur, J.E., Hastie, R., 1978. Learning as accumulation: a reexamination of the learning curve. Psychol. Bull. 85**,** 1256-1274.

McCulloch, M., Jezierski, T., Broffman, M., Hubbard, A., Turner, K., Janecki, T., 2006. Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. Integr. Cancer Ther. 5**,** 30-39.

Mortensen, U., 2002. Additive noise, Weibull functions and the approximation of psychometric functions. Vision Res. 42**,** 2371-2393.

Murre, J.M.J., 2014. S-shaped learning curves. Psychon. Bull. Rev. 21**,** 344-356.

National Institute of Health Research. 2018. Largest ever UK lung cancer screening study launches in London [WWW Document]. URL https://www.nihr.ac.uk/news/largest-ever-uk-lung-cancer-screening-study-launches-in-london/9808 (05. 09. 19)

Nourbakhsh, M.R., Ottenbacher, K.J., 1994. The statistical analysis of single-subject data: a comparative examination. Phys. Ther. 74**,** 768-776.

Pirrone, F., Albertini, M., 2017. Olfactory detection of cancer by trained sniffer dogs: a systematic review of the literature. J. Vet. Behav. 19**,** 105-117.

Pleil, J. & Giese, R. 2017. Integrating exhaled breath diagnostics by disease-sniffing dogs with instrumental laboratory analysis. J. Breath Res. 11**,** 032001.

Rudnicka, J., Walczak, M., Kowalkowski, T., Jezierski, T., Buszewski, B., 2014. Determination of volatile organic compounds as potential markers of lung cancer by gas chromatography–mass spectrometry versus trained dogs. Sens. Actuators B: Chem. 202**,** 615-621.

Saalberg, Y. & Wolff, M. 2016. VOC breath biomarkers in lung cancer. Clin. Chim. 459**,** 5-9.

Stevens, M., 2013. Sensory ecology, behaviour, and evolution. Oxford University Press.

Teodoro-Morrison, T., Diamandis, E.P., Rifai, N., Weetjens, B.J.C., Pennazza, G., De Boer, N.K., Bomers, M.K., 2014. Animal olfactory detection of disease: promises and pitfalls. Clin. Chem. 60**,** 1473-1479.

Walczak, M., Jezierski, T., Górecka-Bruzda, A., Sobczyńska, M., Ensminger, J., 2012. Impact of individual training parameters and manner of taking breath odor samples on the reliability of canines as cancer screeners. J. Vet. Behav. 7**,** 283-294.

World Health Organization. 2019. Cancer [WWW Document]. URL https://www.who.int/en/news-room/fact-sheets/detail/cancer (09. 12. 18)

Zhan, S., Ottenbacher, K., 2001. Single subject research designs for disability research. Disabil. Rehabil. 23**,** 1-8.

Zhang, J., Mueller, S.T., 2005. A note on ROC analysis and non-parametric estimate of sensitivity. Psychometrika, 70**,** 203-212.

Zhou, X.-H., 2011. Statistical methods in diagnostic medicine. Wiley.

Chapter Five

General Discussion

Practical contribution summary

Our project is a novel contribution to the body of literature on this subject. It generated both practical and theoretical findings: using an automated apparatus and operant conditioning, dogs can learn a lung cancer-positive scent target more quickly and accurately with breath samples than saliva samples. We discovered that dogs can detect a lung cancer-positive target in saliva samples at higher than chance levels. We found that breath samples can be used for lung cancer-positive scent detection training purposes up to four times without significant compromise in the dogs' performance. Also, we learnt that a quantitative model could be applied to training data to determine likely maximum performance for individual dogs. This analysis was used to inform our decision to transition the dogs from training to testing the detection of lung cancer using breath samples from patients with respiratory illness. These findings suggest qualified support for the use of dogs for lung cancer screening purposes.

Given the results of previous lung cancer detection studies using breath samples (Buszewski et al., 2012; Fischer-Tenhagen et al., 2018; Guirao et al., 2019; Guirao Montes et al., 2017; McCulloch et al., 2006; Rudnicka et al., 2014; Walczak et al., 2012) we predicted that using breath samples to train the dogs to detect lung cancer was likely to result in above-chance sensitivity and specificity. This proved to be the case. The emergence of saliva as a viable sample type for electronic disease scent detection (Bel'skaya et al., 2020; Chakraborty et al., 2018; Katakura et al., 2015; Ma et al., 2020; Printz, 2012; Qian et al., 2018; Zhang et al., 2012) suggested that saliva too might be a viable medium for disease detector organism training. Our experiment results do not rule out saliva as a potential training medium, but it became impractical for our project to continue training using both sample types once the difference in target acquisition rates became clear. There are a variety of different collection, storage and presentation options available to researchers that may facilitate quicker

acquisition of a disease scent target using saliva and prove more efficacious than our procedures.

Our exploration of the use and re-use of breath samples during scent detection training was informed by two ideas with potentially contradictory outcomes: 1) VOCs dissipate at room temperature, so are likely to lose their potency in breath samples that are repeatedly opened, closed and stored (Wang et al., 2018); and 2) learning effects occur when an organism is repeatedly trained using the same sample (Walczak et al., 2012). We predicted that it was likely that the potency of individual breath samples would decrease over time, creating a situation where dogs were more likely to provide a false negative indication. However, we also predicted that the repeated presentation of specific samples to the dogs would be likely to improve the probability of correct positive or negative indications because the individual scent profile would be learnt through repetition. The extent to which these two dueling processes had been explored in previous experiments was unclear. Response accuracy during repeated exposure to a single sample over a day had been evaluated (Walczak et al., 2012), but no research on the acquisition of a disease scent target by dogs via repeated sample use and extended breaks between exposures had been published at the time our experiment was designed. Our findings seem to suggest that sample decay and learning effects can cancel each other out to an extent during target acquisition training. However, further research is required to confirm this and to clarify the extent of the contribution of each effect.

Using an automated apparatus during training optimized aspects of our work. After successfully shaping each dog's initial interactions with the apparatus, the trainer was no long required in the experiment room. Because the trainer was no longer in the room, the possibility of trainer cueing was eliminated in all subsequent training. Automation allowed the team to identify and apply individual dog's optimal indication thresholds (Edwards, 2019;

Edwards et al., 2022). Automation and the development of standard operating procedures also meant that our experiments were standardized. Standardization increased the chance of future teams being able to precisely replicate our methods and results.

We predicted that the dogs would learn to discriminate lung cancer-positive samples from negative samples with increasing accuracy following a sigmoid performance curve (Murre, 2014). The dogs' acquisition data broadly conformed with this prediction. The range of asymptotic performance identified in article three (Chapter 4) for dogs with extensive training clustered around 70% (mean sensitivity + specificity). A question our project raises for further research is the extent to which each dog's individual performances could be interpreted usefully and accurately in a group. For disease screening purposes, having a team of dogs where the team is treated as a single diagnostic tool, may offer advantages over using only a single dog, There is precedent for such an approach in the tuberculosis screening programmes using teams of giant African pouched rats by APOPO (Bauër et al., 2022).

Screening: Individual and group performance

While average group behaviour is rarely representative of individual behaviour (Adi-Japha et al., 2008; Pamir et al., 2011), there could be advantages to assessing canine group performance as disease detectors using arbitrary cut-off values. The cut-offs could decide what proportion of individual dog responses to a sample must agree with each other before the sample status is fixed. For example, in a team of 4 trained dogs, a cut-off value of 75% could be applied. This would mean that if at least 3 out of 4 of the dogs provided a positive indication in the presence of a sample, the sample status would be identified as positive.

Theoretical contribution summary

Our project addressed theoretical questions about the nature of concept formation and target acquisition. All three experiments concurrently monitored the acquisition of a complex concept, while experiment three focused specifically on the analysis of target concept

acquisition. Each project used samples obtained from participants in a clinical context: the dogs were presented with training samples drawn from the same pool as test samples. While the gender, smoking status and disease status of each participant was recorded, individual characteristics including co-occurring illnesses, current treatment, and diet were not. During multiple exemplar/sample training, we assessed each dog's ability to form a complex concept of the lung cancer scent target and respond accurately to both the "signal + noise" of a true lung cancer scent and the "noise" of extraneous VOCs occurring at random in each new sample.

Concept formation and target acquisition

The acquisition of simple concepts is abrupt (Gallistel et al., 2004). The acquisition of complex concepts involving generalisation of discriminable feature identification to unfamiliar contexts and media is much more nuanced and likely to require exposure to a large range of exemplars (Bhatt et al., 1988; Herrnstein et al., 1976; Keller et al., 1951). In some cases, previous research on canine scent detection of lung cancer focused on targeting specific high frequency VOCs. These VOCs were synthetically generated or analysed in breath samples using mass spectrometry (Buszewski et al., 2012; Fischer-Tenhagen et al., 2018; Rudnicka et al., 2014; Walczak et al., 2012). Being able to identify "marker" VOCs for lung cancer with a high level of accuracy could certainly be a basis of effective screening. However, chemical analysis of biological samples suggests that reliable marker VOCs for this disease do not exist in concentrations available for screening using current technology (Saalberg & Wolff, 2016; Schallschmidt et al., 2016; Shirasu & Touhara, 2011; Zhan et al., 2020). If marker VOCs do not provide a viable explanation for the above-chance lung cancer scent detection shown in previous research, what does?

Wasserman (2016) posited the following two theoretical assumptions about complex concept acquisition in human and non-human animals. Firstly, he suggested that samples or

exemplars being used to train an organism to indicate the presence of a target are likely to have many common elements. After initial training, the detector organism has encountered the elements in a range of other positive samples. During each of these encounters, reinforcement has been used to increase desired responding in the presence of target elements. Therefore, when an organism can perceive several of these (category-specific) elements in a new sample, the probability of a target indication is high. Alternatively, when target elements in samples occur infrequently or are present but difficult to perceive, the probability of an organism indicating the presence of the target is low. These lower-frequency elements are stimuli-specific. Wasserman's second assumption was that complex concepts are developed by strengthening and refining connections between active stimulus elements and organism responding via "error-driven learning". Theoretically, a detector organism responds to a sample on the basis of perceived available elements that have produced the richest reinforcement for responding in the past. If there is a mismatch during training between an organism's response to available elements in a novel sample and the true sample status, the difference between expected and encountered reinforcement (the organism's "prediction error") will shape future behaviour and thus, categorization. In the context of our project, it appeared that the concept-specific elements of lung cancer scent in breath samples were available to dogs more frequently during training than in saliva samples. The repeated use of breath samples during training did not diminish the samples' effectiveness, suggesting that either the individual scent profile of each sample was learnt by the dogs through repeated exposure, or that elements available in the samples for scent detection did not change significantly after each use. Finally, concept-specific characteristics of lung cancer scent in breath samples may be more readily available to dogs when compared with saliva samples, as the dogs' acquired the former target more quickly than the latter.

Learning theory and target acquisition

It is assumed in behavioural science that complex processes and concepts are composed of much simpler, foundational behavioural phenomena that can be discovered through forms of component/composite analysis (Cooper et al., 2020). Learning in the context of the laboratory is precipitated through planning and manipulating the presentation of stimuli, the arrangement of consequences for responses, signaling stimuli, signaling consequences and established the effectiveness of consequences (Catania, 2013). In less controlled circumstances, teasing out the extent to which acquired concepts and skills are setting-specific, and the range of elements under faulty stimulus-control during learning acquisition, is astonishing complex (Kantor, 1988). Underpinning disease scent detection target acquisition lies associative learning theory. Associative learning theory suggests that learning occurs when exposure of an organism to a stimulus changes the perceptual effectiveness of that stimulus (Hall, 2003). When characterized as a single-state phenomenon, associative learning and concomitant changes in responding can be represented accurately with a learning curve. Alternative analyses of associative learning suggest that multiple aspects of learning context are acquired at the same time by animals, and produce step-wise patterns of 'all or nothing' responding that are much more abrupt than previously thought (Blanco & Moris, 2018). Acquisition of new behaviours occurs through conditioning processes – respondent or operant - consistent with associative learning theory. Learning varies depending on the number of opportunities the organism requires before new responding appears, how quickly the organism can attain maximum performance, and the level at which maximum performance occurs. These three features of learning can often be represented with a reasonable degree of fidelity via an algorithm or a learning curve (Gallistel et al., 2004). We fit a curve to our data based on their shape and limits, informed by theoretical assumptions outlined earlier.

Applications to Lung Cancer Screening

Clearly, there is still a great deal to be learnt about quantifying the acquisition of complex concepts. The first experiment in our project generated data showing a quicker acquisition of the lung cancer target using breath samples when compared with saliva samples. This suggests that there could be more concept-specific features of the scent of lung cancer available to dogs in breath samples than saliva samples. Applying concept formation theory, it is possible that the features of lung cancer scent in saliva are less salient to dogs, or that more "noise" is present in samples.

The second experiment explored the extent to which sample decay through re-use can be off-set by the learning effects of repeated exposure. The design does not allow for separation of degradation and learning effects. Information from this design is still useful because, although the dogs' previous exposure to individual samples could enhance their subsequent performance with those samples, there would need to be sufficient volatiles available for them to "recognise" the sample when encountered again. Although it is possible to design a study that could separate these effects (e.g., by opening samples but not using them), this was financially and logistically infeasible in our project. We hope that this may be possible to follow up on in the future by a group with sufficient resources to address this specific issue.

Our finding that breath samples could be re-used up to four times without a significant reduction in the dogs' sensitivity reinforced the theory that concept-specific repetition supports complex concept formation. It is possible but unlikely that the dogs learnt each individual scent after brief exposure several weeks apart. However, it is much more likely that the dogs had learnt sufficient concept-specific features that they recognised these in the available VOCs in repeatedly opened samples.

Concept acquisition was explicitly quantified in our third experiment. The analysis developed in this work was potentially the most impactful of the three, as it could be applied to any learning context, or any form of disease screening where hit, miss, false alarm and correct reject data are collected. The results of this experiment conform with both acquisition modelling using a sigmoid representation of the learning curve, and concept formation through asymptotic responding at values well above chance.

Project limitations

Many of the limitations associated with our project were logistical. Sample supply and analysis issues were prompted by the COVID-19 pandemic. Building renovation and nationwide lockdowns preventing access to the laboratory. Automated apparatus mechanical faults and dog health issues each impeded data collection.

