Remote Medical Scent Detection of Cancer and Infectious Diseases With Dogs and Rats: a Systematic Review

Pierre BAUËR (pierre.bauer.1@gmail.com)
Institut Curie https://orcid.org/0000-0002-2598-593X
Michelle LEEMANS
Universite Paris-Est Creteil Val de Marne Faculte de medecine
Etienne AUDUREAU
Université Paris-Est Créteil Val de Marne Faculté de médecine: Universite Paris-Est Creteil Val de Marne Faculte de medecine
Isabelle FROMANTIN
Institut Curie

Methodology

Keywords: Medical scent detection, VOCs (=volatile organic compounds), diagnostic, screening, cancer, infectious disease, smell, odour

Posted Date: August 3rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-753298/v1

Abstract

Background: Remote medical scent detection of cancer and infectious diseases with dogs and rats has been an increasing field of research these last 20 years. If validated, the possibility of implementing such a technique in the clinic raises many hopes. This systematic review was performed to determine the evidence and performance of such methods and assess their potential relevance in the clinic.

Methods: Pubmed and Web of Science databases were independently searched based on PRISMA standards. We included studies aiming at detecting cancers and infectious diseases affecting humans with dogs or rats. We excluded studies using other animals, studies aiming to detect agricultural diseases, diseases affecting animals, and others such as diabetes and neurodegenerative diseases. Only original articles were included. Data about patients' selection, samples, animal characteristics, animal training and testing configurations, and performances were recorded.

Results: A total of 62 studies were included. Sensitivity and specificity varied a lot among studies: While some publications report low sensitivities of 17% and specificities around 29%, others achieve rates of 100% sensitivity and specificity. Only 6 studies were evaluated in a double-blind screening like situation. In general, the risk of performance bias was high in most evaluated studies, and the quality of the evidence found was low.

Conclusions: Medical detection using animals' sense of smell lacks evidence and performances so far to be applied in the clinic. What odours the animals detect is not well understood. Further research should be conducted, focusing on patient selection, samples (choice of materials, standardization), and testing conditions. Interpolations of such results to free running detection (direct contact with humans) should be taken with extreme caution.

1. Background

1.1. The burden of cancer & infectious diseases worldwide

Cancer and infectious diseases are considered major health issues among men and women and are among the most common cause of morbidity and mortality worldwide. On the one hand, there were an estimated 18.1 million (95% UI: 17.5–18.7 million) new cases of cancer (17 million excluding non-melanoma skin cancer) and 9.6 million (95% UI: 9.3–9.8 million) deaths from cancer (9.5 million excluding non-melanoma skin cancer) worldwide in 2018 (1). On the other hand, infectious diseases such as urinary tract infections (UTI), tuberculosis, Clostridium difficile infections, Methicillin-resistant Staphylococcus aureus (MRSA), pandemic outbreaks like Ebola and lately Severe acute respiratory syndrome coronavirus 2 (SARS-COV-2), are also claiming the lives of many people.

1.2. Disease early and rapid detection

Disease detection is the first step prior to diagnosis and care. However, detection is not easily accessible everywhere for everyone on the planet. Efforts to control infectious diseases or detect cancers early would benefit from new screening technologies (2, 3). For instance, early diagnosis could reduce mortality for many cancer types (4). As well, quick, reliable, and widespread testing is vital to control a pandemic. Validation of

diagnostic tests is therefore crucial. Performances of such tests are evaluated by their accuracy: sensitivity, specificity, predictive values and likelihood ratios (5).

1.3. Need for a noninvasive low cost and reliable detection method

Diagnostic relies on direct imaging or on collecting samples from individuals or contaminated environments, transportation of samples to a laboratory, and subsequent laboratory testing to demonstrate the presence or the absence of the pathogen of interest. This results in a significant delay in response times and containment efforts. These procedures can be invasive and require skilled human resources and costly equipment and consumables depending on the diseases. A desirable screening method should be noninvasive, painless, inexpensive, and easily accessible to many patients. In addition, it should allow diagnosis at early-stage (6).

1.4. Diseases emit VOCs

The human body emits Volatile Organic Compounds (VOCs) linked to its endogenous process, after absorption of external contaminants (e.g. food) and by bacterial metabolism (e.g. armpit odour) (7–10). VOCs are organic chemicals with a high vapour pressure at typical room temperature, resulting in evaporation or sublimation of the molecules into the air surrounding the source. It has been shown previously that some diseases emit specific VOCs (10). Disease-related VOCs may be found in the blood, breath, faeces, skin, sputum, sweat, urine, and vaginal secretions of affected individuals. Such a signal could pave the way to a new detection technique: using VOCs as biomarkers for disease detection. Research investigating the VOCs profiles associated with various human diseases is underway, primarily driven by the goal of developing instrumentation for use in clinical diagnostics.

1.5. VOCs analysis complexity

Currently, intensive studies are being carried out to identify compounds that could be markers of cancer (11, 12) and could eventually support or even replace traditional screening methods. To do so, techniques such as gas chromatography-mass spectrometry (GC-MS) have already been developed, and several research teams and companies aim at developing bioelectronic noses (13). Currently, the development of this technology is limited by the high cost of the necessary laboratory instrumentation and difficulties in standardizing sample collection and preparation procedures in clinical settings (14). These limitations can, for instance, be due to threshold, non-optimized odour capturing materials, low signal-to-noise ratio, costs, the complexity of both the chemical signature and the subsequent data analysis. It is worth noticing that the origin and the nature of VOCs emitted by cancers are not well understood. Whether the chemical signature originates from the tumour, from the tumour environment, or both is still under investigation.

1.6. Animal sense of smell & anecdotal reports

Such technologies might not be the only solution. Indeed, these last decades, several studies evaluating and reporting the potential ability of trained animals to detect certain diseases thanks to their sense of smell have raised many hopes. Sense of smell has been extensively studied and is reported to be highly developed among certain species (15). Primarily, the canine sense of smell has been deeply investigated (16). Dogs have been trained to locate explosives, illicit drugs, banknotes, missing persons, disaster victims (17, 18). Rats and several other animals have also been successfully trained to identify targeted substances (19).

Animal olfactory detection of human diseases has attracted an increasing amount of interest from researchers in recent years. In 1989, a first case was reported where a dog seemed to have detected his masters' melanoma(20). Similar cases have been reported in the following years (21, 22).

1.7. The consistent use of animals to detect diseases

These case reports allow emitting the hypothesis that some animals could potentially be used to detect diseases. However, these anecdotal findings alone do not mean that animals can be used as systematic and reliable tools to detect diseases. This potential new tool must therefore be further explored and developed following a scientific method.

Several structured research programs have reported the abilities of dogs, rats, and other animals to detect diseases such as cancers, diabetes, epilepsy, tuberculosis, malaria, urinary tract infections (UTI), SARS-COV-2 among others. These studies focus on animal capabilities and research and optimize sampling protocols and materials, storage and use of odours, scent lineups parameters, animal welfare, testing conditions, etc. To do so, research programs usually gather several professionals such as medical staff, chemists, biologists, physicists, statisticians, data scientists, vets, ethologists, and dog handlers.

1.8. Still many unanswered questions

Because of the inconsistent findings reported in this body of research and the complexity of scent detection research, it seems complicated to ascertain the potential value of animal detectors in diagnostic. Indeed, despite the number of studies reporting the ability of trained animals to detect diseases, it is not always known what they detect, as hundreds of VOCs are released from the human body. With a rising number of publications tackling this issue, carrying out a structured and objective state of the art seemed necessary.

1.9. Objectives

In this systematic review, we aim at outlining the performances (sensitivity, specificity) of trained dogs and rats in distinguishing cancers or infectious diseases cases from controls in humans, thanks to their sense of smell, published in peer-reviewed research. Additionally, methodological issues leading to inconsistencies among research are reviewed, and further recommendations to improve performances are given. We excluded studies using other animals (nematodes (23), insects (24, 25)), studies aiming at detecting agricultural diseases, animal infections (dogs, cows, ducks (26–29)), and other diseases (hypo/hyper-glycemia, neurodegenerative diseases). Only original articles were included, and reviews were excluded.