Our project relied on a supply of breath and saliva samples from a respiratory clinic at a local hospital. The commitments of the clinicians associated with the project were wide ranging. Dog training needed to be suspended more than once to allow the clinicians time to adjudicate participants' health and assign 'positive', 'negative' or 'unknown' lung cancer status to samples provided. This adjudication process became exponentially more difficult with the advent of COVID-19 and the explosion of responsibilities weighing on our colleagues as respiratory specialists.

The laboratory building was closed for two months while scheduled renovations took place. Soon after this, the global COVID-19 pandemic sparked another shutdown of the laboratory and a lockdown of all parties involved for further two months. These extended interruptions to training had a significant impact on dog performance and the systems involved in maintaining the project.

More frequent, short-term but equally intrusive interruptions occurred as a result of the automated apparatus faults. These were not unexpected, as the apparatus was a prototype. No

models were available to us of standard operating procedures for the training and testing of dogs using an automated apparatus. These evolved through several iterations as we determined how to optimize the dogs' responding. Researchers became adept at resetting aspects of the machine, while the technician involved in the prototype design was often quickly on hand to address any ongoing issues.

The wellbeing of the dogs and the commitments of their owners determined the extent to which they were available for training on a daily basis. Each dog's participation was functionally assent-based. No coercive consequences for non-training behaviours were provided at any stage in the project, so a "disengaged" dog could effectively end training at will. Two of the dogs experienced extended periods of ill health, while another dog developed an aversion to the feeder. Care was taken to monitor each dog's weight. In the case of three dogs, adjustments were made to diets outside of the laboratory where it became apparent that they were gaining weight while participating in the project. Changes in feeding routines at home allowed the dog's weight to reduce and then stabilize in each instance.

In short, the research team fostered ongoing relationships with both campus technicians and dog owners that accommodated short-term plan changes while minimizing disruption to training as much as possible.

Beyond practical considerations, we were unable to isolate the source of contamination that influenced the dogs' initial acquisition accuracy. It was difficult to gauge the level of sensitivity and specificity of previous canine scent detection work with a lung cancer target as the methodologies of previous projects varied so significantly. Because we had no closely-corresponding models against which to compare the early acquisition results, it was difficult to be certain that our procedure was flawed in some way. We needed to rely on concept formation acquisition data across a range of detector organisms and targets to confirm the unlikely nature of our dogs' early performance.

Further investigation of target acquisition could be developed using an alternative approach to capturing missed positive samples. Because we dropped the indication threshold of any lung-cancer positive sample the dog "missed" during a session, only the first response provided an accurate picture of the dog's performance. The manipulation of the threshold in subsequent sessions meant that we could not analyse the dogs' evaluation of samples during the course of one training day.

Some features of our experiments could be characterized as either limitations or strengths, depending on the expectations of a reviewer. For example, we premised the search for detector organisms on the understanding that there is no such thing as a "scent detection dog". Single subject research methodology allowed us to focus instead on each dog's individual (and potentially idiosyncratic) target acquisition and responding over time, comparing each animal's performance with itself. Applying screening criteria that had been developed for other scent-detection applications (e.g., for border security applications) would not be appropriate for this novel method and setting. Recruiting dogs that had been trained using other methods would be similarly problematic. All training was specific to our unique automated approach and dog suitability was assessed with respect to this approach. While some breeds would be physically mismatched with the apparatus (e.g., pugs, bull dogs, mastiffs), we did not select for breed. Using pets provided us with two advantages: 1) high social validity – dog owners described the "enthusiasm" of their dogs to arrive at the laboratory for training, and 2) a much larger pool of potential trainees. Because we used pet dogs across a range of breeds, our data were potentially more broadly applicable to dogs generally.

Another potentially challenging feature of our research involved the breath and saliva samples. Dog-related VOC contamination was a limitation of the research conducted in our experiments. Although we had the means of controlling for contamination for the purpose of

operational/blind testing of the dogs (e.g., we only paid attention to the dogs' first encounter with novel samples from novel patients and we changed the order in which the dogs trained using a random number generator each day), it wasn't possible to control for or eliminate this issue in the present investigation. However, in line with the applied aims of this study, dog-related VOC contamination would be a standard issue when samples are re-used and, therefore, we have produced data that are relevant to the applications that we are most interested in understanding.

Our project relied upon the application of single-subject research design for the optimization of practices in a field without any current gold standard practice or universally recognised methodology or measures. There is potentially a strong argument to be made of a burgeoning of single-subject research in this field before any meaningful meta-analysis is possible.

Final reflection

Effective screening for the early detection of lung cancer could increase the efficacy of treatment and survivability of this disease. From a theoretical basis informed by learning acquisition and concept formation, we were able to ascertain that dogs learn a lung cancer target scent from breath samples more easily than saliva samples; that breath samples can be used up to four times for training purposes without significantly impacting sensitivity; and quantitative measures can be generated to standardize the transition from training to testing. On the basis of what we have learnt in our project about dogs' ability to acquire complex concepts like the scent of lung cancer, it is not necessary for dogs to perform like either an eNose or a mass spectrometer to provide effective screening. However, they can certainly perform in tandem with other technologies. The chemical analysis of a breath sample using technology is a task distinct from the discrimination performed by a dog during scent detection. Using a group of dogs as a single diagnostic tool for lung cancer screening is a

potentially low-risk, non-invasive option to increase the chances of early disease detection. The costs and absolute risks of not being screened need to be compared against both the risks of higher than necessary exposure to x-rays or biopsy via more intrusive screening, and the maximum demonstrated sensitivity and specificity of dogs as lung cancer detectors. Increasing the power of sample evaluation by providing weighted analysis of group results, as described earlier, may be one way of tipping the balance in favour of canine scent detection.

References[6]

Adi-Japha, E., Karni, A., Parnes, A., Loewenschuss, I., & Vakil, E. (2008). A shift in task routines during the learning of a motor skill: Group-averaged data may mask critical phases in the individuals' acquisition of skilled performance. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(6), 1544-1551. https://doi.org/10.1037/a0013217

Alasaad, S., Permunian, R., Gakuya, F., Mutinda, M., Soriguer, R., & Rossi, L. (2012). Sarcoptic-mange detector dogs used to identify infected animals during outbreaks in wildlife. *BMC Veterinary Research*, *8*, 110. https://doi.org/10.1186/1746-6148-8-110

Amundsen, T., Sundstrøm, S., Buvik, T., Gederaas, O. A., & Haaverstad, R. (2014). Can dogs smell lung cancer? First study using exhaled breath and urine screening in unselected patients with suspected lung cancer. *Acta Oncologica*, *53*(3), 307-315. https://doi.org/10.3109/0284186X.2013.819996

Angle, C., Waggoner, L. P., Ferrando, A., Haney, P., & Passler, T. (2016). Canine detection of the volatilome: A review of implications for pathogen and disease detection. *Frontiers of Veterinary Science*, *3*, 47. https://doi.org/10.3389/fvets.2016.00047

Baer, D. M., & Parsonson, B. S. (1981). Applied changes from steady state: Still a problem in the visual analysis of data. *Symposium on Recent Developments in the Quantification of Steady-State Operant Behaviour*, Owens Park, Manchester.

Baker, S. G. (2003). The central role of Receiver Operating Characteristic (ROC) curves in evaluating tests for the early detection of cancer. *JNCI: Journal of the National Cancer Institute*, *95*(7), 511-515. https://doi.org/10.1093/jnci/95.7.511

[6] References not included in the reference lists of the relevant manuscripts

Bauër, P., Leemans, M., Audureau, E., Gilbert, C., Armal, C., & Fromantin, I. (2022). Remote medical scent detection of cancer and infectious diseases with dogs and rats: A systematic review. *Integrative Cancer Therapies*, *21*, 15347354221140516-15347354221140516. https://doi.org/10.1177/15347354221140516

Bel'skaya, L. V., Sarf, E. A., Kosenok, V. K., & Gundyrev, I. A. (2020). Biochemical markers of saliva in lung cancer: Diagnostic and prognostic perspectives. *Diagnostics*, *10*(4). https://doi.org/10.3390/diagnostics10040186

Bhatt, R. S., Wasserman, E. A., Reynolds, W. F., & Knauss, K. S. (1988). Conceptual behavior in pigeons: Categorization of both familiar and novel examples from four classes of natural and artificial stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*(3), 219-234. https://doi.org/10.1037/0097-7403.14.3.219

Blanco, F., & Moris, J. (2018). Bayesian methods for addressing long-standing problems in associative learning: The case of PREE. *Quarterly Journal of Experimental Psychology*, *71*(9), 1844-1859. https://doi.org/10.1080/17470218.2017.1358292

Blandin, S. K., Crosbie, P. A., Balata, H., Chudziak, J., Hussell, T., & Dive, C. (2017). Progress and prospects of early detection in lung cancer. *Open Biology*, *7*(170070). https://doi.org/http://doi.org/10.1098/rsob.170070

Bomers, M. K., van Agtmael, M. A., Luik, H., van Veen, M. C., Vandenbroucke-Grauls, C. M. J. E., & Smulders, Y. M. (2012). Using a dog's superior olfactory sensitivity to identify *Clostridium difficile* in stools and patients: Proof-of-principle study. *BMJ (Online)*, *345*(7888), 7-9. https://doi.org/10.1136/bmj.e7396

Bomers, M. K., van Agtmael, M. A., Luik, H., Vandenbroucke-Grauls, C. M. J. E., & Smulders, Y. M. (2014). A detection dog to identify patients with *Clostridium difficile*

infection during a hospital outbreak. *The Journal of Infection*, *69*(5), 456-461. https://doi.org/10.1016/j.jinf.2014.05.017

Bosch, S., Berkhout, D. J., Ben Larbi, I., de Meij, T. G., & de Boer, N. K. (2019). Fecal volatile organic compounds for early detection of colorectal cancer: Where are we now? *Journal of Cancer Research and Clinical Oncology*, *145*(1), 223-234. https://doi.org/10.1007/s00432-018-2821-3

Bryce, E. M. D. F., Zurberg, T. R. C. T., Zurberg, M. R. N., Shajari, S. B., & Roscoe, D. M. D. F. (2017). Identifying environmental reservoirs of *Clostridium difficile* with a scent detection dog; Preliminary evaluation. *The Journal of Hospital Infection*, *97*(2), 140-145. https://doi.org/10.1016/j.jhin.2017.05.023

Buszewski, B., Ligor, T., Jezierski, T., Wenda-Piesik, A., Walczak, M., & Rudnicka, J. (2012). Identification of volatile lung cancer markers by gas chromatography–mass spectrometry: Comparison with discrimination by canines. *Analytical and Bioanalytical Chemistry*, *404*(1), 141-146. https://doi.org/10.1007/s00216-012-6102-8

Byosiere, S.-E., Feng, L. C., & Rutter, N. J. (2019). Factors that may affect the success of scent detection dogs: Exploring non-conventional models of preparation and deployment. *Comparative Cognition & Behavior Reviews*, *14*, 81-86. https://doi.org/10.3819/CCBR.2019.140009

Cambau, E., & Poljak, M. (2020). Sniffing animals as a diagnostic tool in infectious diseases. *Clinical Microbiology and Infection*, *26*(4), 431-435. https://doi.org/10.1016/j.cmi.2019.10.036

Campbell, L. F., Farmery, L., George, S. M. C., & Farrant, P. B. J. (2013). Canine olfactory detection of malignant melanoma. *BMJ Case Reports*. https://doi.org/10.1136/bcr-2013-008566

Cantor, B. (2020). *The Equations of Materials*. Oxford University Press.

Catala, A., Grandgeorge, M., Schaff, J.-L., Cousillas, H., Hausberger, M., & Cattet, J. (2019). Dogs demonstrate the existence of an epileptic seizure odour in humans. *Scientific Reports*, *9*(1), 4103-4103. https://doi.org/10.1038/s41598-019-40721-4

Catania, A. C. (2013). *Learning* (5th ed.). Sloan Publications.

Centre for Disease Control United States of America. (2018). *Who should be screened for lung cancer?* Retrieved July 19, 2019, from https://www.cdc.gov/cancer/lung/basic_info/screening.htm

Chakraborty, D., Viveka, T. S., Arvind, K., Shyamsundar, V., Kanchan, M., Alex, S. A., Chandrasekaran, N., Vijayalakshmi, R., & Mukherjee, A. (2018). A facile gold nanoparticle–based ELISA system for detection of osteopontin in saliva: Towards oral cancer diagnostics. *Clinica Chimica Acta*, *477*, 166-172. https://doi.org/10.1016/j.cca.2017.09.009

Charles, M. K., Wang, Y., Zurberg, T., Kinna, J., & Bryce, E. (2019). Detecting *Clostridioides (Clostridium) difficile* using canine teams: What does the nose know? *Infection Prevention in Practice*, *1*(1), 100005-100005. https://doi.org/10.1016/j.infpip.2019.100005

Church, J., & Williams, H. (2001). Another sniffer dog for the clinic? *The Lancet*, *358*(9285), 930. https://doi.org/10.1016/S0140-6736(01)06065-2

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied Behavior Analysis* (3rd ed.). Pearson Education, Inc.