2. Methods

2.1. Literature Search

Pubmed and Web of Science were independently searched based on the standards of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (30). Studies about detecting cancer and infectious diseases (were excluded hypo/hyper glycemia, neurodegenerative diseases) in humans (agricultural diseases, animal infections and animal cancers were excluded) by dogs, rats or mice sense of smell (we excluded studies using other animals, for instance, nematodes or bees), in the databases from 01/01/2000 to 01/05/2021 were retrieved. The search strategy was adjusted to each database. For PubMed the following string was employed:

(("Dogs"[Mesh]) OR (canine*) OR ("Rats"[Mesh]) OR ("Mice"[Mesh])) AND (("Volatile Organic Compounds"[Mesh]) OR (Volatile AND Organic AND Compound*) OR ("Odorants"[Mesh]) OR (odor*) OR (odour*) OR ("Smell"[Mesh]) OR (nose) OR (scent*) OR (sniff*) OR (olfact*)) AND (("Disease"[Mesh]) OR ("Neoplasms"[Mesh]) OR (cancer)). For Web of Science the subsequent sequence was researched: (TI= ((dog* OR canine* OR rat* OR mouse OR mice) AND (cancer* OR neoplasm* OR disease*) AND (smell* OR scent* OR sniff* OR olfact* OR odo$r* OR volatomic* OR (volatile organic compound*) OR (volatile* AND organic* AND compound*)))) OR (AB = ((dog* OR canine* OR rat* OR mouse OR mice) AND (cancer* OR neoplasm* OR disease*) AND (smell* OR scent* OR sniff* OR olfact* OR odo$r* OR volatomic* OR (volatile organic compound*) OR (volatile* AND organic* AND compound*)))). The references of retrieved literature were further reviewed for original articles. Non-human diseases detection and articles written in a language other than English were excluded.

2.2. Study selection and eligibility criteria

In total, 5,665 records were identified, 2,057 from PubMed, and 3,608 from Web of Science, of which 210 were duplicates (Fig. 1). We reviewed the remaining 5,455 titles and abstracts to identify studies relevant to the topic. Three authors (P.B., M.L. and I.F.) reviewed the abstracts and/or full-text manuscripts independently and selected those that were regarded to be relevant. No disagreement on the selection of articles was seen between the three reviewers. The inclusion criteria were for studies on diseases detection with dogs, rats or mice, in original articles. Articles that described original research involving animal olfactory detection of human disease using samples collected from human participants were selected for inclusion. Review articles, articles not directly relevant to the topic were excluded. Sixty-two full-text papers were reviewed for inclusion; none were excluded after full-text review. A total of sixty-two papers were included in the systematic review.

2.3. Data extraction

The relevant data were extracted from the 62 selected articles. A standardized table was designed to abstract the studies of interest. Details abstracted from each study included: Title, authors, DOI, journal, year, country (where detection took place), type of publication, disease(s) of interest, number of hospitals, sample type, the material used for sampling, sampling protocol, sample conservation, sample presentation, animal class, animal number, animal details, animals' prior experience, animal other uses in parallel, animal living conditions, animal training type, animal training duration, animal % selected successfully, animal training & test setting, non-blinded single-blinded or double-blinded test, temperature, hygrometry, pressure, sensitivity, specificity, positive samples type, control samples type, number of samples for training, number of positive samples for training, number of negative samples for training, number of samples for testing, number of positive samples for testing, number of negative samples for testing, controls reused or not, gold standard, potential bias, and remarks.

3. Results

3.1. Systematic presentation and synthesis, of the characteristics and findings of the included studies.

The last 20 years have seen an increase in the number of publications dealing with diseases detection with dogs and rats (Fig. 2). Starting with case reports in 1989, several proofs of principles in the early 2000's have

been reported, followed by more complexified studies this last decade. A summary of key information from each of the studies is provided in Table 1, including the disease targeted for detection, the type of body fluid used, the animal detector, and the sensitivity and specificity reported.

Cancer detection has received the most attention, with 2/3 of the studies targeting one or more cancers (6, 31–69). The remaining studies targeted tuberculosis, MRSA (70), Malaria (71), UTI (72), Clostridium difficile (73–77), and Covid 19 (78–85) (Fig. 3). However, since the Covid-19 outbreak in 2020, already eight original articles reporting the ability of dogs to detect Covid 19 have been published, and more will possibly follow.

3.2. Patients' selection

A diagnostic test is designed to accurately discriminate patients from controls. Therefore, the choice of patients and controls is critical. Patients and controls description among reviewed studies is reported in Table 2.

Positive patients selected are subjects diagnosed with the disease of interest before any treatment. Diagnostic is mainly done with a reference test corresponding to the gold standard (histology, imaging, PCR & immunoassays). Histopathologic diagnostic is usually the reference test for cancer. The accuracy of the reference tests is, however, not systematically reported among reviewed studies.

Controls are of several types: (i) Healthy volunteers (healthy = absence of the disease of interest) who do not have and never had the disease; (ii) Healthy volunteers who do not have the disease anymore; (iii) Volunteers diagnosed with other diseases than the one of interest.

The absence of the targeted disease is not the only criteria for controls selection. Some teams often report to match age, gender, skin colour, smoker status, diet, symptoms, and other comorbidities to limit confounders. Several studies, however, included controls with unmatched criteria compared to patients. In some cases, animal performances (sensibility and specificity) were recalculated on a subpopulation of controls matching criteria such as sex and age.

A major drawback is the absence of controls screening in most reviewed studies. This can lead to false negatives samples.

3.3. Samples types and logistics

3.3.1. Body fluids & diversity of samples used

When detection was designed to be done without contact between patients and animals, several types of body fluids have been collected to present odours to the animal detector. Urine is the main body fluid used (n = 20), followed by breath (n = 18), saliva (n = 10), skin secretions (n = 8), cell cultures (n = 7), faeces (n = 6), blood/serum (n = 5), tissue (n = 4), and smear (n = 1). Direct contact with patients or infected areas was conducted in four publications. The total is superior to the number of publications, as some studies report to have used several types of samples. These data are reported in Table 3, as well as in Fig. 4. Three studies performed detection in direct contact between animals and humans (patients and controls).

3.3.2. Sampling materials and protocols

Sampling materials and devices to capture VOCs are reported in Table 3. For urine, blood, and faeces, no material was specially designed to capture VOCs efficiently. Only a receptacle (recipient, jar, cup, vial) was used. For sweat samples, cotton pads are usually used. The composition of these pads is not always well-described. For breath, three types of materials/recipients were used: (i) Either only a container (ex: breath sampling bag (Otsuka Pharmaceutical Company, Tokyo, Japan)(68)); (ii) Either a tube filled with an absorber (ex: cylindrical polypropylene organic vapour sampling tube (Defencetek, Pretoria, South Africa). Each tube is open at either end, is 6 inches long, has an outer diameter of 1 inch, has an inner diameter of 0.75 inches, and has removable end caps. A removable 2-inch-long insert of silicone oil-coated polypropylene “wool” captures volatile organic compounds in exhaled breath as breath passes through the tube (42)); (iii) Other materials (ex: Face mask taken off and placed into a Ziploc® bag (46)). Tissues do not require specific sampling materials. In general, the choices of sampling materials are poorly motivated, and material characterization and properties to capture VOCs are inaccurately described.

Sampling protocols are essential for reproducibility and limit bias. They are reported for each study in Table 3. They are well described primarily for two types of samples: exhaled air and skin secretions. For instance, Thuleau et al 2018 report that all patients and controls must shower with the identical odourless soap before skin secretions sampling (49). As well, some studies add information about fasting requirements prior to sampling, to limit biases.

3.3.3. Sample storage conditions and conservation duration

Samples’ storage conditions are reported in Table 3. Temperature storage conditions are described for 73% of the reviewed articles. We chose to classify them into three categories: (i) room temperature; (ii) cold: 0°C < T < 8°C; (iii) frozen: T < 0°C (Fig. 5). However, the choices of temperature storing conditions are not motivated except in Willis et al 2011: The team primarily stored samples at − 80◦C, which has been the most desirable for retaining chemical volatiles (86).