Cornu, J.-N., Cancel-Tassin, G., Ondet, V., Girardet, C., & Cussenot, O. (2010). Olfactory detection of prostate cancer by dogs sniffing urine: A step forward in early diagnosis. *European Urology*, *59*(2), 197-201. https://doi.org/10.1016/j.eururo.2010.10.006

Crawford, M. A., Chang, C. L., Hopping, S., Browne, C. M., & Edwards, T. L. (2023). Influences of breath sample re-use on the accuracy of lung cancer detection dogs. *Journal of Breath Research, 17*(1), 16001. https://doi.org/10.1088/1752-7163/ac9b7f

Crawford, M. A., Perrone, J. A., Browne, C. M., Chang, C. L., Hopping, S., & Edwards, T. L. (2022). Transitioning from training to testing with scent detection animals: Application to lung cancer detection dogs. *Journal of Veterinary Behavior*. https://doi.org/https://doi.org/10.1016/j.jveb.2022.07.004

Davies, J. C., Alton, E., Simbo, A., Murphy, R., Seth, I., Williams, K., Somerville, M., Jolly, L., Morant, S., & Guest, C. (2019). Training dogs to differentiate *Pseudomonas aeruginosa* from other cystic fibrosis bacterial pathogens: Not to be sniffed at? *European Respiratory Journal*, *54*(5), 1900970. https://doi.org/10.1183/13993003.00970-2019

Dehlinger, K., Tarnowski, K., House, J. L., Los, E., Hanavan, K., Bustamante, B., Ahmann, A. J., & Ward, W. K. (2013). Can trained dogs detect a hypoglycemic scent in patients with type 1 diabetes? *Diabetes Care*, *36*(7), e98-e99. https://doi.org/10.2337/dc12-2342

Dorman, D. C., Foster, M. L., Fernhoff, K. E., & Hess, P. R. (2017). Canine scent detection of canine cancer: A feasibility study. *Veterinary Medicine , 8*, 69-76. https://doi.org/10.2147/VMRR.S148594

Edwards. (2019). Automated canine scent-detection apparatus: Technical description and training outcomes. *Chemical Senses*, *44*(7), 449-455. https://doi.org/10.1093/chemse/bjz039

Edwards, T. L., Browne, C. M., Schoon, A., Cox, C., & Poling, A. (2017). Animal olfactory detection of human diseases: Guidelines and systematic review. *Journal of Veterinary Behavior: Clinical Applications and Research*, 59-73. https://doi.org/https://doi.org/10.1016/j.jveb.2017.05.002

Edwards, T. L., Giezen, C., & Browne, C. M. (2022). Influences of indication response requirement and target prevalence on dogs' performance in a scent-detection task. *Applied Animal Behaviour Science*, *253*, 105657. https://doi.org/https://doi.org/10.1016/j.applanim.2022.105657

Ehmann, R., Boedeker, E., Friedrich, U., Sagert, J., Dippon, J., Friedel, G., & Walles, T. (2012). Canine scent detection in the diagnosis of lung cancer: Revisiting a puzzling phenomenon. *European Respiratory Journal*, *39*(3), 669. https://doi.org/10.1183/09031936.00051711

Elliker, K., Sommerville, B., Broom, D., Neal, D., Armstrong, S., & Williams, H. (2014). Key considerations for the experimental training and evaluation of cancer odour detection dogs: Lessons learnt from a double-blind, controlled trial of prostate cancer detection. *BMC Urology*, *14*(1), 22. https://doi.org/10.1186/1471-2490-14-22

Ellis, H., Mulder, C., Valverde, E., Poling, A., & Edwards, T. (2017). Reproducibility of African giant pouched rats detecting *Mycobacterium tuberculosis*. *BMC Infectious Diseases*, *17*(1). https://doi.org/10.1186/s12879-017-2347-3

Ellis, H. M., Watkins, E. E., Cox, C., Sargisson, R. J., & Edwards, T. L. (2019). Trialing a semi-automated line cage for scent detection by African pouched rats. *Behavior Analysis: Research and Practice*, *19*(2), 150-163. https://doi.org/10.1037/bar0000095

Eskandari, E., Ahmadi Marzaleh, M., Roudgari, H., Hamidi Farahani, R., Nezami-Asl, A., Laripour, R., Aliyazdi, H., Dabbagh Moghaddam, A., Zibaseresht, R., Akbarialiabad,

H., Yousefi Zoshk, M., Shiri, H., & Shiri, M. (2021). Sniffer dogs as a screening/diagnostic tool for COVID-19: A proof-of-concept study. *BMC Infectious Diseases*, *21*(1), 243-248. https://doi.org/10.1186/s12879-021-05939-6

Fischer-Tenhagen, C., Johnen, D., Nehls, I., & Becker, R. (2018). A proof-of-concept: Are detection dogs a useful tool to verify potential biomarkers for lung cancer? *Frontiers in Veterinary Science*, *5*, 52. https://doi.org/10.3389/fvets.2018.00052

Fischer-Tenhagen, C., Theby, V., Krömker, V., & Heuwieser, W. (2018). Detecting *Staphylococcus aureus* in milk from dairy cows using sniffer dogs. *Journal of Dairy Science*, *101*(5), 4317-4324. https://doi.org/10.3168/jds.2017-14100

Fisher, R. S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J. H., Elger, C. E., Engel, J., Forsgren, L., French, J. A., Glynn, M., Hesdorffer, D. C., Lee, B. I., Mathern, G. W., Moshé, S. L., Perucca, E., Scheffer, I. E., Tomson, T., Watanabe, M., & Wiebe, S. (2014). ILAE official report: A practical clinical definition of epilepsy. *Epilepsia*, *55*(4), 475-482. https://doi.org/10.1111/epi.12550

Gadbois, S., & Reeve, C. (2016). The semiotic canine: Scent processing dogs as research assistants in biomedical and environmental research. *Dog Behavior*, *2*(3), 26–32. https://doi.org/10.4454/db.v2i3.43

Gallistel, Fairhurst, S., & Balsam, P. (2004). The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(36), 13124-13131. https://doi.org/10.1073/pnas.0404965101

Gonder-Frederick, L. A., Grabman, J. H., Shepard, J. A., Tripathi, A. V., Ducar, D. M., & McElgunn, Z. R. (2017). Variability of diabetes alert dog accuracy in a real-world setting. *Journal of Diabetes Science and Technology*, *11*(4), 714-719. https://doi.org/10.1177/1932296816685580

Gordon, R. T., Schatz, C. B., Myers, L. J., Kosty, M., Gonczy, C., Kroener, J., Tran, M., Kurtzhals, P., Heath, S., Koziol, J. A., Arthur, N., Gabriel, M., Hemping, J., Hemping, G., Nesbitt, S., Tucker-Clark, L., & Zaayer, J. (2008). The use of canines in the detection of human cancers. *The Journal of Alternative and Complementary Medicine* , *14*(1), 61-67. https://doi.org/10.1089/acm.2006.6408

Goss, K.-U. (2019). The physical chemistry of odors — consequences for the work with detection dogs. *Forensic Science International*, *296*, 110-114. https://doi.org/10.1016/j.forsciint.2019.01.023

Grandjean, D., Sarkis, R., Lecoq-Julien, C., Benard, A., Roger, V., Levesque, E., Bernes-Luciani, E., Maestracci, B., Morvan, P., Gully, E., Berceau-Falancourt, D., Haufstater, P., Herin, G., Cabrera, J., Muzzin, Q., Gallet, C., Bacqué, H., Broc, J.-M., Thomas, L., Lichaa, A., Moujaes, G., Saliba, M., Kuhn, A., Galey, M., Berthail, B., Lapeyre, L., Capelli, A., Renault, S., Bachir, K., Kovinger, A., Comas, E., Stainmesse, A., Etienne, E., Voeltzel, S., Mansouri, S., Berceau-Falancourt, M., Dami, A., Charlet, L., Ruau, E., Issa, M., Grenet, C., Billy, C., Tourtier, J.-P., & Desquilbet, L. (2020). Can the detection dog alert on COVID-19 positive persons by sniffing axillary sweat samples? A proof-of-concept study. *PLoS ONE*, *15*(12), e0243122-e0243122. https://doi.org/10.1371/journal.pone.0243122

Guerrero-Flores, H., Apresa-García, T., Garay-Villar, Ó., Sánchez-Pérez, A., Flores-Villegas, D., Bandera-Calderón, A., García-Palacios, R., Rojas-Sánchez, T., Romero-Morelos, P., Sánchez-Albor, V., Mata, O., Arana-Conejo, V., Badillo-Romero, J., Taniguchi, K., Marrero-Rodríguez, D., Mendoza-Rodríguez, M., Rodríguez-Esquivel, M., Huerta-Padilla, V., Martínez-Castillo, A., Hernández-Gallardo, I., López-Romero, R., Bandala, C., Rosales-Guevara, J., & Salcedo, M. (2017). A non-invasive tool for detecting

cervical cancer odor by trained scent dogs. *BMC Cancer*, *17*(1), Article 79. https://doi.org/10.1186/s12885-016-2996-4

Guest, C., Harris, R., Sfanos, K. S., Shrestha, E., Partin, A. W., Trock, B., Mangold, L., Bader, R., Kozak, A., McLean, S., Simons, J., Soule, H., Johnson, T., Lee, W.-Y., Gao, Q., Aziz, S., Stathatou, P. M., Thaler, S., Foster, S., & Mershin, A. (2021). Feasibility of integrating canine olfaction with chemical and microbial profiling of urine to detect lethal prostate cancer. *PLoS ONE*, *16*(2), e0245530-e0245530. https://doi.org/10.1371/journal.pone.0245530

Guest, C., Pinder, M., Doggett, M., Squires, C., Affara, M., Kandeh, B., Dewhirst, S., Morant, S. V., D'Alessandro, U., Logan, J. G., & Lindsay, S. W. (2019). Trained dogs identify people with malaria parasites by their odour. *The Lancet Infectious Diseases*, *19*(6), 578-580. https://doi.org/10.1016/S1473-3099(19)30220-8

Guest, C. M., Harris, R., Anjum, I., Concha, A. R., & Rooney, N. J. (2020). A lesson in standardization – subtle aspects of the processing of samples can greatly affect dogs' learning. *Frontiers in Veterinary Science*, *7*, Article 525. https://doi.org/10.3389/fvets.2020.00525

Guirao, A., Molins, L., Ramón, I., Sunyer, G., Viñolas, N., Marrades, R., Sánchez, D., Fibla, J. J., Boada, M., Hernández, J., Guzmán, R., Libreros, A., Gómez-Caro, A., Guerrero, C., & Agustí, A. (2019). Trained dogs can identify malignant solitary pulmonary nodules in exhaled gas. *Lung Cancer*, *135*, 230-233. https://doi.org/10.1016/j.lungcan.2019.06.008

Guirao Montes, Á., Molins López-Rodó, L., Ramón Rodríguez, I., Sunyer Dequigiovanni, G., Viñolas Segarra, N., Marrades Sicart, R. M., Hernández Ferrández, J., Fibla Alfara, J. J., & Agustí García-Navarro, Á. (2017). Lung cancer diagnosis by trained dogs.

European Journal of Cardio-Thoracic Surgery, *52*(6), 1206-1210. https://doi.org/10.1093/ejcts/ezx152

Hackner, K., Errhalt, P., Mueller, M. R., Speiser, M., Marzluf, B. A., Schulheim, A., Schenk, P., Bilek, J., & Doll, T. (2016). Canine scent detection for the diagnosis of lung cancer in a screening-like situation. *Journal of Breath Research*, *10*(4), 046003. https://doi.org/10.1088/1752-7155/10/4/046003

Hall, G. (2003). Learned changes in the sensitivity of stimulus representations: Associative and nonassociative mechanisms. *The Quarterly Journal of Experimental Psychology Section B*, *56*(1b), 43-55. https://doi.org/10.1080/02724990244000151

Hall, N. J. (2017). Persistence and resistance to extinction in the domestic dog: Basic research and applications to canine training. *Behavioural Processes*, *141*(1), 67-74. https://doi.org/10.1016/j.beproc.2017.04.001

Hardin, D. S., Anderson, W., & Cattet, J. (2015). Dogs can be successfully trained to alert to hypoglycemia samples from patients with Type 1 diabetes. *Diabetes Therapy*, *6*(4), 509-517. https://doi.org/10.1007/s13300-015-0135-x

Harshman, S. W., Mani, N., Geier, B. A., Kwak, J., Shepard, P., Fan, M., Sudberry, G. L., Mayes, R. S., Ott, D. K., Martin, J. A., & Grigsby, C. C. (2016). Storage stability of exhaled breath on Tenax TA. *Journal of Breath Research*, *10*(4), 046008. https://doi.org/10.1088/1752-7155/10/4/046008

Hautus, M. (2015). Signal detection theory. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed.). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.43090-4

Hayes, J. E., McGreevy, P. D., Forbes, S. L., Laing, G., & Stuetz, R. M. (2018). Critical review of dog detection and the influences of physiology, training, and analytical methodologies. *Talanta*, *185*, 499-512. https://doi.org/https://doi.org/10.1016/j.talanta.2018.04.010

Herrnstein, R. J. (1990). Levels of stimulus control: A functional approach. *Cognition*, *37*(1), 133-166. https://doi.org/10.1016/0010-0277(90)90021-B

Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon. *Science*, *146*(3643), 549-550. https://doi.org/10.1126/science.146.3643.549

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*(4), 285-302. https://doi.org/10.1037/0097-7403.2.4.285

Hirotsu, T., Sonoda, H., Uozumi, T., Shinden, Y., Mimori, K., Maehara, Y., Ueda, N., & Hamakawa, M. (2015). A highly accurate inclusive cancer screening test using *Caenorhabditis elegans* scent detection. *PLoS ONE*, *10*(3), e0118699-e0118699. https://doi.org/10.1371/journal.pone.0118699

Horvath, G., Andersson, H., & Nemes, S. (2013). Cancer odor in the blood of ovarian cancer patients: A retrospective study of detection by dogs during treatment, 3 and 6 months afterward. *BMC Cancer*, *13*(1), 396-396. https://doi.org/10.1186/1471-2407-13-396

Horvath, G., Andersson, H., & Paulsson, G. (2010). Characteristic odour in the blood reveals ovarian carcinoma. *BMC Cancer*, *10*(1), 643-643. https://doi.org/10.1186/1471-2407-10-643

Horvath, G., Järverud, G. a. K., Järverud, S., & Horváth, I. (2008). Human ovarian carcinomas detected by specific odor. *Integrative Cancer Therapies, 7*(2), 76-80. https://doi.org/10.1177/1534735408319058

Hudson, D. T., Curran, A. M., & Furton, K. G. (2009). The stability of collected human scent under various environmental conditions. *Journal of Forensic Science, 54*(6), 1270-1277. https://doi.org/10.1111/j.1556-4029.2009.01153.x

Iyengar, G. V., Subramanian, K. S., & Woittiez, J. R. W. (2020). *Element Analysis of Biological Samples: Principles and Practices* (Vol. II). CRC Press.