When stored at room temperature, all studies precise that samples are stored in the absence of light. Even if not specified for cold and frozen conditions, we assumed it was stored in the absence of light in a fridge or a freezer. Such parameter is essential, as light is known to alter VOCs (87). Hygrometry and atmospheric pressure were not described.

Also, sample conservation duration is poorly described among the reviewed articles and varies from a few days to several months. No data has been found about the quantification and the variation of VOCs captured in samples.

3.4. Animal types

The data concerning animal details can be found in Table 4.

3.4.1. Dogs

Dogs are used in 92% of the reviewed studies (n = 57). In total, 226 dogs have participated in studies. Among these 226 dogs, 186 (82%) have completed the whole study process (i.e. have undergone full training and have participated in final testing). A majority of the studies reported that dogs have been trained by professional dog handlers (data not shown).

Dogs seem to be the first choice when it comes to using animals to detect diseases. This choice might come from the extensive use of dogs in drugs and explosives detection, the availability of dog trainers in many countries worldwide, and therefore the accumulated knowledge concerning their education. However, little information is provided about dog selection, except that choices are based on motivation (willing to search, willing to play) and sense of smell. Standard selection tests to evaluate animal capacities are not described.

Some breeds such as German and Belgian shepherds, Labradors and Springers seem to be extensively used (see Table 4). The repartition between males and females is the following: 52% males, 48% females. The number of dogs per study varies between 1 and 10, with an average of 3,96 (SD = 2,84) dogs per study. These numbers are reported in Fig. 5 and Table 4.

3.4.2. Rats

African Giant Pouched Rats have been extensively used for tuberculosis detection in Tanzania (5 studies conducted with the same organization: APOPO). Rats were chosen for their sense of smell, easiness of operant conditioning, and availability in Tanzania. It is reported that such animals can live approximately eight years and they can be trained within a few weeks.

We excluded one study that used mice (88).

3.5. Animal training & testing

All reviewed studies report positive operant conditioning methods for training, the reward being food or a toy. A clicker training method is reported in 45% of the studies. Animal living conditions were, however, not or poorly described. A few teams mentioned dogs housing conditions (for instance, dogs being hosted in families). Education durations vary from a few weeks (Mcculloch et al 2006) to 5 years (Sonoda et al 2011) depending on the teams and difficulty of the exercise. The frequency of training ranged from once a week to two sessions per day, each day. These results are reported in Table 4.

3.5.1. Sample presentation / stations

How individual samples were presented to the dogs is reported in Table 5. A trade-off between odour intensity and contact avoidance between samples and the dogs' nose to limit sample pollution or dog contamination is usually reported to have led stations design. Stations' cleaning is not always described. When reported, no rationale is given. For instance, Horvath et al. (2008) chose to clean both the boxes and the containers with hot water after each exercise. Two years later, the same team (Horvath et al 2010) switched and cleaned with 95% alcohol. No standardized protocol has been identified.

3.5.2. Scent line-up characteristics

Scent line-ups characteristics are reported in Table 5. Scent line-ups are usually composed of 2 to 10 stations, disposed in line or a circle. Most of the studies report a forced-choice design, i.e. a fixed number of positive samples (> 0) per scent line-up. This concept of forced-choice design has been described by Edwards et al 2017 (89). In a forced-choice design, the handler knows that the animal must find a fixed number of samples, which can induce bias. Some studies report the possibility of having zero positive samples per line (called "*blank runs*"), which corresponds to an unforced choice. Another type of unforced choice is to be able to vary the

number of positive samples per line. Unforced choices are less common and often lead to worse performances (see Table 6).

Within scent line-ups, several types of samples can be found: (i) positive samples; (ii) controls (healthy, other diseases); (iii) distractors. Distractors are samples different from positive samples and controls samples. For instance, Murarka et al. (2019) used paper clips, paper towels, cotton balls, screws as distractors (58). They are used to stimulate the animals to search.

3.5.3. Number of samples used for training and testing

The average number of samples used for training and testing is reported in Table 7 and standard deviation, minimum, and maximum numbers. These numbers are not systematically reported among studies, especially for training. Indeed, only 37% of cancer studies and 8% of infectious diseases studies gave information about the number of samples used for training. Information was more exhaustive for testing: 80% of cancer studies and 68% of infectious diseases studies gave the exact number of samples used (). Considering only studies which provided information, the mean number of samples used for training per study was 258 (SD = 560; min = 20; max = 2600), and the mean number for testing was 184 (SD = 186; min = 14; max = 902).

The number of times samples are used is not well reported. For training, some studies report training with only new samples to avoid two biases: 1) the memory effect (i.e. animals don't learn to generalize, but remember each sample), and 2) the "novel object preference" effect (i.e. animals select every sample they never encountered before). For instance, Ehmann et al 2011 report that during the training and later in the testing, every test tube containing a human breath sample was used only once to preclude simple memory recognition of participants' unique odour signatures (69). Even if not always described, when looking at the numbers, it is evident that most of the studies reuse some samples for training.

For testing, most teams used samples (positive and controls) only once per animal during testing. However, a few studies report sample reuses (ex: Cornu et al 2011, cf Table 6). The potential issues brought by reusing samples are discussed below.

3.5.4. Blind conditions

Several studies report to work in blinded or double-blinded conditions. However, these terms do not seem to be used the same way among studies. In this review, we chose to classify the blinded conditions with the following terms:

- Unblinded conditions (UB): the dog handler knows the nature or the position of the sample to evaluate.
- Single blinded conditions (SB): an operator in the room (visible by the dog) knows the nature and the position of the samples to analyze, but the dog handler does not.
- Double-blinded conditions (DB): nobody in the room knows the nature nor the position of the samples to analyze. This can be subdivided as follows:
- Someone outside the room (or at least completely hidden) knows the nature of the sample to analyze (DB1) and can give feedback
- In this configuration, the animals' indication can be evaluated each time, and therefore the handler can:
- Reward his animal (= positive reinforcement)

- Decide to continue or not the evaluations (because he knows if his animal is doing well or not)
- Nobody knows the nature of the sample to analyze, or at least, cannot communicate it to the field (DB2)
- In this configuration, the animals' indication cannot be evaluated each time, and therefore the handler:
- Does not know whether to reward his animal or not
- Does not know when to continue or to stop testing

In 78% of reviewed studies, evaluations were done in double-blinded conditions to avoid the "Clever Hans" bias (90, 91). When well described, DB1 is the major double-blinded subtype reported (42%). Such conditions have limitations (see discussion). In 59% of studies, scent line-up had a forced-choice configuration. These results are reported in Tables 5 & 6.

3.6. Performances

Sensitivity and specificity varied widely, ranging from perfect to chance performance, with considerable variation among studies examining the same disease, sample, and detector. Results are reported in Table 6 and represented in Fig. 7 when both sensitivity and specificity were available.

For "Forced Choices" situations, we reported the specificity numbers from the original articles. However, these scores can be considered as not relevant (see discussion). Configurations using an unforced choice line-up in double-blind type DB2 correspond to a true "screening like situation". Only six studies were found with the latter design: three with rats detecting tuberculosis, two with dogs detecting cancer, one with dogs detecting C. difficile infections.

4. Discussion

4.1. General comments

Since 2004, many proofs of concept have been published about the ability of dogs or rats to detect diseases. However, there are often great discrepancies among results. While some publications report low sensitivities of 17% (Gordon et al., 2008) and specificities around 29% (Amundsen et al., 2014), others achieve rates of 100% sensitivity (Horvath et al., 2008; Cornu, et al., 2011; Sonoda, et al., 2011) and specificity (Sonoda, et al., 2011; Yamamoto, et al., 2020).

Only a few studies reported testing performed in screening conditions (DB2, unforced choice), and those usually enrolled small numbers of animals. This could be explained by the fact that screening conditions in double-blind testing combined with unforced choices are more challenging for the animals, the handlers, and the operators, limiting the amount of data required to validate such a method. These results are discussed in the following paragraphs.