Jaine, Kvizhinadze, G., Wilson, N., & Blakely, T. (2018). *Screening for lung cancer in NZ is highly unlikely to be cost-effective*. Retrieved March 22, 2019, from https://blogs.otago.ac.nz/pubhealthexpert/2018/09/03/screening-for-lung-cancer-in-nz-is-highly-unlikely-to-be-cost-effective-new-nz-study/

Jendrny, P., Schulz, C., Twele, F., Meller, S., von Köckritz-Blickwede, M., Osterhaus, A. D. M. E., Ebbers, J., Pilchová, V., Pink, I., Welte, T., Manns, M. P., Fathi, A., Ernst, C., Addo, M. M., Schalke, E., & Volk, H. A. (2020). Scent dog identification of samples from COVID-19 patients - a pilot study. *BMC Infectious Diseases, 20*(1), 536-536. https://doi.org/10.1186/s12879-020-05281-3

Jendrny, P., Twele, F., Meller, S., Osterhaus, A., Schalke, E., & Volk, H. A. (2021). Canine olfactory detection and its relevance to medical detection. *BMC Infectious Diseases, 21*(1), 838. https://doi.org/10.1186/s12879-021-06523-8

Johnen, D., Heuwieser, W., Fischer-Tenhagen, C. (2017). An approach to identify bias in scent detection dog testing. *Applied Animal Behaviour Science, 189*, 1-12. https://doi.org/https://doi.org/10.1016/j.applanim.2017.01.001

Juge, A. E., Foster, M. F., & Daigle, C. L. (2022). Canine olfaction as a disease detection technology: A systematic review. *Applied Animal Behaviour Science*, *253*, 105664. https://doi.org/10.1016/j.applanim.2022.105664

Junqueira, H., Quinn, T. A., Biringer, R., Hussein, M., Smeriglio, C., Barrueto, L., Finizio, J., & Huang, X. Y. (2019). Accuracy of canine scent detection of non-small cell lung cancer in blood serum. *Journal of the American Osteopathic Association*, *119*(7), 413-418. https://doi.org/10.7556/jaoa.2019.077

Kantor, J. R. (1988). *The Logic of Modern Science.* (3rd ed.). Principia Press.

Katakura, A., Yamamoto, N., Sakuma, T., Sugahara, K., Onda, T., Noguchi, S., & Shibahara, T. (2015). A screening test for oral cancer using saliva samples: Proteomic analysis of biomarkers in whole saliva. *Journal of Oral and Maxillofacial Surgery, Medicine, and Pathology*, *27*(1), 1-5. https://doi.org/10.1016/j.ajoms.2013.05.013

Kauczor, H. U., Baird, A. M., Blum, T. G., Bonomo, L., Bostantzoglou, C., Burghuber, O., Cepicka, B., Comanescu, A., Couraud, S., Devaraj, A., Jespersen, V., Morozov, S., Agmon, I. N., Peled, N., Powell, P., Prosch, H., Ravara, S., Rawlinson, J., Revel, M. P., Silva, M., Snoeckx, A., Ginneken, B. v., Meerbeeck, J. P. v., Vardavas, C., Stackelberg, O. v., & Gaga, M. (2020). ESR/ERS statement paper on lung cancer screening. *European Radiology*, *30*(6), 3277-3294. https://doi.org/10.1007/s00330-020-06727-7

Keep, B., Pike, T. W., Moszuti, S. A., Zulch, H. E., Ratcliffe, V. F., Porritt, F., Hobbs, E., & Wilkinson, A. (2021). The impact of training method on odour learning and generalisation in detection animals. *Applied Animal Behaviour Science*, *236*, 105266. https://doi.org/10.1016/j.applanim.2021.105266

Keller, B. M., S., F., & Schoenfeld, W. N. (1951). Principles of psychology: A systematic text in the science of behavior. *American Journal of Psychology*, *64*, 473-474. https://doi.org/10.2307/1419029

Kitiyakara, T., Redmond, S., Unwanatham, N., Rattanasiri, S., Thakkinstian, A., Tangtawee, P., Mingphruedhi, S., Sobhonslidsuk, A., Intaraprasong, P., & Kositchaiwat, C. (2017). The detection of hepatocellular carcinoma (HCC) from patients' breath using canine scent detection: A proof-of-concept study. *Journal of Breath Research*, *11*(4), 046002-046002. https://doi.org/10.1088/1752-7163/aa7b8e

Klaus, H., & Joachim, P. (2017). Canine olfaction as an alternative to analytical instruments for disease diagnosis: Understanding 'dog personality' to achieve reproducible results. *Journal of Breath Research*, *11*(1), 012001. http://stacks.iop.org/1752-7163/11/i=1/a=012001

Koivusalo, M., Vermeiren, C., Yuen, J., Reeve, C., Gadbois, S., & Katz, K. (2017). Canine scent detection as a tool to distinguish meticillin-resistant *Staphylococcus aureus*. *The Journal of Hospital Infection*, *96*(1), 93-95. https://doi.org/10.1016/j.jhin.2017.03.005

Kokocińska-Kusiak, A., Woszczyło, M., Zybala, M., Maciocha, J., Barłowska, K., & Dzięcioł, M. (2021). Canine olfaction: Physiology, behavior, and possibilities for practical applications. *Animals*, *11*(8), 2463. https://doi.org/10.3390/ani11082463

Koskinen, A., Koskinen, H., Bäck, L., Saxen, H., & Klockars, T. (2017). A detection dog for paediatric urinary tract infection caused by *Escherichia coli*. *Infectious Diseases 49*(11-12), 874-877. https://doi.org/10.1080/23744235.2017.1331466

Krause, M. A., Udell, M. A. R., Leavens, D. A., & Skopos, L. (2018). Animal pointing: Changing trends and findings from 30 years of research. *Journal of Comparative Psychology*, *132*(3), 326-345. https://doi.org/10.1037/com0000125

Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation, 24*(3-4), 445-463. https://doi.org/10.1080/09602011.2013.815636

Leitch, O., Anderson, A., Kirkbride, K. P., & Lennard, C. (2013). Biological organisms as volatile compound detectors: A review. *Forensic Science International, 232*(1-3), 92-103. https://doi.org/10.1016/j.forsciint.2013.07.004

Lit, L., Schweitzer, J. B., & Oberbauer, A. M. (2011). Handler beliefs affect scent detection dog outcomes. *Animal Cognition, 14*(3), 387-394. https://doi.org/10.1007/s10071-010-0373-2

Loeb, G. E. (2012). Optimal isn't good enough. *Biological Cybernetics, 106*(11-12), 757-765. https://doi.org/10.1007/s00422-012-0514-6

Los, E. A., Ramsey, K. L., Guttmann-Bauman, I., & Ahmann, A. J. (2017). Reliability of trained dogs to alert to hypoglycemia in patients with Type 1 diabetes. *Journal of Diabetes Science and Technology, 11*(3), 506-512. https://doi.org/10.1177/1932296816666537

Lourenço, C., & Turner, C. (2014). Breath analysis in disease diagnosis: Methodological considerations and applications. *Metabolites, 4*(2), 465-498. https://doi.org/10.3390/metabo4020465

McCulloch M, Jezierski T, Broffman M, Hubbard A, Turner K, Janecki T. (2006). Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. *Integrative Cancer Therapies, 5*(1):30-9. https://doi.org/10.1177/1534735405285096

Ma, L., Zhang, Z., & Li, X. (2020). Non-invasive disease diagnosis using surface-enhanced Raman spectroscopy of urine and saliva. *Applied Spectroscopy Reviews, 55*(3), 197-219. https://doi.org/10.1080/05704928.2018.1562938

Maa, E., Arnold, J., Ninedorf, K., & Olsen, H. (2021). Canine detection of volatile organic compounds unique to human epileptic seizure. *Epilepsy & Behavior, 115*, 107690. https://doi.org/10.1016/j.yebeh.2020.107690

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). Taylor & Francis Group. https://doi.org/10.4324/9781410611147

Mahoney, A. M., Weetjens, B. J., Cox, C., Beyene, N., Mgode, G., Jubitana, M., Kuipers, D., Kazwala, R., Mfinanga, G. S., Durgin, A., & Poling, A. (2011). Using giant African pouched rats to detect tuberculosis in human sputum samples: 2010 findings. *The Pan African Medical Journal*, *9*(1), 28-28. https://doi.org/10.4269/ajtmh.10-0180

Mahoney, A., Edwards, T. L., Weetjens, B. J., Cox, C., Beyene, N., Jubitana, M., Durgin, A., & Poling, A. (2013). Giant African pouched rats (*Cricetomys gambianus*) as detectors of tuberculosis in human sputum: Two operational improvements. *The Psychological Record, 63*(3), 583-593. https://doi.org/10.11133/j.tpr.2013.63.3.012

Maughan, M. N., Best, E. M., Gadberry, J. D., Sharpes, C. E., Evans, K. L., Chue, C. C., Nolan, P. L., & Buckley, P. E. (2022). The use and potential of biomedical detection dogs during a disease outbreak. *Frontiers in Medicine, 9*, 848090-848090. https://doi.org/10.3389/fmed.2022.848090

Maurer, M., McCulloch, M., Willey, A. M., Hirsch, W., & Dewey, D. (2016). Detection of bacteriuria by canine olfaction. *Open Forum Infectious Diseases, 3*(2), ofw051-ofw051. https://doi.org/10.1093/ofid/ofw051

Mazzola, S. M., Pirrone, F., Sedda, G., Gasparri, R., Romano, R., Spaggiari, L., & Mariangela, A. (2020). Two-step investigation of lung cancer detection by sniffer dogs. *Journal of Breath Research*, *14*(2), 026011. https://doi.org/10.1088/1752-7163/ab716e

Mendel, J., Frank, K., Edlin, L., Hall, K., Webb, D., Mills, J., Holness, H. K., Furton, K. G., & Mills, D. (2021). Preliminary accuracy of COVID-19 odor detection by canines and HS-SPME-GC-MS using exhaled breath samples. *Forensic Science International: Synergy*, *3*, 100155-100155. https://doi.org/10.1016/j.fsisyn.2021.100155

Mgode, G. F., Cox, C. L., Mwimanzi, S., & Mulder, C. (2018). Pediatric tuberculosis detection using trained African giant pouched rats. *Pediatric Research*, *84*(1), 99-103. https://doi.org/10.1038/pr.2018.40

Mgode, G. F., Weetjens, B. J., Nawrath, T., Cox, C., Jubitana, M., Machang'u, R. S., Cohen-Bacrie, S., Bedotto, M., Drancourt, M., Schulz, S., & Kaufmann, S. H. E. (2012). Diagnosis of tuberculosis by trained African giant pouched rats and confounding impact of pathogens and microflora of the respiratory tract. *Journal of Clinical Microbiology*, *50*(2), 274-280. https://doi.org/10.1128/JCM.01199-11

Murarka, M., Vesley-Gross, Z. I., Essler, J. L., Smith, P. G., Hooda, J., Drapkin, R., & Otto, C. M. (2019). Testing ovarian cancer cell lines to train dogs to detect ovarian cancer from blood plasma: A pilot study. *Journal of Veterinary Behavior*, *32*, 42-48. https://doi.org/10.1016/j.jveb.2019.04.010

Murre, J. M. J. (2014). S-shaped learning curves. *Psychonomic Bulletin and Review*, *21*(2), 344-356. https://doi.org/10.3758/s13423-013-0522-0

National Institute of Health Research. (2018). *Largest ever UK lung cancer screening study launches in London*. Retrieved May 9, 2019, from https://www.nihr.ac.uk/news/largest-ever-uk-lung-cancer-screening-study-launches-in-london/9808

Neto, M. F., Nguyen, Q. H., Marsili, J., McFall, S. M., & Voisine, C. (2016). The nematode *Caenorhabditis elegans* displays a chemotaxis behavior to tuberculosis-specific odorants. *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases*, *4*, 44-49. https://doi.org/10.1016/j.jctube.2016.06.001

Ostrin, E. J., Sidransky, D., Spira, A., & Hanash, S. M. (2020). Biomarkers for lung cancer screening and detection. *Cancer Epidemiology Biomarkers and Prevention*, *29*(12), 2411. https://doi.org/10.1158/1055-9965.EPI-20-0865

Pacik, D., Plevova, M., Urbanova, L., Lackova, Z., Strmiska, V., Necas, A., Heger, Z., & Adam, V. (2018). Identification of sarcosine as a target molecule for the canine olfactory detection of prostate carcinoma. *Scientific Reports*, *8*(1), 4958-4955. https://doi.org/10.1038/s41598-018-23072-4

Pamir, E., Chakroborty, N., Stollhoff, N., Gehring, K., Antemann, V., Morgenstern, L., Felsenberg, J., Eisenhardt, D., Menzel, R., & Nawrot, M. (2011). Average group behavior does not represent individual behavior in classical conditioning of the honeybee. *Learning and Memory*, *18*, 733-741. https://doi.org/10.1101/lm.2232711

Pennazza, G., & Santonico, M. (2018). *Breath Analysis*. Elsevier Science & Technology.