4.2. Considerations about patients' selection and samples

4.2.1. Patient and control selection: reference test and populations matching

First, careful diagnosis of patients and controls is critical to avoid bias. Making sure that patients have the disease of interest is usually confirmed with the gold standard. The accuracy of such a test must be high to avoid false-positive inclusions. Also, confirmed negative samples are critical, and all controls should be tested in an ideal situation. However, very few studies report having rigorously tested controls. This can be explained by the fact that asking volunteers to perform non-required detection tests is costly, time-consuming, tedious and invasive, poses ethics issues, and could lead to volunteer disengagement. However, from a scientific point of view, non-tested volunteers could be a source of false-negative samples. Such samples would be detrimental for animal training and testing. Indeed, the animal must be educated with samples with known status. An inaccurate reference test might lead to sample status errors and mislead the detector. For instance, Thuleau et al. (2018) reported they educated dogs to detect breast cancer from patients with cancer confirmed by histology and from volunteers with a recent (< 12 months) negative mammography. Even if mammography is reliable, false negatives can occur, or cancer can appear within a few months following the screening. Dogs are trained to ignore such samples, which can lead to other mistakes.

If the reference test has poor accuracy, then animal training can be impacted. For instance, the results reported by Cornu et al 2011 show that training a dog with potential “rogue” controls affected final performances (66). Selected controls were patients aged > 50 with elevated Prostate-Specific Antigen (PSA, comparable with cancer patients regarding these characteristics). Control patients had mean PSA value of 8.3 +/- 4.1 [range: 2–16.8]. Given these values, it can be considered that 20–30% of these control patients with negative prostate biopsies had prostate cancer.

Similarly, Willis et al 2004 reported they were concerned that “rogue” control specimens from people with undiagnosed cancer elsewhere in the body might be inadvertently added to pooled samples. They did have an occasion during training in which all dogs unequivocally indicated as positive a sample from a participant recruited as a control based on negative cystoscopy and ultrasonography. After further tests, a transitional cell carcinoma was discovered. As such detection method with animals is not yet validated, not all false positives indicated by animals can be double-checked. More recently, Grandjean et al. 2020 had a similar issue, with two of their supposed SARS-CoV-2 negative controls turned out to be positive.

Second, the importance of matching the characteristics of patients and controls groups to make sure that animals detect the disease itself and not a confounding factor is known (89, 92). Matching has been reported with age, sex, skin colour, other diseases, comorbidities, symptoms, smoker status, diet.

For instance, Bomers et al 2012 worked on C. difficile detection with dogs at a hospital. They reported that on the day of the detection round all cases had diarrhoea compared with 6% of the controls. In such a situation, we can wonder if the dog successfully indicated the targeted disease (C. difficile), or just the presence of diarrhoea.

To prevent such bias, Willis et al 2004 exposed the dogs to urine from patients presenting with a broad range of transitional cell carcinomas, in terms of grade and stage, to increase their likelihood of recognizing the common factor or factors. They took particular care to train the dogs with control samples containing elements likely to be present in urine from patients with bladder cancer and commonly occurring in other non-malignant pathologies. This way, they could teach the dogs to ignore non-cancer specific odours. This led to the inclusion of urine samples from a variety of patients, such as people with diabetes to control for glucose, those with

chronic cystitis to deal with the influence of leucocytes and protein, and healthy menstruating women to control for blood.

Several years later, the same team (Willis et al 2011) assumed that body fluids, tissues and emissions from young, healthy individuals differ in composition from those of older cancer patients to a greater extent than do samples from age-matched individuals with the non-cancerous disease of the same organ. They performed an electronic nose study in which the classification accuracy dropped once more diseased individuals were added to the healthy control group (93). This shows that the choice of controls can markedly affect the level of specificity achieved.

4.2.2. Disease-specific odour and types of body fluids chosen

Research teams made the hypothesis that a specific odour was present in the samples they chose. However, so far and to our knowledge, in the case of cancer, it is not known whether a specific cancer has a specific chemical signature or not, and, if so, what is the source of such signature. Indeed, the odour of cancer could come from the tumour itself, or the modified environment surrounding the tumour, or both. Moreover, it is still not known yet whether all cancers have shared odours or not. For instance, McCulloch group reported good dogs' performances trained to alert to two cancers rather than for single cancer discrimination. This could mean that there is a general biochemical marker common to all cancers, with individual-specific cancers having additional markers (53).

There are different interpretations considering the localization of disease odour within the body: is it localized, organ-specific or spread? For instance, Horvath et al. (2008) report that one important observation during the training period was that use of fat from the same individuals from whom the carcinomas were removed did not increase the number of failures. The absence of reaction by the dog suggests that a general body odour including all organs did not exist. However, two years later, the same team (Horvath et al 2010) reported that for the same cancer (ovarian), dogs trained with tumours could discriminate blood samples and vice versa. Their study strongly suggests that the characteristic odour emitted by ovarian cancer samples is also present in the blood (plasma). Similarly, after observing that canine scent judgement can be used on both breath samples and watery stool samples, Sonoda et al 2011 concluded that chemical compounds may be circulating throughout the body for colorectal cancer.

Murarka et al. (2019) comment that Yoel et al. (2015) found that after being trained on the breast cancer cell line, the dogs were able to detect both skin cancer and lung cancer cell lines, suggesting the possible presence of a general cancer olfactory cue within cancer cell lines. However, this study did not explore whether these dogs could also then detect cancer in patient-derived samples. In this case, there is also the possibility that the dog learnt to disregard control samples (which were probably similar) instead of recognizing malignant cell cultures. This seems in according to observations from Murarka et al 2019, whose research suggests that after training on cell lines to prepare the dogs, there was no spontaneous switch to blood plasma.

From these observations, four situations can be considered depending on odour specificity and localization, which are presented in Table 8.

Table 8
Disease odours localization and specificity hypothesis

	Disease odour localized	Disease odour widespread
Disease odour: specific	Sample choice critical High test specificity	Sample choice is less critical High test specificity
Disease odour: common to several diseases	Sample choice critical Localization can give alert to a shortlist of diseases	Sample choice is less critical Low test specificity

Table 8 shows that body fluid choice is critical. This also affects control choice. From this table, we see that an odour widespread throughout the body and non-specific to a disease will lead to low specificity tests. In such situations, indications of a sample by a trained animal will not give much information on what disease to look for, and therefore will have low added value.

Body fluids used in the reviewed articles are dominated by breath and urine (Fig. 4). These have the advantage of being easy to sample (liquid, air, noninvasive), easy to split into several samples, and therefore allow several trainings and tests per sample without encountering pollution or odour decrease. Liquids like urine are also easy to dilute, for instance, to increase detection difficulty by reducing the amount of VOCs per sample. These dilutions also allowed to study animal detection thresholds (88), and comparisons with GC-MS and e-noses. However, we regret that the reasons that lead to the choices of body fluids were not or poorly documented.

4.2.3. Sampling protocols

After body fluid and sampling localization choice, sampling protocols and materials are key to have high-quality samples. Most of the studies report the importance of applying the same sampling procedures both for patients and controls to eliminate potential bias and confounders. For instance, Ehmann et al 2012 showed that, at first, trained dogs were not discriminating disease state, but sampling location which was different for patients (at the hospital) vs healthy volunteers (at home).

If the sampling protocol is made at home with no supervision, risk of error can occur, leading to poor samples quality. Thuleau et al. (2018) report that to sample skin secretion they asked patients and volunteers to shower with an odourless soap, before sleeping with a cotton pad on the breast overnight. In this case, researchers cannot be sure that the person has followed each step correctly or that no incident occurred. In this example, the pad could have felt during the night, resulting in pollution and a limited contact time of the pad with the skin, and therefore in a limited amount of VOCs. As well, other odours could have been impregnated on the sample, such as bedsheets' odours, partners' odour, pets' odours. Such unsupervised sampling protocols add difficulties and should be controlled as much as possible.

A non-exhaustive list of parameters that can induce bias are smokers status, sex, age, ethnicity, diet, different sampling locations, different sampling protocols for patients and controls, treatments. For instance, to limit diet

bias, Hackner et al. (2016) report that for homogeneous sampling, the tested persons were constrained not to drink, eat and smoke within 90 minutes before breath sample collection.

4.2.4. Odour sampling materials

All types of body fluids do not necessarily require odour sampling materials. For instance, urine, faeces, and blood can be sampled and presented untransformed to animal detectors. However, breath and skin secretions need optimized materials to capture VOCs without releasing other odours that could disturb detection. Some sampling materials have been presented in Sect. 3.5.2 and Table 3. For instance, Willis et al 2016 report that their choice of material comprising their patches came from studies on canine scenting in forensic science. In terms of the greatest variety and quantity of skin surface VOCs collected and readily released, the optimum fibre appeared at the outset of their study to be 100% cotton, so they employed a widely available, sterile, pure cotton gauze throughout. For the chosen sampling time of 15 minutes, they were again guided by the forensic science literature.

However, such description is an exception, and as for the choice of body fluids, we can regret that the choice of materials is little documented. The vast discrepancies among material types strongly suggest this part of research is still empirical and needs better understanding, characterization, and standardization. In the future, this field of research would benefit from a better description of material parameters, as it is often done in publications reporting VOCs detection by GC-MS.

4.2.5. Sample conservation

In chemistry, it is known that temperature variations, light and hygrometry can modify VOC profiles. Such parameters are crucial but not well described and yet not consensual.

Most of the reviewed studies stored samples at low temperatures (< 0°C), and only a few stored them at room temperature (see Table 3 and Fig. 5). This choice is usually not motivated, except in a few studies. Willis et al. (2011) report that samples were stored primarily at − 80◦C, which has been the most desirable for retaining chemical species (86). Mahoney et al 2012 report that their samples were frozen at − 20◦C until the evaluation day (up to seven days). Though there is some controversy surrounding the cellular impact of freezing and thawing sputum, past research suggests that samples may be kept frozen without significant alteration of cell quality or cell counts (94). Not much information is given about light. However, most studies report storing samples in a fridge or in a freezer, where an absence of light is evident. No information has been found about hygrometry or pressure. Conservation time and the number of sample openings lack description. The heterogeneity about VOCs conservation procedures shows this part is still empirical and needs better understanding and evaluation. Guidelines about minimal, maximal, and optimal conservation conditions would undoubtedly be helpful for standardization.

4.2.6. Considerations about odour threshold

Selected animals have a superior sense of smell compared to humans (15). For instance, Horvath et al. (2010) observed that trained dogs could detect a quantity of 20 ovarian carcinoma cells on the abdominal fat. However, the sense of smell is not unlimited, and it loses efficiency below a certain VOCs threshold. This threshold effect has been studied in Sato et al 2017 (article excluded from this systematic review). Willis et al 2004 also report that they had to consider the physical state of the urine when presented to the dog. They opted

to train one cohort of dogs on wet samples and another on dry samples. When tested, the dogs trained on liquid urine performed significantly better, suggesting that the more volatile molecules are important in the cancer odour signature.

Odour threshold also plays a role in dog training progression. Some teams chose to directly use the same types of samples at training start and for testing. On the contrary, others started detection work with samples with a higher amount of VOCs and decreased the intensity step by step. The latter strategy is supposed to be easier for the animals prior to lowering the threshold. These samples with more VOCs can be (i) bigger (bigger in volume, surface, quantity); (ii) more concentrated (exhaled air, sweat, etc); (iii) other types of samples, such as tumours or materials directly in contact with the tumour. However, the diversity of samples used before the final configuration is not systematically reported within studies.

In addition, there may have differences in odour intensity between diseases, especially infectious and viral diseases with strong diffusion (to be related to contagion) vs. hidden tumours. Hence the importance of odour sampling procedures and materials, as well as sample conservation.

4.2.7. Sample number of uses: pollution and memory effect

The number of times samples are used is not always well reported. It is evident, however, that some studies reused samples at least for some training. Here, two types of “reuses” are to consider:

- Case 1: The same sample is presented several times to the same animal detector
- Case 2: The same sample is presented to several dogs (several times per dog or not)
- Case 3: Sample replicates of the same patient are presented to an animal detector

In cases 1 and 2, there is a risk of pollution (by direct contact with the animal or by its breath, by the atmosphere), which lead to sample alteration each time the sample is used. Therefore, once smelled, samples are not identical to “new” samples. Moreover, opening a sample several times can lead to a decrease in VOCs quantity. In cases 1 and 3, samples from the same person are presented several times to an animal. By doing so, there is a risk of training animal’s memory instead of discrimination. This latter issue has been reported by several teams who saw their results plummet in double-blind situations with only new samples.

On the contrary, however, Willis et al 2016 report that multiple uses of the same sample during training did not appear to lead to a significant loss of volatile signature since the dog continued to successfully select known melanoma samples used up to 15 times over a period of 18 months post-collection. With such observation, one can assume that the dog did not learn to discriminate samples but instead memorized one specific sample.

Ideally, an animal should smell only new (uncontaminated) samples, only once per patient (to avoid memory effect). The advantage of urine, faeces, blood and breath is that these body fluids are easy to sample or to aliquote, allowing to have several samples very quickly. This way, several dogs can be trained with samples from the same person, while preserving their quality.

In some studies (ex: Cornu et al 2011;), some control samples were reused during testing. This does not seem to be a problem in an unforced choice configuration (cf scent line-ups characteristics, part 3.7). However, in a

forced-choice configuration, reusing some control samples might reduce the number of new possibilities for the dogs, leading to an easier design and higher success rate just by chance.

4.3. Animals

4.3.1. Animals

Except for dogs, giant pouched rats have been extensively used by one team working on tuberculosis detection in Tanzania. Little literature report reasons considering animal choice except for their high sense of smell. Dogs are the most used animals worldwide. This choice can be justified by the availability and experience of dog trainers in many countries, for instance, for drugs and explosives detection. Dogs have the advantage of being adaptable to different fields (battle, airports, rescue, remote scent tracing, contact with humans). However, for remote disease detection only (detection done in a controlled configuration, at a distance from patients), there is no need for such adaptation, and to our knowledge, no validated study to prefer dogs than rats. Authors generally report looking for motivated dogs with high olfaction capabilities. However, there seems to be no standard validated tests for dog selection, which so far remains empirical in the absence of clear guidelines.

Gordon et al 2008 mention that it has been an ongoing theory that certain breeds are better at scent detection than others (95). However, studies have shown a greater difference in scenting ability between dogs within a breed than between breeds (96). We observe performances variations in selected studies between breeds and within the same breeds. This has been described in Jamieson et al 2017, who concluded that a dog should not be solely chosen based on its breed (95) due to individual variation. In addition, if we consider that evaluated dogs were for the majority selected among the best, under the watchful eyes of an experienced professional, we can assume that even more discrepancies would exist without such selection. There are an estimated 500 million dogs worldwide and, so far, less than 200 have been considered potentially adapted to conduct disease screening tasks in controlled studies and achieved varying results. Such method seems to have huge potential; however these low numbers preclude extrapolation.

4.3.2. Selection success

In Elliker et al 2014, only three out of ten dogs initially recruited for the study passed the first stage of training. According to them, high failure rates are common when training dogs for specialist roles because of the specific behaviour/temperament attributes required (97, 98).

Despite this low selection rate, 82% of the dogs mentioned in the studies completed all the exercises requested. This number may seem high but hide several parameters:

- It is likely that some studies only mention the dogs who performed well and do not mention all the dogs they evaluated before selecting their champions.
- Some of the dogs, even after completion of the whole program, have poor results.
- The loss rate is greater when the difficulty of the exercise increases (blank runs, double-blind). As most of the studies report forced choices scent line-ups, more dogs succeed.

Interestingly, Murarka et al 2019 report that all dogs leaving the disease detection program and switched to other odours (for example, narcotics, bed bugs, accelerants, blood plasma) have been rapidly and successfully

trained. This strongly illustrates the difficulty of disease detection with dogs compared to other odours.

Elliker et al 2014 report that it has been suggested that it may be useful to breed dogs specifically for cancer odour detection (99), which may help to increase the proportion of suitable dogs available for future studies of this type.

4.3.3. Training duration

Considerable differences in training durations are observed within studies, going from a few weeks to several years. Such differences can be explained by the type of disease to detect, the difference between patients and controls, the choice of body fluids, the quality of samples, training differences, animal abilities. No correlation was observed between training duration and success rates among studies. However, Ehmann et al 2012 identified an improvement of lung cancer identification capabilities along with the test series and conclude that an ongoing training effect must be assumed, calling for even more extended dog training in future studies.