Pickel, D., Manucy, G. P., Walker, D. B., Hall, S. B., & Walker, J. C. (2004). Evidence for canine olfactory detection of melanoma. *Applied Animal Behaviour Science*, *89*(1), 107-116. https://doi.org/10.1016/j.applanim.2004.04.008

Pirrone, F., & Albertini, M. (2017). Olfactory detection of cancer by trained sniffer dogs: A systematic review of the literature. *Journal of Veterinary Behavior*, *19*, 105-117. https://doi.org/10.1016/j.jveb.2017.03.004

Poling, A., Weetjens, B., Cox, C., Beyene, N., Durgin, A., & Mahoney, A. (2011). Tuberculosis detection by giant African pouched rats. *The Behavior Analyst, Spring 34*(1), 47–54. https://doi.org/10.1016/j.jveb.2017.03.004

Poling, A., Weetjens, B. J., Cox, C., Mgode, G., Jubitana, M., Kazwala, R., Mfinanga, G. S., & Huis In 'T Veld, D. (2010). Short report: Using giant African pouched rats to detect tuberculosis in human sputum samples: 2009 findings. *The American Journal of Tropical Medicine and Hygiene, 83*(6), 1308-1310. https://10.4314/pamj.v9i1.71204

Printz, C. (2012). Saliva yields clues to early cancer detection. *Cancer*, *118*(13), 3224-3224. https://doi.org/10.1002/cncr.27700

Qian, K., Wang, Y., Hua, L., Chen, A., & Zhang, Y. (2018). New method of lung cancer detection by saliva test using surface-enhanced Raman spectroscopy. *Thoracic Cancer*, *9*(11), 1556-1561. https://doi.org/10.1111/1759-7714.12837

Reeve, C. P., Cummings, E. M. D., McLaughlin, E. P., Smith, S. B., & Gadbois, S. P. (2020). An idiographic investigation of diabetic alert dogs' ability to learn from a small sample of breath samples from people with Type 1 diabetes. *Canadian Journal of Diabetes, 44*(1), 37-43.e31. https://doi.org/10.1016/j.jcjd.2019.04.020

Reid, P. J. (2009). Adapting to the human world: Dogs' responsiveness to our social cues. *Behavioural Processes*, *80*(3), 325-333. https://doi.org/10.1016/j.beproc.2008.11.002

Reither, K., Jugheli, L., Glass, T. R., Sasamalo, M., Mhimbira, F. A., Weetjens, B. J., Cox, C., Edwards, T. L., Mulder, C., Beyene, N. W., & Mahoney, A. (2015). Evaluation of giant African pouched rats for detection of pulmonary tuberculosis in patients from a high-endemic setting. *PLoS ONE, 10*(10), e0135877-e0135877. https://doi.org/10.1371/journal.pone.0135877

Richards, K. M., Cotton, S. J., & Sandeman, Richard M. (2008). The use of detector dogs in the diagnosis of nematode infections in sheep feces. *Journal of Veterinary Behavior*, *3*(1), 25-31. https://doi.org/10.1016/j.jveb.2007.10.006

Rodionova, E. I., Kochevalina, M. Y., Kotenkova, E. V., Morozova, O. V., Kogun', G. A., Bataeva, E. L., & Ambaryan, A. V. (2015). Detection of volatile organic compounds associated with hepatocellular carcinoma by macrosmatic animals: Approaches to the search for new tumor markers. *Biology Bulletin of the Russian Academy of Sciences*, *42*(3), 239-245. https://doi.org/10.1134/S1062359015030103

Rooney, N. J., Guest, C. M., Swanson, L. C. M., & Morant, S. V. (2019). How effective are trained dogs at alerting their owners to changes in blood glycaemic levels? Variations in performance of glycaemia alert dogs. *PLoS ONE*, *14*(1), e0210092-e0210092. https://doi.org/10.1371/journal.pone.0210092

Rudnicka, J., Walczak, M., Kowalkowski, T., Jezierski, T., & Buszewski, B. (2014). Determination of volatile organic compounds as potential markers of lung cancer by gas chromatography–mass spectrometry versus trained dogs. *Sensors & Actuators B: Chemical*, *202*, 615-621. https://doi.org/10.1016/j.snb.2014.06.006

Saalberg, Y., & Wolff, M. (2016). VOC breath biomarkers in lung cancer. *Clinica Chimica Acta*, *459*, 5-9. https://doi.org/10.1016/j.cca.2016.05.013

Samhita, L., & Gross, H. J. (2013). The "Clever Hans Phenomenon" revisited. *Communicative & Integrative Biology*, *6*(6), 27122. https://doi.org/10.4161/cib.27122

Schallschmidt, K., Becker, R., Jung, C., Bremser, W., Walles, T., Neudecker, J., Leschber, G., Frese, S., & Nehls, I. (2016). Comparison of volatile organic compounds from lung cancer patients and healthy controls—challenges and limitations of an observational

study. *Journal of Breath Research*, *10*(4), 046007. https://doi.org/10.1088/1752-7155/10/4/046007

Schallschmidt, K., Becker, R., Zwaka, H., Menzel, R., Johnen, D., Fischer-Tenhagen, C., Rolff, J., & Nehls, I. (2015). In-vitro cultured lung cancer cells are not suitable for animal-based breath biomarker detection. *Journal of Breath Research*, *9*(2), 027103-027103. https://doi.org/10.1088/1752-7155/9/2/027103

Schoon, G. A. A. (2005). The effect of the ageing of crime scene objects on the results of scent identification line-ups using trained dogs. *Forensic Science International*, *147*(1), 43-47. https://doi.org/10.1016/j.forsciint.2004.04.080

Schoon, G. A. A., De Jonge, D., & Hilverink, P. (2020). How dogs learn to detect colon cancer—Optimizing the use of training aids. *Journal of Veterinary Behavior*, *35*, 38-44. https://doi.org/10.1016/j.jveb.2019.10.006

Seo, I.-S., Lee, H.-G., Koo, B., Koh, C. S., Park, H.-Y., Im, C., & Shin, H.-C. (2018). Cross-detection for odor of metabolic waste between breast and colorectal cancer using canine olfaction. *PLoS ONE*, *13*(2), e0192629-e0192629. https://doi.org/10.1371/journal.pone.0192629

Shirasu, M., & Touhara, K. (2011). The scent of disease: Volatile organic compounds of the human body related to disease and disorder. *The Journal of Biochemistry*, *150*(3), 257-266. https://doi.org/10.1093/jb/mvr090

Singh, H., & Pazdur, R. (2021). Contribution of early clinical benefit end points to decreased lung cancer mortality rates. *JAMA Oncology*, *7*(6), 829-830. https://doi.org/10.1001/jamaoncol.2020.8090

Sonoda, H., Kohnoe, S., Yamazato, T., Satoh, Y., Morizono, G., Shikata, K., Morita, M., Watanabe, A., Morita, M., Kakeji, Y., Inoue, F., & Maehara, Y. (2011). Colorectal cancer screening with odour material by canine scent detection. *Gut*, *60*(6), 814-819. https://doi.org/10.1136/gut.2010.218305

Staddon, J. E. R., & Cerutti, D. T. (2003). Operant conditioning. *Annual Review of Psychology*, *54*, 115-144. https://doi.org/10.1146/annurev.psych.54.101601.145124

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137-149. https://doi.org/10.3758/BF03207704

Suckling, D. M., & Sagar, R. L. (2011). Honeybees *Apis mellifera* can detect the scent of *Mycobacterium tuberculosis*. *Tuberculosis, 91*(4), 327-328. https://doi.org/10.1016/j.tube.2011.04.008

Taverna, G., Tidu, L., Grizzi, F., Stork, B., Seveso, M., Bozzini, G., Sardella, P., Latorre, G., Lughezzani, G., Buffi, N., Guazzoni, G., & Mandressi, A. (2016). Highly-trained dogs' olfactory system for detecting biochemical recurrence following radical prostatectomy. *European Urology Supplements 15*(3), 734. https://doi.org/10.1016/S1569-9056(16)60736-4

Taylor, M. T., McCready, J., Broukhanski, G., Kirpalaney, S., Lutz, H., & Powis, J. (2018). Using dog scent detection as a point-of-care tool to identify toxigenic *Clostridium difficile* in stool. *Open Forum Infectious Diseases*, *5*(8), 179. https://doi.org/10.1093/ofid/ofy179

Thuleau, A., Gilbert, C., Bauër, P., Alran, S., Fourchotte, V., Guillot, E., Vincent-Salomon, A., Kerihuel, J.-C., Dugay, J., Semetey, V., Kriegel, I., & Fromantin, I. (2019). A new

transcutaneous method for breast cancer detection with dogs. *Oncology*, *96*(2), 110-113. https://doi.org/10.1159/000492895

Torri, L., & Piochi, M. (2016). Sensory methods and electronic nose as innovative tools for the evaluation of the aroma transfer properties of food plastic bags. *Food Research International*, *85*, 235-243. https://doi.org/10.1016/j.foodres.2016.05.004

Troisi, C. A., Mills, D. S., Wilkinson, A., & Zulch, H. E. (2019). Behavioral and cognitive factors that affect the success of scent detection dogs. *Comparative Cognition & Behavior Reviews*, *14*, 51-76. https://doi.org/10.3819/CCBR.2019.140007

Urbanová, L., Vyhnánková, V., Krisová, Š., Pacík, D., & Nečas, A. (2015). Intensive training technique utilizing the dog's olfactory abilities to diagnose prostate cancer in men. *Acta Veterinaria Brno*, *84*(1), 77-82. https://doi.org/10.2754/avb201585010077

van der Schee, M. P., Fens, N., Brinkman, P., Bos, L. D. J., Angelo, M. D., Nijsen, T. M. E., Raabe, R., Knobel, H. H., Vink, T. J., & Sterk, P. J. (2012). Effect of transportation and storage using sorbent tubes of exhaled breath samples on diagnostic accuracy of electronic nose analysis. *Journal of Breath Research*, *7*(1), 016002. https://doi.org/10.1088/1752-7155/7/1/016002

Walczak, M., Jezierski, T., Górecka-Bruzda, A., Sobczyńska, M., & Ensminger, J. (2012). Impact of individual training parameters and manner of taking breath odor samples on the reliability of canines as cancer screeners. *Journal of Veterinary Behavior: Clinical Applications and Research*, *7*(5), 283-294. https://doi.org/10.1016/j.jveb.2012.01.001

Wang, X. R., Cassells, J., & Berna, A. Z. (2018). Stability control for breath analysis using GC-MS. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, *1097-1098*, 27-34. https://doi.org/10.1016/j.jchromb.2018.08.024

Wasserman, E. A. (2016). Conceptualization in pigeons: The evolution of a paradigm. *Behavioural Processes*, *123*, 4-14. https://doi.org/10.1016/j.beproc.2015.09.010

Williams, J., & Koppmann, R. (2007). Volatile organic compounds in the atmosphere: An overview. In *Volatile Organic Compounds in the Atmosphere* (pp. 1-32). Wiley Online Books. https://doi.org/https://doi.org/10.1002/9780470988657.ch1

Willis, C. M., Britton, L. E., Harris, R., Wallace, J., & Guest, C. M. (2011). Volatile organic compounds as biomarkers of bladder cancer: Sensitivity and specificity using trained sniffer dogs. *Cancer Biomarkers: Section A of Disease Markers*, *8*(3), 145-153. https://doi.org/10.3233/CBM-2011-0208

Willis, C. M., Britton, L. E., Swindells, M. A., Jones, E. M., Kemp, A. E., Muirhead, N. L., Gul, A., Matin, R. N., Knutsson, L., & Ali, M. (2016). Invasive melanoma in vivo can be distinguished from basal cell carcinoma, benign naevi and healthy skin by canine olfaction: A proof-of-principle study of differential volatile organic compound emission. *British Journal of Dermatology*, *175*(5), 1020-1029. https://doi.org/10.1111/bjd.14887

Willis, C. M., Church, S. M., Guest, C. M., Cook, W. A., McCarthy, N., Bransbury, A. J., Church, M. R. T., & Church, J. C. T. (2004). Olfactory detection of human bladder cancer by dogs: Proof-of-principle study. *BMJ*, *329*(7468), 712-714. https://doi.org/10.1136/bmj.329.7468.712

Wong, H. B., & Lim, G. H. (2011). Measures of diagnostic accuracy: Sensitivity, specificity, PPV and NPV. *Proceedings of Singapore Healthcare*, *20*(4), 316-318. https://doi.org/10.1177/201010581102000411

World Health Organisation. Cancer 2021. Available from: https://www.who.int/news-room/fact-sheets/detail/cancer.

Yamamoto, A., Kamoi, S., Kurose, K., Ito, M., Takeshita, T., Kure, S., Sakamoto, K., Sato, Y., & Miyashita, M. (2020). The trained sniffer dog could accurately detect the urine samples from the patients with cervical cancer, and even cervical intrapithelial neoplasia grade 3: A pilot study. *Cancers*, *12*(11), 3291. https://doi.org/10.3390/cancers12113291

Zentall, T. R., Galizio, M., & Critchfield, T. S. (2002). Categorisation, concept learning, and behavior analysis: An introduction. *Journal of the Experimental Analysis of Behavior*, *78*(3), 237-248. https://doi.org/10.1901/jeab.2002.78-237

Zhan, X., Wang, Z., Yang, M., Luo, Z., Wang, Y., & Li, G. (2020). An electronic nose-based assistive diagnostic prototype for lung cancer detection with conformal prediction. *Measurement*, *158*, 107588. https://doi.org/https://doi.org/10.1016/j.measurement.2020.107588

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika*, *70*(1), 203-212. https://doi.org/10.1007/s11336-003-1119-8

Zhang, L., Xiao, H., Zhou, H., Santiago, S., Lee, J., Garon, E., Yang, J., Brinkmann, O., Yan, X., Akin, D., Chia, D., Elashoff, D., Park, N.-H., & Wong, D. (2012). Development of transcriptomic biomarker signature in human saliva to detect lung cancer. *Cellular and Molecular Life Sciences*, *69*(19), 3341-3350. https://doi.org/10.1007/s00018-012-1027-0

Appendix A: Publication Co-Authorship Forms

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work. Chapter 2: Unpublished Manuscript	
Nature of contribution by PhD candidate	Crawford led the investigation, visualization and drafting of the manuscript. She contributed in equal parts with co-authors to conceptualization, formal analysis, methodology, project administration, and the editing and review of the manuscript. Crawford played a supporting role in the administration of resources.
Extent of contribution by PhD candidate (%)	75

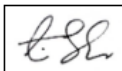
CO-AUTHORS

Name	Nature of Contribution
Dr Timothy L. Edwards	Led data curation, funding acquisition, resource administration and supervision. Edwards contributed equally with co-authors in conceptualization, formal analysis, methodology, project administration, and the editing and review of the manuscript. He supported visualization and the development of the initial manuscript.
Dr Michael Jameson	Contributed equally with co-authors to develop conceptualization, and review and editing the manuscript. Jameson played a supporting role in formal analysis, funding acquisition, investigation, and supervision.
Dr Clare M. Browne	Contributed equally with co-authors to develop conceptualization, methodology, and review and editing the manuscript. Browne played a supporting role in formal analysis, funding acquisition, investigation, supervision and the writing of the draft manuscript.
Dr Catherina Chang	Contributed equally with co-authors to develop conceptualization, project administration, and the review and editing of the manuscript. Chang played a supporting role in formal analysis, funding acquisition and investigation.
Sandra Hopping	Contributed equally with co-authors to review and editing of the manuscript. Hopping played a supporting role in the administration of resources.