4.4. Scent line-up

4.4.1. Scent line-up: Number of samples and line vs circle, the distance between samples

The number of samples presented to animals ranges from 2 (Bomer et al 2012) to 12 (Essler et al 2021). No justification was provided considering these numbers. No study was performed with only one sample. It has been shown in the literature that dogs were able to perform tests with one sample only (100). In such a test, dogs have to make an absolute choice. They are asked to “evaluate”. On the contrary, when several samples are presented, the dog can perform a discrimination task and is probably more stimulated. In this situation, they are asked to “search”. All studies reviewed used the latter configuration.

Samples were presented in the line, circle, or randomly (Table 5). The choice of a line can be motivated by the easiness of designing “blank” runs, such that, at the end of the line, the dog can indicate that no positive sample was found. Blank runs can also be done in a circle configuration. The advantage with the latter is that there is no start nor end, so all samples are equivalent.

Space between samples is fundamental for several reasons. The most obvious reason is to preclude cross-contamination between samples. Another less apparent reason is that it gives enough time for latency and persistent olfaction times. Latency is defined as the necessary time to get an olfactive stimulus, estimated at 0,5 seconds for dogs. Persistence time is the duration the olfactive sensation stays. If samples are too closed, these durations cannot be respected, and dogs risk either missing a sample or mixing signals.

4.4.2. Scent line-up configurations: forced vs unforced choice

Using forced vs unforced choices scent line-ups have a strong influence on performances. Unforced choice exercises are more complex. In a forced-choice exercise, the animal learns only one configuration. They know they must “find” the odour of the disease. As a result, they chose the sample which resembles the most to the target or the one which is the odd one out. Moreover, Bomers et al 2012 report that anticipation of a single positive result could have influenced the trainer’s behaviour, thereby unintentionally influencing the dog’s response (90). Such configuration is therefore not only easier for the animal but also for the handler. On the

contrary, animals must evaluate each sample in an unforced choice configuration and cannot choose only by simple comparison. This is a difficulty that not all animals can overcome. An unforced choice situation is, however, the only one that could be applied for screening.

With the particular configuration reported by Murarka et al 2019 (see results), the dog has only one sample to evaluate, while the distractor is here for stimulation (58). Such configuration is an interesting tradeoff between one vs several samples scent line-ups described above and can easily be applied for screening.

4.4.3. Atmospheric conditions

Atmospheric conditions during training and testing are known to affect dogs sense of smell. These conditions are poorly documented within the reviewed studies. Those who did, however, reported working with controlled temperatures between 12°C and 20°C (see Table 5). It can also be seen that, when not under control, this can negatively impact scent detection work, like for instance reported by Sonoda et al 2011 where tests were conducted from 13 November 2008 to 15 June 2009 because the dog's concentration tended to decrease during the hot summer season (68). As well, Hackner et al 2016 observed that some limiting influences included high humidity and elevated ambient temperature, which were found to be detrimental to the dogs' performance. They suggest that testing should not be performed during unfavourable weather conditions (43).

4.4.4. Blind conditions

4.4.4.1. Proofs of principle vs double blind clinical trials in a screening like situation:

For a potential deployment of disease detection with animals, only double-blind clinical trials in screening-like situation (i.e. unforced choice) might be useful (see Sect. 3.9 for blinded conditions). Up to date, only 6 studies meet expectations (Table 6). Focusing on these studies, the results usually decreased at first when shifting to double-blind. This fall between training and double-blind testing has often been explained by the Clever Hans effect (90). To avoid failure, teams must train as much as they can in blind situations, as suggested by Gordon et al 2008 who report that the use of blinding during the training should be initiated early to preclude unintended clues by the trainers that may contaminate the process (53). Willis et al 2016 reported that after training the dog in a non-blinded situation, their trainer reported back a near 100% success rate in identifying the melanomas. It was decided to begin a series of double-blind tests. However, after 13 runs, the dog had successfully identified only one of the melanoma samples (44). Implementing blinded conditions is not easy during training because dog handlers need to know when to reinforce positive behaviour. To do so, a non-blinded assistant hidden from the dog and who can quickly tell the handler when to reinforce is needed.

4.4.4.2. Rewarding or not the dogs in screening-like situation: a puzzling question

In a screening like situation, nobody knows whether the animal's indication is correct or not, which can be an issue for the reward. Indeed, if the trainer decides not to reward the animal, the latter can little by little lose interest. On the contrary, if the animal is rewarded every time, this might reinforce biases in case of incorrect indications. Therefore, several strategies are adopted among teams.

For instance, McCulloch et al 2006 report that, since the experimenters no longer knew the status of the target breath sample, they did not activate the clicker device after a sitting indication by the dog, and therefore the handler did not reward the dog with any food. Bomers et al 2012, in the case of C. difficile infections, search in hospital wards, confirms that surveillance is principally different from the type of case directed diagnosis in their study design because the dog cannot immediately receive a reward after a positive identification, potentially extinguishing the trained alert. The same solution was adopted by Willis et al 2011: “Both the trainers and researchers remained blinded throughout the trial, only breaking the sample and positional codes at the very end, meaning that the dogs could not be rewarded for a correct indication immediately after each test run. The trainers reported that, over time, this led to a loss of confidence in the dogs, with a deterioration in their performance”. On the contrary, Elliker et al 2014 performed two types of tests. On the first one, they were in a DB2 situation and decided to reward the dog for each indication. However, during three rigorously controlled double-blind tests involving urine samples from new donors, the dogs did not indicate cancer samples more frequently than expected by chance. The team finally switched to a DB1 situation, to be able to reward the dogs only for positive responses. These are exceptions because most of the studies were conducted in DB1 configuration, which allowed to know whether to reward the dog or not after each line.

According to Biehl et al 2019, rewarding dogs’ work has to be independent of the results achieved and should refer only to the work done. If dogs are only rewarded for positive indications, they will quickly learn to achieve more rewards through positive indications, which could easily lead to higher false-positive results. Hackner et al 2016 attributed the inferior results to the true double-blind and screening-like conditions. They report that this factor posed immense stress on the dogs and their handlers, and therefore suggest positive feedback mechanisms for future study designs. According to them, it seems to be favourable to confront dogs relatively often with the pattern odours. Their results report that a test situation where dogs will always find an unblinded positive and ignore an unblinded negative sample in the line-up would probably be better. The positive sample would create the opportunity to earn a reward and would reinforce the dogs’ motivation. The negative sample assures the handler that the dog is still performing well. The other samples in the line-up should be the blinded test samples.

Another similar solution would be to alternate training lines and test lines. It could be decided that one test line has to be performed only after an amount (to determine) of successful training lines. Another training line could be performed right after the test to ensure the dog is still doing well. Such a pattern is feasible for implementation; however, it would slow the testing throughput.

This subject is crucial for implementing such a method, and no consensus nor solution has been admitted so far.

4.5. Applications / Implementation

Pickel et al 2004 published a proof of concept with dogs sniffing humans. Even if scientifically feasible, such a technique seems hardly applicable in the field. Since then, several studies using remote disease detection have step by step built a new scientific discipline. This review shows that no scientific study has validated that animals can be used as a first-line remote detection tool prior to existing technologies. Only APOPO, the organization supervising Giant Pouched Rats detecting tuberculosis in Tanzania, has found its place as a second line screener, which makes sense for tuberculosis detection (101, 102).

4.5.1. How many evaluations are needed to validate a sample?

Most studies focus on the performances of each animal separately. However, as animals are living organisms, their performances can be subject to variations. Biehl et al 2019 reported that literature data show that some dog trainers included only one dog in scent detection, whereas others had five to six dogs and collected the individual dogs' data (54). McCulloch et al. 2006 state that the sniffing quality of all dogs was comparable, and therefore the results obtained were similar. However, Ehmann et al. 2012 found differences in hit rates between individual dogs and consequently defined a 'corporate dog decision' that required at least 3 out of 5 dogs with an identical decision. Amundsen et al., as well as Hackner et al., also showed considerable variations in single dogs' results. These variations might be due to the dogs' different sniffing capabilities and the dogs' different daily conditions and training.