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

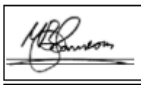
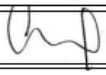
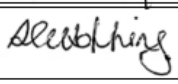
Name	Signature	Date
Dr Timothy L. Edwards		14/04/2023

July 2015

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

Dr Michael Jameson		18/04/2023
Dr Clare M. Browne		21/04/2023
Dr Catherina Chang		19/04/2023
Sandra Hopping		12/04/2023

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 3, Published Article: Crawford, M. A., Chang, C. L., Hopping, S., Browne, C. M., & Edwards, T. L. (2022). Influences of breath sample re-use on the accuracy of lung cancer detection dogs. *Journal of Breath Research*, *17*(1), 16001–. https://doi.org/10.1088/1752-7163/ac9b7f

Nature of contribution by PhD candidate	Crawford led the investigation, visualization and drafting of the manuscript. She contributed in equal parts with co-authors to conceptualization, formal analysis, methodology, project administration, and the editing and review of the manuscript. Crawford played a supporting role in the administration of resources.
Extent of contribution by PhD candidate (%)	75

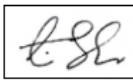
CO-AUTHORS

Name	Nature of Contribution
Dr Timothy L. Edwards	Led data curation, funding acquisition, resource administration and supervision. Edwards contributed equally with co-authors in conceptualization, formal analysis, methodology, project administration, and the editing and review of the manuscript. He supported visualization and the development of the initial manuscript.
Dr Clare M. Browne	Contributed equally with co-authors to develop conceptualization, methodology, and review and editing the manuscript. Browne played a supporting role in formal analysis, funding acquisition, investigation, supervision and the writing of the draft manuscript.
Dr Catherina Chang	Contributed equally with co-authors to develop conceptualization, project administration, and the review and editing of the manuscript. Chang played a supporting role in formal analysis, funding acquisition and investigation.
Sandra Hopping	Contributed equally with co-authors to review and editing of the manuscript. Hopping played a supporting role in the administration of resources.

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

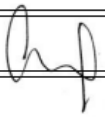
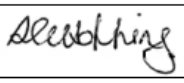
Name	Signature	Date
Dr Timothy L. Edwards		14/04/2023

July 2015

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

Dr Clare M. Browne		21/03/2023
Catherina Chang		19/4/2023
Sandra Hopping		12/04/2023

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.	Chapter 4, Published Article: Crawford, M. A., Perrone, J. A., Browne, C. M., Chang, C. L., Hopping, S., & Edwards, T. L. (2022). Transitioning from training to testing with scent detection animals: Application to lung cancer detection dogs. *Journal of Veterinary Behavior*, *55-56*, 23–34. https://doi.org/10.1016/j.jveb.2022.07.004
Nature of contribution by PhD candidate	Crawford led the investigation and drafting of the manuscript. She contributed in equal parts with co-authors to conceptualization, formal analysis, methodology, project administration, visualization, editing and review of the manuscript. Crawford played a supporting role in the administration of resources.
Extent of contribution by PhD candidate (%)	65

CO-AUTHORS

Name	Nature of Contribution
Dr Timothy L. Edwards	Led data curation, funding acquisition, resource administration and supervision. Edwards contributed equally with co-authors in conceptualization, formal analysis, methodology, project administration, visualization, editing and review of the manuscript. He supported development of the initial manuscript.
Prof. John A. Perrone	Contributed equally with co-authors to develop formal analysis, methodology, visualization, editing and review of the final document. Perrone played a supporting role in conceptualization, data curation and supervision.
Dr Clare M. Browne	Contributed equally with co-authors to develop conceptualization, and review and editing the manuscript. Browne played a supporting role in formal analysis, funding acquisition, investigation and supervision.
Catherina Chang	Contributed equally with co-authors to develop conceptualization, project administration, and the review and editing of the manuscript. Chang played a supporting role in formal analysis, funding acquisition and investigation.
Sandra Hopping	Contributed equally with co-authors to review and editing of the manuscript. Hopping played a supporting role in the administration of resources.

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

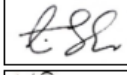
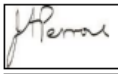
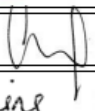
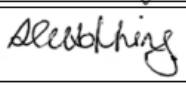
Name	Signature	Date

July 2015

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

Dr Timothy L. Edwards		14/04/2023
Prof. John A. Perrone		13/04/2023
Dr Clare M. Browne		21/04/2023
Catherina Chang		19/4/2023
Sandra Hopping		12/04/2023

Appendix B: Chapter Two Supplementary Material

Supplementary File 1 (S1)

Table 1

Breath Sample Collection and Storage Methods for Canine Scent Detection of Lung Cancer Research

Experiment	Collection Method [Number of Novel Samples]	Storage Method	Sensitivity	Specificity
McCulloch (1)	Exhaled 3-5 times through tube [169]	Polypropylene organic vapor sampling tube; silicone-coated polypropylene wool, stored in sealed bag at room temperature.	0.99 (95% CI [0.99, 1.0])	0.99 (95% CI [0.96, 1.0])
Buszewski (2)	Exhaled 2-3 times through tube [73]	Disposable polypropylene sampling tubes; removable inserts were removed from tubes and stored in sterile polypropylene boxes. No temperature information.	0.82 (CI not available)	0.82 (CI not available)
Ehmann (3)	Exhaled 5 times through tube [220]	Glass tubes; rubber caps, polypropylene wool with silicone treatment, stored in light-tight box at room temperature.	0.71 (95% CI [0.51, 0.88])	0.93 (95% CI [0.81, 0.98])
Walczak (4)	Unclear [627]	Polypropylene sample tubes; cotton wool, polypropylene box, sealed bag, stored at room temperature.	0.79 (CI not available)	0.78 (CI not available)
Rudnicka (5)	3 deep exhalations into tube [253]	Polypropylene tube; oil-coated polypropylene inserts. Inserts transferred to glass vessels for training. Wrapped in a sterile plastic bag, stored at room temperature.	0.86 (CI not available)	0.72 (CI not available)
Fischer-Tenhagen et al. (6)	Exhaled 5 times through tube [60]	Glass tubes; rubber caps, polypropylene wool with silicone treatment. Stored at 8°C.	9/9, 8/9 (CI not available)	8/10, 4/10 (CI not available)

References

1. McCulloch M, Jezierski T, Broffman M, Hubbard A, Turner K, Janecki T. Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. Integr. Cancer. Ther. 2006;5(1):30-9.

2. Buszewski B, Ligor T, Jezierski T, Wenda-Piesik A, Walczak M, Rudnicka J. Identification of volatile lung cancer markers by gas chromatography–mass spectrometry: comparison with discrimination by canines. Anal. Bioanal. Chem. 2012;404(1):141-6.

3. Ehmann R, Boedeker E, Friedrich U, Sagert J, Dippon J, Friedel G, et al. Canine scent detection in the diagnosis of lung cancer: revisiting a puzzling phenomenon. Eur. Respir. J. 2012;39(3):669.

4. Walczak M, Jezierski T, Górecka-Bruzda A, Sobczyńska M, Ensminger J. Impact of individual training parameters and manner of taking breath odor samples on the reliability of canines as cancer screeners. J. Vet. Behav. 2012;7(5):283-94.

5. Rudnicka J, Walczak M, Kowalkowski T, Jezierski T, Buszewski B. Determination of volatile organic compounds as potential markers of lung cancer by gas chromatography–mass spectrometry versus trained dogs. Sens. Actuators. B. Chem. 2014;202:615-21.

6. Fischer-Tenhagen C, Johnen D, Nehls I, Becker R. A proof of concept: are detection dogs a useful tool to verify potential biomarkers for lung cancer? Front. Vet. Sci. 2018;5:52.

Supplementary File 2 (S2)

SOP Number: 20

Handling and Care of Pet Dogs for Research

General:

This document outlines the general procedure for the handling and care of pet dogs being used in research. This protocol has been developed with reference to the Animal Welfare (Dogs) Code of Welfare (Animal Welfare Advisory Committee, 2010, http://www.mpi.govt.nz/protection-and-response/animal-welfare/codes-of-welfare/)

Location/Equipment:

- A maximum of 5 dogs will be present in any of the facilities at any one time.
- Dogs will not be held at the facility overnight. Arrangements will be made by the researcher to meet owners when they drop off their dog before each research session and for pickup afterwards. Under exceptional circumstances, after consultation with supervisors, a researcher may pick up or drop off the dog from the dog owner's home.
- Each dog will be held in a separate kennel, crate, or tie-up station containing bedding, toys, and a bowl of fresh water.
- A logbook will be kept with the name of the dog, arrival time and collection time, and contact details of owner. Dog's owner will sign in upon arrival and sign out when they have picked up their dogs.
- Dogs must wear a collar and be on a lead at all times, except when they are in the kennels or in some cases in the experimental room (depending on the exact requirements of the particular study).
- All dogs must be fully vaccinated up to date for the core diseases: distemper, hepatitis, parvovirus, and leptospirosis. Kennel cough vaccination is also required. Vaccinations are confirmed by sighting and photographing the vaccination certificate; all vaccination records will be stored in a database and reviewed monthly – if a dog's vaccination is due, the owner will be asked to bring the dog's updated vaccination record into the facility and the photograph/record updated accordingly.
- All food will be stored in labelled sealed containers within the storage room.

Record Keeping:

- Pet dogs will be used in studies. A roster of pet dogs that are available will be maintained and kept up-to-date regarding their history of participation and suitability for this type of research.
- The following records will be kept for each dog:
    - Animal's name
    - Owner's name, address and telephone number
    - Emergency/contact telephone number of the owner or their nominee
    - A photo of the dog

- A description of the animal including:
    - sex
    - breed
    - colour
    - age
    - distinguishing features
    - any collars, leads, or belongings brought in with the animal
    - their vaccination status (photograph of the vaccination certificate and storage of expiry dates in a database)
    - Microchip number (if the dog is microchipped)
    - the name and contact number of the veterinarian who normally attends to the animal
    - the dog's normal diet (food type and amount per day)
    - if the dog has any allergies, or any other relevant health issues (e.g., medication).

- The date and number of trials during each experimental session will be recorded for each dog. In addition, the time the dog was last fed and any adverse/unusual observations will be noted.
- During the recruitment process, owners will be asked if their dog has shown aggressive behaviour in the past, and if it is aggressive/protective around food. If a dog has a history of aggression or shows signs of aggression at any time after recruitment, they will not be used for this research. Owners will also be asked if their dog has any other relevant behaviour issues (e.g., fear of certain noises).
- Owners will be asked to sign in their dogs at the lab at the beginning of the day and out at the end.
- Data will be kept for a minimum of 5 years after data collection has ceased.

Cleaning:

- Dog pens and the experimental room will be vacuumed and soiled surfaces will be cleaned with disinfectant at the end of each day after dogs have left.
- If any faeces or urine is deposited inside the facility, they will be disposed of appropriately and cleaned thoroughly. For example, disposable gloves will be worn, and a plastic scoop will be used to remove any faeces which will then be bagged and disposed of in an external rubbish bin outside of the dog facility. The area will then be cleaned thoroughly using an appropriate cleaner.
- A foot pedal rubbish bin will be used to contain general waste within the dog facility, and this will be emptied as appropriate.
- Bowls will be washed thoroughly each day with a disinfectant that is suitable for food surfaces that kills both viruses and bacteria and then rinsed.
- Pest control for insects and rodents will be applied as necessary. These control methods will be used in such a way that dogs cannot access them. If there is any risk of dogs accessing them, a non-toxic (to dogs) control method must be used.

Animal Handling/General Care:

- One person will be in attendance when dogs are present at the laboratory. Another person will be aware and available on call during the running days.
- Before removing a dog from its kennel, external doors to the facility will be closed, and the dog must be put on a lead.
- Dogs will be taken out to walk and toilet every 2 hours, and dog waste disposed of appropriately. Volunteer dog walkers who have been recruited and trained according to Appendix 2 may be used to walk dogs involved in research at the University of Waikato.
- Dog walkers will take care to prevent dogs from accessing bait stations that are positioned around campus. An up-to-date map of the location of these stations will be posted at the exit of each building. Walkers will also prevent dogs from accessing discarded food/trash and, if there is any indication that they have eaten something inappropriate, they will immediately report this to the lab supervisor.
- Any dog showing signs of aggression towards the researcher will cease participation in the study.
- Any dogs showing persistent signs of distress or fear will cease participation in the research and the owner will be contacted.
- Dogs that are transported in vehicles by researchers will always be transported in a manner that is safe and approved by the dog's owners. Dogs will only be transported individually, with the exception of dogs that live together.
- If a dog becomes ill/injured, the laboratory manager will call the owner's vet and take the dog to the vet immediately. If the owner's vet is unavailable, the dog will be taken to the local vet (Newstead Vet) or the after-hours clinic if the local vet is closed. The owner will be contacted as soon as practical. A vehicle must be available (i.e., on campus) for emergency transportation at all times when dogs are present at the facility. If the dog's illness/injury occurred as a result of involvement in the research project, the School of Psychology will pay for the veterinary services. Otherwise, the owner will pay for the services.
- Multiple dogs will only be present at the facility at the same time if the owners give permission for this to occur, and if they state that their dog is friendly towards other dogs (as part of the consent form). The dogs will be kept on leads around each other and held in separate kennels/crates/tie up stations (unless they are from the same household and the owner prefers them held together). Only dogs that are confirmed to be reliably friendly toward each other (including those who live together) will be walked together. If any conflict behaviour is seen between dogs, then the dogs will be separated from each other and the senior researcher will be contacted.

Building Security:

All researchers will ensure the building is locked and secure (windows and curtains shut) before leaving. If there are any security concerns, University of Waikato security will be called on 07 838-4444.

Emergency Evacuation Procedures:

If the personal safety of the staff or researchers are not compromised the dogs will be lead one at a time (if possible) to the fence adjacent to the FMD building where tie-stations have been installed for this purpose. Each dog will be secured to one of the tie-stations, which are situated such that no dog can be in physical contact with another. Five chew-proof leads designated for this purpose are hanging by the main exit of both buildings. Dog owners will be contacted as soon as possible and temporary provision of water will be made to the dogs.