Biehl et al 2019 report that in their study, single dogs' results showed great differences concerning sensitivity in the range of 22–67% and concerning the specificity of 71–89%. They conclude that it is advisable not to rely on a single dog's decision but to define a corporate decision to minimize variations arising from the single dogs. This choice is not straightforward. Indeed, Mahoney et al 2012 report that considering the use of two animals only, if they consider that a sample is positive if indicated by both animals, test sensitivity will decline, but specificity will rise. On the contrary, if only the indication of one of the two dogs is needed, the sensitivity will increase, but specificity will fall. The argument can be declined for more animals and indications. For instance, Gordon et al 2008 report that, at the time, their study was the only one to incorporate replicates for assessing specificity. There were 3 and 2 replicates (33 and 18 runs) for the prostate and breast arms, respectively. The team adds that any study, ultimately attempting to prove canine superiority over conventional cancer screening, must include replicates and, in the future, go head to head with standard screening methods. Another example is Mgode et al 2012, where for tuberculosis detection, a sample is considered positive if selected by two rats. Such a corporate decision is a tradeoff that has not found a consensus yet.

4.5.2. Number of samples to train a dog and maintain performances

The number of samples available for training is crucial. Indeed, many samples are needed so that animals learn to generalize and do not memorize each sample. Quantity is essential to work as often as possible with new (non-polluted) samples and limit the "novel object preference". Willis et al 2011 report their protocol also avoids the phenomenon of novel-object preference, whereby dogs preferentially chose unfamiliar items over familiar ones (103).

This is not straightforward, as organizing efficient logistics to gather samples continuously can be challenging to implement. For instance, Gordon et al 2008 report that it took longer than anticipated to obtain enough samples to prepare for the final testing. This resulted in the training being spread over an extended period, 12–14 months. Possibly, the animals were periodically memorizing individual patients rather than recognizing an "odour signature" for cancer despite utilizing a large number of training samples. An ongoing system of recruitment of patients with cancer and control patients needs to be established, so the dogs have adequate numbers of new samples to maintain their proficiency even after the conclusion of the study. This has also been reported by Ehmann et al 2011, who wrote that during the training and also later in the testing, every test

tube containing a human breath sample was used only once to preclude simple memory recognition of participants' unique odour signatures.

This need for a continuous arrival of new samples is a huge limitation. Indeed, if intended to be implemented in countries with low access to diagnostics, this arrival of new samples from screened patients and controls will be limited. This implies continuous logistics and partnerships with hospitals that might not be cost-effective.

4.5.3. Field implementation

If scientifically validated, remote scent medical detection implementation will have to overcome several issues. First, if implemented in populations with low health access, such detection will have sense only if care can follow. We saw that many known samples, both from patients and controls, are required to train animals. If implemented in an area with low access to gold standard detection, sample recruitment might be compromised.

Routine adoption of such detection raises the question of the number of samples which can be screened every day and its cost. From the studies reviewed, it seems that one dog, if efficient, could screen roughly a dozen of new samples per day. Willis et al 2016 report that only one new test was conducted per week, with training sessions in between, which is not very efficient for mass screening. Rats, however, seem to be able to screen more samples, as reported by Weetjens et al 2009: "The use of trained rats to detect tuberculosis is reliable, potentially cheaper and faster than sputum smear microscopy. One evaluation cage can contain more than 12 rats per day, and one rat can screen 140 samples in 40 minutes. The evaluation set-up can therefore process up to 1,680 samples per day, while a microscopist can process up to only a maximum of 40 samples per day (WHO recommends an average of 20 samples per day) (104, 105)".

Another important consideration is the prevalence of the disease to be detected. Indeed, if very few positive samples are present, this could lower animals' motivation and accuracy. Hence the importance of training sessions with regular new known positive samples.

As discussed in Sect. 4.2.2, such detection will be helpful if the odour and/or the sampling localization is specific to a shortlist of diseases. If not, then in the case of an alert, medical staff will not know what to look for.

Free running rapid detection might be useful for infectious diseases. Free running proofs of concept have been published for C. difficile infections detection with encouraging results (74). However, such detection has not been proven yet to work in the field for other diseases. So far, published articles report successful proofs of concept in remote conditions (like for cancer). Free running detection has recently been presented as an objective by several teams working on SARS-COV-2 detection. For instance, Guest et al 2021 report that their preparatory work indicates that two dogs could screen 300 people in 30 min, for example, the time it takes to disembark from a plane, and PCR would only need to be used to test those individuals identified as positive by the dogs. However, no study has demonstrated such application in real screening conditions in contact with people in public places so far. On a different disease, Maureen et al 2018 report that despite being highly trained, dogs are vulnerable to distractions and other foreign stimuli in a unique social environment (106). Concerning their study, Essler et al 2021 report that though dogs have previously been shown to be able to discriminate between saliva samples of SARS-CoV-2 positive and negative patients, these studies are also

using repeated presentations of the same samples. Thus, it is possible dogs can discriminate between their training set of positive and negative patient samples but are unable to generalize this odour to new samples. These considerations are major limitations that preclude short term implementation. However, this relatively new field of research is progressing quickly, and future studies may address and outcome these issues.

Finally, disease diagnostic can be expensive and complex to implement because of costly infrastructure and instruments, the need for consumables and high-skilled professionals (MSc, PhD, MD). In this context, several teams claim that medical remote scent detection with animals might be cheap, however, this has yet to be proven. Cornu et al 2011 report that in their proof-of-principle study, they tested a limited number of subjects in a costly, long study that makes it difficult to conceive of extended use for this test in clinical practice (66). Similarly, Sonoda et al 2011 declare “it may be difficult to introduce canine scent judgement into clinical practice owing to the expense and time required for the dog trainer and dog education” (68). No socio-economic study on the subject was found.

5. Conclusions

According to Hackner et al 2016, a suitable screening method should provide a true negative rate of near to 100% to be sufficient for safe use. Despite the number of studies reporting the potential capacity of trained animals to be used as diseases detectors in a clinical setting, no validation has been issued so far. Willis et al 2016 alert that introducing canine diagnosis of cancer in the absence of adequate validation, and without external quality control assurance mechanisms in place, may raise some of the same patient safety issues like those highlighted by the British Medical Association in their 2005 report on unregulated screening tests (107).

Interestingly, several teams do not recommend the use of such a technique routinely. For instance, Horvath et al 2008 wrote that they do not believe that dogs may be used in clinical practice. Dogs as “living instruments” may be influenced by several factors before and during their work, leading to changes in the accuracy rates. However, under controlled circumstances, they may be used in experiments to further explore this exciting new property of malignancies. In Willis et al 2011, researchers do not advocate the use of dogs in a clinical setting. They hope that a greater understanding of the VOC biomarkers associated with bladder cancer, and urological disease more widely, will help optimize the design of an electronic nose. This has been suggested by Taverna et al 2016, that dogs could be used to explore the response to cancer treatment or relapses in conjunction with VOCs identification.

The implementation of dogs to detect infections in a free-running setting (in contact with humans) has still to prove efficiency. For instance, Bomers et al 2012 report that a limitation of using an animal as a diagnostic tool is that behaviour is not fully predictable. The dog’s reaction to other stimuli (for example, children’s play, being beckoned, being offered a treat) illustrates that dogs are still prone to distraction despite a high level of training.

However, this research field has made considerable progress since 2004, research teams, programs and networks are constituted, and the main scientific obstacles seem to have been identified. By carrying out studies on materials, VOCs conservation conditions, and by better mastering the selection and variability of dogs, a rigorous process will undoubtedly lead to possible implementation. Medico economic studies still need to be conducted.

Finally, the work done in chemistry on the olfactory signature of diseases is complementary and will probably help to understand better and standardize research conducted with animals. Subsequent progress on this subject should determine more clearly what will be possible to implement in the future.

Abbreviations

DB: Double-blinded conditions

DB1: Double-blinded sub-conditions 1

DB2: Double-blinded sub-conditions 2

FC: Forced choice

GC-MS: Gas chromatography-mass spectrometry

MRSA: Methicillin-resistant Staphylococcus aureus

N/A: Not applicable

PSA: Prostate-Specific Antigen

SARS-COV-2: Severe acute respiratory syndrome coronavirus 2

SB: Single blinded conditions

SD: Standard deviation

VOC: Volatile organic compound

UB: Unblinded conditions

UFC: Unforced choice

UTI: Urinary tract infections

Declarations

Ethics approval and consent to participate:

Not applicable.