Versions and Reviews: Version 2.1
Date revised: October 2019
Date approved: 18 October 2019
Next revision due: 19 October 2022

Standard Operating Procedure for Training Dogs to Use

Automated Scent-Detection Apparatus

NOTE: Procedures for dog selection, habituation, handling, and care have been omitted, as requirements are likely to vary among laboratories. The complete standard operating procedures specific to the author's laboratory are available upon request.

Apparatus setup

Position the apparatus in a room without other objects that might distract the dog. Only the front panel of the apparatus should be accessible to dogs. Moveable partitions may be used to block access to the other sides of the apparatus. The room must have a door that closes/latches and should be equipped with one or more cameras that can be used to monitor the dogs. The computer(s) used to control the apparatus and monitor the dogs should be positioned in an adjacent room.

Training

1. Introduction
   Once the dog is habituated to the environment and researchers, training sessions can be started. During the shaping and early training process, at the first sign of fatigue/disinterest, the session should be terminated, ideally immediately following a correct response and reinforcement. Early shaping/training sessions should not exceed 10 minutes.

2. Conditioned reinforcer establishment
   Dispense food from the automatic feeder using the remote/hand-switch until the dog immediately approaches the feeder upon hearing the sound made when it is activated. The dog should approach the automatic feeder and consume the food within 3 seconds 3 times in a row to continue to the next stage of training.

3. Shaping: nose in sample port
   Once the sound of the feeder is established as a conditioned reinforcer, the remote/hand-switch is used to train the dog to put its nose into the sample port of the apparatus. Use the method of differential reinforcement of successive approximations to this target behaviour (see Appendix). Prompting (e.g., pointing) may be used but the prompt must be faded and removed before proceeding to the next step (lever activation).

   As soon as the nose is first inserted into the port, the dog is taken out of the room and the apparatus is loaded with positive samples and set to treat indications of 1000ms as indications (500ms as observations). Continue shaping as required until the dog begins to trigger the feeder automatically. Once a session (17 samples) at the 1000ms threshold is complete, increase the threshold to 1500ms. Once a round at 1500ms is complete, continue to the next step.

4. Shaping: lever activation
   With the apparatus turned off, use the method of differential reinforcement of successive approximations to shape lever pressing (see Appendix). Depending on the size and the behavioural tendencies of the specific dog, an appropriate topography should be selected for shaping (e.g., use of paw or nose to activate lever). Prompting (e.g., pointing) may be used but the prompt must be faded and removed before proceeding to the next step. Once the lever has been activated 10 times without prompts (and reinforced via manual activation of the feeder), proceed to the next step.

5. Discrimination training
   Load the apparatus with approximately half positive and half negative samples (e.g., 8 and 9 samples,

respectively), alternating positive and negative with the first sample positive (enter the status of the samples into the session configuration file). Ideally, samples should contain a high concentration of the target/control substance. Set the apparatus with a threshold of 1500ms for the indication response and 500ms for the observation response.

Put the dog into the experimental room and stand beside the apparatus. If there is no response to the apparatus within 20 seconds, prompt as required. On the first negative trial, also allow 20 seconds before prompting to see if the lever press occurs without prompting. Continue prompting, if necessary, but fade prompts as soon as possible (i.e., wait for an increasing amount of time before prompting). Prompt with a consistent cue - finger pointing to sample port or lever.

Once a round (of 17 samples) is completed without prompting, randomize the sample arrangement in subsequent sessions. Continue until hit rate is above 80% and the dog is working reliably without prompting. The experimenter should gradually remove themselves from the room and, once the dog is successfully working on its own, systematically increase the indication threshold in 500 or 1000ms increments until the target threshold is reached (5500ms is generally optimal based on our preliminary research, but this may vary depending on the dog/application).

With a standard sample (e.g., amyl acetate) at a high concentration, hit rate and correct rejection rate should reach and stay at approximately 100%. At this point, additional rounds may be added (i.e., the samples can be presented two or more times in one session) or the sample concentration, type, or distribution (e.g., positive sample prevalence) can be changed systematically as required.

6. Troubleshooting

Problem	**Solution**
The dog is performing poorly in training.	- Make sure the dog is healthy. Deal with any health-related issues first. - Confirm that food is an effective reinforcer by evaluating approach and consumption and/or by attempting to shape a simple response. If current food is ineffective, confirm that dog is not being overfed prior to training and, if confirmed, select a different food (using a paired-choice preference assessment procedure[1]). - Return to earlier stages of training as required (e.g., if the lever press is not occurring reliably in discrimination training, conduct another lever press shaping session in isolation). - Check factors related to sample quality (make sure that samples have been prepared and arranged as specified in the specific sample preparation SOP) - If the dog continues to perform poorly, the dog will cease participation in the study. Consult with supervisor.

[1]Vicars, S. M., Miguel, C. F., & Sobie, J. L. (2014). Assessing preference and reinforcer effectiveness in dogs. *Behavioural Processes*, *103*, 75-83.

Appendix: Guidelines for Shaping

1. Introduction

   This document outlines the basic training hierarchy for shaping by successive approximations. As a general rule, each step must be completed 3 times in a row before progressing to the next stage of training. Some dogs, however, may require additional learning trials before progressing. Keep sessions short (5-10 mins) and finish on a positive note when possible to ensure that the process is enjoyable for the dog.

2. Procedure

   2.1. Researcher is to position themselves near the apparatus, ideally near the door, avoiding the dog's gaze to reduce unintentional cueing. This will facilitate fading of the researcher's presence during later trials when the dog is required to be in the experimental room alone. Gestural prompts may be used to facilitate training, but these should be used only as needed as they must be faded out before training is complete.

   2.2. Shaping of sample port entry

      2.2.1. Turn apparatus on and start a session so that the sample port beam breaks will produce a "beep."

      2.2.2. Reinforce moving further and further away from the feeder until the dog is reliably approaching the other side of the room (near apparatus).

      2.2.3. Reinforce attending to the apparatus (putting nose near any part of the front panel).

      2.2.4. Reinforce placing nose near port.

      2.2.5. Reinforce placing nose in port.

      2.2.6. Reinforce breaking the beam for any length of time (indicated by "beeping" sound).

      2.2.7. Reinforce touching the segment flap with nose.

      2.2.8. Reinforce pushing the flap inwards.

   2.3. Shaping of lever press

      2.3.1. Turn apparatus off.

      2.3.2. Reinforce any movement toward lever.

      2.3.3. Reinforce movement of nose or paw toward lever (as appropriate).

      2.3.4. Reinforce any contact with the lever (nose or paw, as appropriate).

      2.3.5. Reinforce any movement of the lever.

      2.3.6. Reinforce movement of the lever that produces the "click" (microswitch closure).

Supplementary File 3 (S3)

Methods

Apparatus and Setting

The dogs completed initial scent detection training (target: amyl acetate) to use the automated apparatus with a trainer present in the experiment room, then transitioned to working alone. Once the dogs demonstrated fluency using the apparatus, the samples were switched to breath and saliva samples, with samples from lung cancer-positive patients serving as the positive samples.

Breath and saliva samples were presented to dogs on an automated apparatus, consisting of a carousel with 17 removable aluminium chambers sandwiched between two stainless steel plates. A single sample was placed in each chamber. Each chamber was fitted with an inward-opening hinged flap that remained closed unless pushed open. A 10 cm-diameter aperture in the centre of a 1 m^2 acrylic panel mounted on the front of the apparatus allowed the dogs access to one chamber at a time. A grid of infrared sensors directly behind the aperture measured the duration of the dogs' access to each chamber. The dogs' response durations were used to determine whether they had evaluated samples; a duration of at least 250ms constituted an "observation" response. A switch on the right-hand side of the front panel could be activated by the dogs to rotate the carousel by a single chamber once the observation response occurred. A second, longer response duration, typically 4500 ms but adjusted for individual dogs, constituted an "indication" response. Indication responses to positive samples were reinforced using a remote-controlled kibble dispenser (Treat and Train Remote Dog Trainer, by Pet Safe), positioned 1.5 m from the apparatus. The experiment room housing the apparatus was 3.2 m by 4.3 m, maintained at approximately 20°C. A computer in an adjacent room controlled the apparatus and was connected to two cameras

that provided a live feed of the dogs' activities. See (1,2) for further details of this apparatus and details of training processes and results.

Sample Information

Samples provided by 490 human participants recruited from the Chronic Obstructive Pulmonary Disease clinic of a local hospital were used for training purposes. Human ethics approval was obtained from New Zealand's Health and Disability Ethics Committees (17/NTB/178/AM02). Following informed consent, breath and saliva samples were collected from each participant by two nursing staff in a dedicated consulting room. Additional data including demographics, the subject's smoking/vaping status, time to last meal, time to last smoking were also collected. Participants were referred by their general practitioners to the general respiratory clinic and had not yet been assigned a diagnosis and participation was offered as part of the clinic's diagnostic process. Thus neither study staff nor participants were aware of the participants' cancer status at recruitment. Procedures, materials, personnel, and location remained the same throughout the recruitment period.

Breath samples for each participant were collected in four annealed glass tubes (Duran tubing, 120 mm long, 26 mm OD, 2 mm wall thickness) packed with 1 g of L10Y4 polypropylene fibre (manufactured by IFG Asota) sandwiched between two halves of a split single cotton wool ball. Participants were instructed to "Take a deep breath and breathe slowly through the tube. Exhale completely, until there is no air left in your lungs. Breathe three times through each of the tubes." The nurses sealed each tube with plastic caps (tapered plug/cap LDPE red 23.62 D MX, manufactured by Hi-Q Components), placed each tube in its own sealable plastic bag, then sealed all four of the capped and bagged samples from the participant into a larger plastic bag.

A single saliva sample was also collected from each participant. 10 ml of tap water in a disposable plastic cup was offered to the participant with the request to “rinse your mouth with the water provided (do not swallow the water).” This mixture of saliva and tap water was then expectorated into a hospital-issue 100 ml polypropylene sample cup which was immediately closed, sealed in a plastic bag, and placed in the larger plastic bag holding the breath samples.

Samples were immediately stored in a -80°C freezer at the hospital, then transferred using dry ice to -80°C or -60°C freezers at the University of Waikato Scent Detection Laboratory. The disease status of the patients was prospectively adjudicated by a specialist respiratory physician and medical oncologist independently, with any disagreements resolved by consensus. Lung cancer diagnoses were confirmed and staged by the local lung cancer multi-disciplinary team; these data were anonymised and then shared with the scent-detection researchers. Each week 17 samples were identified for training purposes and presented to the dogs in the apparatus in a randomized array. Between 2 and 5 positive samples were presented in each set of 17 on training days, while negative samples made up the remaining 12-15 samples. The number of positive and negative samples from smokers were presented in equivalent proportions to prevent the dogs from developing a smoking-related bias. When a saliva sample from a specific patient was selected for the first time, the sample was thawed and aliquoted into 3 x 3 ml volumes into annealed 5 ml glass jars (28 mm interior diameter, 9 mm interior depth) sealed with silicone septa prior to the experimental day. Two of the three saliva aliquots were used for training purposes; the third was retained for chemical analysis. All were frozen again immediately after redistribution. On the day of the experimental sessions, after reaching room temperature, caps were removed from breath sample tubes or the saliva jars, which were placed directly into the apparatus. Training samples were thawed, opened, used, closed, and refrozen up to four times (3).

Procedure

The dogs visited the laboratory and trained two days per week, either mornings or afternoons. Each day, dogs completed between 2 and 10 sessions. A session consisted of one complete rotation of the 17-segment carousel. Saliva samples were used on the first of the two days; breath samples from the same participants were used on the second day. After 18 weeks of training, breath and saliva sample presentation was reversed to evaluate possible order effects. A combination of novel and used samples were included in each training array. Used samples had been presented to the dogs for at least one training day previously, then refrozen.

Training and Error Correction

During training, indication duration requirements for a sample to be treated as "indicated" by the dogs began at 501 ms and were increased by 100 ms after each subsequent successful session at that value. These values were increased to an upper indication threshold of 3500 – 4500 ms, based on optimal performance for each dog. The indication duration requirements were adjusted separately for each sample type.

In instances where a dog missed (failed to indicate) a positive sample, the indication threshold for that individual positive sample was lowered to facilitate acquisition of the lung cancer target. The size of the threshold reduction was determined by the length of the initial sub-threshold indication; the new threshold was set 500 ms below the duration of the dog's initial indication or reduced to 500 ms if the initial value was lower than 1000 ms. For example, if an indication threshold of 4000 ms was set for a dog's initial indication response to a novel sample, but the dog only indicated for 1200 ms before withdrawing, the threshold for that specific sample during the next training session would be lowered to 700 ms. In the same scenario, if the dog only indicated for 900 ms before withdrawing, a requirement of 500

ms would be applied in the next session. The indication threshold was raised by 100 ms in each subsequent session as the dog experienced reinforcement for correct indication of the sample. If the dog missed a sample with a lowered threshold, the threshold was reduced further, following the same procedure.

References

1. Edwards. Automated canine scent-detection apparatus: technical description and training outcomes. Chem. Senses. 2019;44(7):449-55.

2. Edwards TL, Giezen C., & Browne CM. Influences of indication response requirement and target prevalence on dogs' performance in a scent-detection task. Appl. Anim. Behav. Sci. 2022;105657.

3. Crawford MA, Chang CL, Hopping S, Browne C., Edwards, TL. Influences of breath sample re-use on the accuracy of lung cancer detection dogs. J. Breath Res. 2022; 016001–

Supplementary File 4 (S4)

Data Analysis

$$\log d = \frac{1}{2}\log\left[\left(\frac{hit}{miss}\right)\left(\frac{correct\ rejection}{false\ alarm}\right)\right]$$

$$\log b = \frac{1}{2}\log\left[\left(\frac{hit}{miss}\right)\left(\frac{false\ alarm}{correct\ rejection}\right)\right]$$

Appendix C: Chapter Three Supplementary Material

Supplementary Data

Table 1

Individual Hit Rate/Sensitivity Values

Name	0	1	2	3
Tui	76.02941	67.95455	35.25499	
Katie	74.53113	77.83654	59.77273	
Bramble	76.66667	85.89744		
Onyx	76.05311	74.29617	75.26991	
Rylea	81.25889	89.89899		
BJ	73.06608	65.48345	66.01831	50
Rocky	51.78819	61.81284	58.40125	61.90476
Mean	72.7705	74.74	58.94344	55.95238

Table 2

Individual Correct Rejection Rate/Specificity Values

Name	0	1	2	3
Tui	94.57891	75.78755	89.94969	86.15385
Katie	65.03038	57.51838	78.92512	67.5
Bramble	73.82366	78.0303		
Onyx	82.27538	83.41766	77.18131	85.83333
Rylea	45.20238	59.36911		
BJ	67.33092	79.80226	87.03704	
Rocky	76.43438	76.57516	77.66813	73.12253
Mean	72.09657	72.92863	68.46021	78.15243

Table 3

Individual log B Values

Name	0	1	2	3
Tui	-0.37107	-0.08424	-0.60838	
Katie	0.097826	0.20708	-0.20883	
Bramble	0.034504	0.114659		
Onyx	-0.08052	-0.12072	-0.02008	
Rylea	0.370207	0.393314		
BJ	0.059349	-0.15532	-0.25856	
Rocky	-0.24287	-0.15051	-0.19731	-0.12344
Mean	-0.01894	0.029182	-0.21553	-0.12344

Table 4

Individual log d Values

Name	0	1	2	3
Tui	0.872426	0.411377	0.344246	
Katie	0.368984	0.338805	0.389647	
Bramble	0.486192	0.677129		
Onyx	0.586458	0.582237	0.509197	
Rylea	0.286023	0.558231		
BJ	0.374123	0.447709	0.603162	
Rocky	0.274051	0.364317	0.344659	0.336425
Mean	0.464037	0.482829	0.396937	0.336425

Lung Cancer Detection Project

Sample Handling and Presentation
Standard Operating Procedure
05/05/2019
Margaret Crawford

Preparation – Breath Samples

1) Selecting Samples – Weekly Laboratory Preparation

1.1 Determine number of positive, control and/or blank samples for training session.
1.2 Select samples.
1.3 Create sample list sorted by freezer/box number.
1.4 Retrieve bags from freezer boxes and place in numerical order in work box/es, ready for use.
1.5 Complete random configuration form (1 – 17) to determine where the samples will be placed in the apparatus.

2) On Training Day

2.1 Wipe large work table with isopropanol at least 8 hours earlier. Air room.
2.2 Postit notes recording sample number and date of use placed on the table in order of presentation to indicate placement of trays in 6/6/5 configuration with as much space between trays as possible.
2.3 Place 17 trays on table

3) Transferring Breath Samples from the Freezer for Training

3.1 Note the time.
3.2 Put on latex gloves.
3.3 Open freezer, then work box. Remove sample 1 *main bag* containing 2 x individually bagged and labelled breath samples in sealed tubes,
  3.3.1 Use b*oth hands* to open *main bag* and r*ight hand* take breath sample from first sample *main bag* and place on first sample tray,
  3.3.2 Use *left hand* to seal and return *main bag* to training box in numerical order and close freezer.
3.4 Change latex gloves, left hand first simultaneously inverting and removing right hand from the base of glove.
3.5 Repeat this process 16 more times.
3.6 Finally, close the freezer.

Allow 30 minutes to elapse between the removal of the first sample from the freezer and the transfer of samples from the trays to the apparatus. This allows all sealed samples to return to room temperature.

While waiting:

3.7 Ensure the door to the experiment room is sitting ajar.
3.8 Chambers should be placed on the apparatus, ready to receive sample containers.
3.9 Run AC in experiment room at 20C. When ambient temperature reaches 20C, record time.

4) Transferring Breath Samples from Tray to Apparatus

4.1 Note the time.
4.2 Put on latex gloves.
4.3 Open first sample bag and remove the sample tube (located in Tray 1). Remove tube lids (use box cutter if necessary to open tape) then carry sample to chamber, and place both lids interior sides up in chamber next to the tube. Both tube openings should be exposed.
4.4 Replace gloves. Repeat process 16 times.

When finished, note the time, place the lid on the carousel so the sample chambers are enclosed and leave for 10 minutes while headspace VOCs accrue. Record the ambient room temperature.

5) Unloading Apparatus from the Carousel to the Trays

5.1 Remove the carousel lid. Open freezer. Remove lid from work box.
5.2 Put on latex gloves.
5.3 Retrieve sample from chamber 1. Secure lids using tape, being careful not to touch the tape dispenser. Carry sample to corresponding tray. Place sealed tube in bag **along with the corresponding tray postit note** in the sample bag, being sure to twist the bag around the sample and expel as much air as possible.
5.4 Replace gloves.
5.5 Repeat until all samples are bagged and waiting on trays.

6) From the Trays to the Freezer

6.1 Change Gloves.
6.2 Remove first sample *main bag* from work box, use *both hands to* open it, *right hand* under base of main bag, use *left hand* to drop in bagged sample 1 tube using thumb and forefinger. Middle finger and base of hand used to hold main bag while *right hand* closes zip lock.
6.3 Sample 1 *main bag* is returned to work box with *right hand* and filed numerically.
6.4 Repeat process until all samples are returned to the freezer.
6.5 Note the time when the last sample is returned to the work box and the freezer is closed.

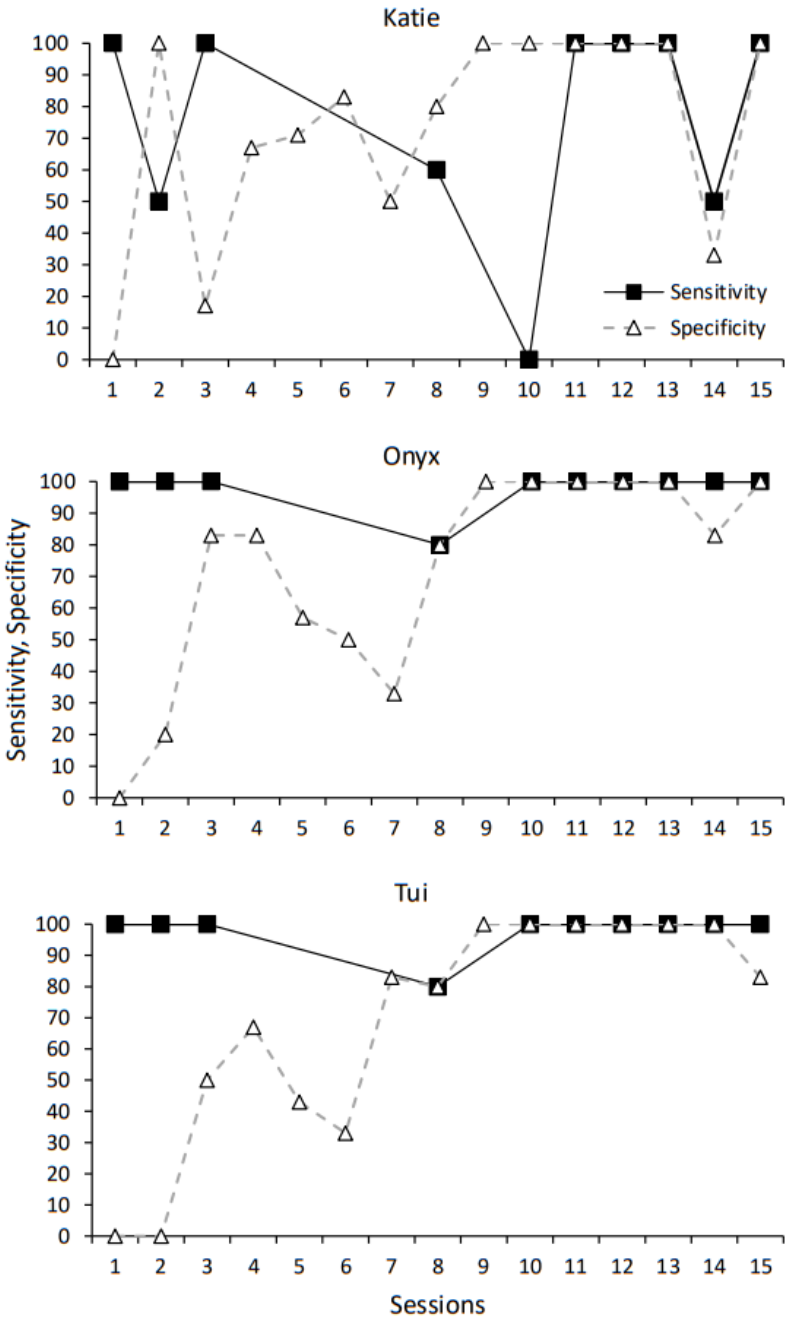
Miscellaneous: Destruction of Waste Materials

- Any waste products exposed to breath that cannot be sterilized via acid bath will be stored in 'sharps' medical waste containers or sealed plastic bags and returned to the hospital for disposal.

Appendix D: Chapter Four Supplementary Material

Supplementary File S1

Sensitivity and specificity of three dogs when evaluating novel patients at the outset of training, demonstrating anomalous, rapid approach to near-perfect accuracy.



Appendix E: Chapter Four Quantitative Analysis Model Selection

How our quantitative model was selected

The researchers reviewed and discarded a range of dependent and independent measures as they evaluated which form of analysis might provide the most pertinent information about optimal transition from training to testing. The purpose of model fitting was two-fold; firstly, to determine the structure of the dogs' concept formation as they acquired the lung cancer-positive target; and secondly, to discover a quantitative estimation of the rate of acquisition. This rate could be used to estimate the impact of challenges like breaks in the training schedule, use and re-use of sample materials, and/or individual differences in participants. The key considerations of scale, limits/no limits, shape, number of parameters and how to measure acquisition for binary responses informed decision-making.

Dependent measures – structural building blocks

Before selecting the sum of sensitivity and specificity measures, other diagnostic accuracy measures including positive predictive value and negative predictive value were considered. Signal detection theory includes *d'*, the Receiver Operator Characteristic (ROC), A and A' as alternatives.

Positive predictive value, negative predictive value, sensitivity and specificity

Positive predictive value (PPV) identifies how many known positive samples are present as a percentage of all samples indicated positive. This is done by dividing correct positive indications by the sum of correct indications and false alarm indications. Similarly, negative predictive value (NPV) identifies how many known negative samples are present as a percentage of all the samples indicated negative. PPV and NPV calculations rely on comparable disease prevalence across populations for meaningful comparisons. Meanwhile, sensitivity is calculated by dividing correct positive indications by the sum of correct positive indications and missed indications of positive samples. So long as there are no significant variations in diagnostic emphasis or major differences in participant characteristics, actual

disease prevalence does not influence sensitivity. This is also the case for specificity, a measure generated by dividing correct rejections by the sum of correct rejections and false alarms (Wong & Lim, 2011).

Signal detection theory

d' is a statistic that represents the extent to which "noise" and "signal + noise" (represented as curves) are discriminable to a detector (Stanislaw & Todorov, 1999). It assumes a normal distribution and similar variance of both the noise and signal + noise distributions during detection activities. We discovered that the proportion of hits and false alarms in individual dog's responding to the lung cancer disease target during acquisition did not follow a pattern of normal distribution.

We also considered applying the Receiver Operator Characteristic (ROC) to understand the dogs' target acquisition more clearly (Baker, 2003). The ROC is generated by plotting false alarms (X-axis) versus sensitivity (Y-axis). By manipulating biases in responding, false alarm and sensitivity values can be obtained which, when graphed, show a curve. This curve can be used to identify performance cut-off points depending on the nature of the measure required from the detector organism. As we did not systematically manipulate bias, and required a representation of change over time, we rejected the ROC. *A* and *A'* are non-parametric best estimates of area under the ROC curve calculated by averaging a range of highest to lowest ROC curve values that can be generated from a single data point (Zhang & Mueller, 2005). However, we concluded that *A* and *A'* are neither intuitive nor transparent measures that would strengthen our analysis.

Independent measures – structural alternatives

The following models of acquisition were considered before the Weibull function was selected: split-middle line of progress; linear regression; cumulative Gaussian function, recursive algorithm, the Avrami equation and the Boltzmann equation.

The structure of acquisition

Split-middle line of progress and linear regression

The Split-middle line of progress is a multi-step measure that identifies a trendline using median values (Lane & Gast, 2014). It is unable to accommodate asymptote, which is a defining feature of the acquisition data associated with target acquisition. Analysis using linear regression is also unable to accommodate asymptote, therefore neither of these measures were selected.

Cumulative Gaussian function

The cumulative Gaussian function (as the name suggests) assumes a relatively normal distribution of data; our data were not normally distributed. Also, cumulative Gaussian function analysis involves the fitting of a curve that begins at 0 and ends at 1.0. Meanwhile, our project data began at chance responding, or approximately 0.5 and (although a maximum value of 1.0 was possible), appeared to show asymptotic performance for fully trained dogs at a value of approximately 0.7.

Recursive algorithm

The recursive algorithm used by Gallistel et al. (2004) was generated by using binary data. The dogs' individual responses in our project were recorded in zeros and ones, but our preferred unit of analysis was the session rather than the individual response. Session data were generated by calculating the mean of sensitivity plus specificity results derived from 17 consecutive responses to a random mixture of positive and negative samples. In practice, a session involved one full rotation of the automatic training apparatus. This decision to use session data was made to reduce the noise inherent in too molecular a unit of analysis. As is apparent from the data included in Crawford et al. (2022), even session data generated during the training process were very noisy prior to further interpretation.

Avrami equation

The Avrami equation is traditionally used to understand the kinetics of crystallisation and other changes in the phase of materials (Cantor, 2020). It was selected because of its ability to produce a sigmoid curve with potentially predictive qualities. However, the Avrami equation was rejected when it became apparent that the plotted curve must start at 0 and end at 1.0 for the function to work.

Boltzmann equation

The Botlzmann equation (Cantor, 2020) was selected when it appeared that the sigmoid curve it fitted could start at different values on the Y-axis, potentially accommodating our 'at chance' responding of 0.5 when training commenced. Visual analysis of the results of its application to our project showed that the fit was poor – especially in instances where data were intermittently unavailable. On the basis of these findings, we settled on the Weibull function, the features of which are described in the manuscript in Chapter 4.