Consent for publication:

Not applicable.

Availability of data and materials:

All data generated or analysed during this study are included in this published article.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' contributions:

P.B.: Protocol writing, title and abstracts screening, data extraction, manuscript writing, discussion, and submission. M.L.: title and abstracts screening, data extraction, discussion, and full-text reading. E.A.: Discussion and full-text reading. I.F.: Research idea, data extraction, discussion, and full-text reading. All authors read and approved the final manuscript.

Acknowledgements:

Not applicable

References

1. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Znaor A, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. 2018;
2. 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 [Internet]. 2016;3(June):1–7. Available from: http://journal.frontiersin.org/Article/10.3389/fvets.2016.00047/abstract
3. Angle TC, Passler T, Waggoner PL, Fischer TD, Rogers B, Galik PK, et al. Real-Time Detection of a Virus Using Detection Dogs. Front Vet Sci [Internet]. 2016 Jan 8;2. Available from: www.frontiersin.org
4. Yu C, Fan S, Sun Y, Pickwell-Macpherson E. The potential of terahertz imaging for cancer diagnosis: A review of investigations to date. Quant Imaging Med Surg [Internet]. 2012;2(1):33–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23256057%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3496499
5. Shreffler J, Huecker MR. Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios [Internet]. StatPearls. StatPearls Publishing; 2020 [cited 2021 Jun 14]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/32491423
6. 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 [Internet]. 2012 Jul 3;404(1):141–6. Available from: http://link.springer.com/10.1007/s00216-012-6102-8

7. Jain RB. Levels of selected urinary metabolites of volatile organic compounds among children aged 6-11 years. Environ Res [Internet]. 2015 Oct 1 [cited 2021 Jun 4];142:461–70. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0013935115300438
8. Blanchet L, Smolinska A, Baranska A, Tigchelaar E, Swertz M, Zhernakova A, et al. Factors that influence the volatile organic compound content in human breath. J Breath Res [Internet]. 2017 Feb 22 [cited 2018 Oct 15];11(1):016013. Available from: http://stacks.iop.org/1752-7163/11/i=1/a=016013?key=crossref.bd4e80d433fa94f5085fabd0e4e13cdf
9. Abd El Qader A, Lieberman D, Shemer Avni Y, Svobodin N, Lazarovitch T, Sagi O, et al. Volatile organic compounds generated by cultures of bacteria and viruses associated with respiratory infections. Biomed Chromatogr [Internet]. 2015 Dec [cited 2019 Jun 2];29(12):1783–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26033043
10. 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.
11. LIBARDONI M. Analysis of human breath samples with a multi-bed sorption trap and comprehensive two-dimensional gas chromatography (GC×GC). J Chromatogr B [Internet]. 2006 Sep 14;842(1):13–21. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1570023206003825
12. Miekisch W, Schubert JK, Noeldge-Schomburg GF. Diagnostic potential of breath analysis—focus on volatile organic compounds. Clin Chim Acta [Internet]. 2004 Sep;347(1–2):25–39. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0009898104002256
13. Nakhleh MK, Amal H, Jeries R, Broza YY, Aboud M, Gharra A, et al. Diagnosis and Classification of 17 Diseases from 1404 Subjects via Pattern Analysis of Exhaled Molecules. ACS Nano [Internet]. 2017 [cited 2018 Oct 22];11:42. Available from: www.acsnano.org
14. Sethi S, Nanda R, Chakraborty T. Clinical Application of Volatile Organic Compound Analysis for Detecting Infectious Diseases. Clin Microbiol Rev [Internet]. 2013 Jul;26(3):462–75. Available from: https://journals.asm.org/doi/10.1128/CMR.00020-13
15. Niimura Y, Matsui A, Touhara K. Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. Genome Res [Internet]. 2014 Sep;24(9):1485–96. Available from: http://genome.cshlp.org/lookup/doi/10.1101/gr.169532.113
16. Lesniak A, Walczak M, Jezierski T, Sacharczuk M, Gawkowski M, Jaszczak K. Canine Olfactory Receptor Gene Polymorphism and Its Relation to Odor Detection Performance by Sniffer Dogs. J Hered [Internet]. 2008 May 8;99(5):518–27. Available from: https://academic.oup.com/jhered/article-lookup/doi/10.1093/jhered/esn057
17. Browne C, Stafford K, Fordham R. The use of scent-detection dogs. Ir Vet J. 2006;Volume 59(February).
18. Williams M, Johnston JM. Training and maintaining the performance of dogs (Canis familiaris) on an increasing number of odor discriminations in a controlled setting. Appl Anim Behav Sci [Internet]. 2002 Aug;78(1):55–65. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0168159102000813

19. Poling A, Weetjens BJ, Cox C, Beyene NW, Sully A. Using Giant African Pouched Rats (Cricetomys Gambianus) to Detect Landmines. Psychol Rec [Internet]. 2010 Oct 29;60(4):715–28. Available from: http://link.springer.com/10.1007/BF03395741
20. Williams H, Pembroke A. SNIFFER DOGS IN THE MELANOMA CLINIC? Lancet [Internet]. 1989 Apr 1 [cited 2019 Jun 2];333(8640):734. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2564551
21. Church J, Williams H. Another sniffer dog for the clinic? Lancet [Internet]. 2001 Sep;358(9285):930. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0140673601060652
22. Campbell LF, Farmery L, George SMC, Farrant PBJ. Canine olfactory detection of malignant melanoma. Case Reports [Internet]. 2013 Oct 14;2013(oct14 1):bcr2013008566–bcr2013008566. Available from: https://casereports.bmj.com/lookup/doi/10.1136/bcr-2013-008566
23. Hirotsu T, Sonoda H, Uozumi T, Shinden Y, Mimori K, Maehara Y, et al. A Highly Accurate Inclusive Cancer Screening Test Using Caenorhabditis elegans Scent Detection. Lee M-H, editor. PLoS One [Internet]. 2015 Mar 11;10(3):e0118699. Available from: https://dx.plos.org/10.1371/journal.pone.0118699
24. Desikan P. Rapid diagnosis of infectious diseases: The role of giant African pouched rats, dogs and honeybees. Indian J Med Microbiol. 2013;31(2):114–6.
25. Suckling DM, Sagar RL. Honeybees Apis mellifera can detect the scent of Mycobacterium tuberculosis. Tuberculosis [Internet]. 2011 Jul;91(4):327–8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1472979211000904
26. Fischer-Tenhagen C, Wetterholm L, Tenhagen B, Heuwieser W. Training dogs on a scent platform for oestrus detection in cows. Appl Anim Behav Sci [Internet]. 2011 Apr;131(1–2):63–70. Available from: http://dx.doi.org/10.1016/j.applanim.2011.01.006
27. Fischer-Tenhagen C, Tenhagen B-A, Heuwieser W. Short communication: Ability of dogs to detect cows in estrus from sniffing saliva samples. J Dairy Sci [Internet]. 2013 Feb;96(2):1081–4. Available from: http://dx.doi.org/10.3168/jds.2012-5683
28. Golden GJ, Grady MJ, McLean HE, Shriner SA, Hartwig A, Bowen RA, et al. Biodetection of a specific odor signature in mallard feces associated with infection by low pathogenic avian influenza A virus. PLoS One [Internet]. 2021;16(5):e0251841. Available from: http://dx.doi.org/10.1371/journal.pone.0251841
29. Dorman DC, Foster ML, Fernhoff KE, Hess PR. Canine scent detection of canine cancer: a feasibility study. 2017 [cited 2018 Sep 26]; Available from: http://dx.doi.org/10.2147/VMRR.S148594
30. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med [Internet]. 2009 Jul 21;6(7):e1000097. Available from: http://www.prisma-statement.
31. Pickel D, Manucy GP, Walker DB, Hall SB, Walker JC. Evidence for canine olfactory detection of melanoma. Appl Anim Behav Sci [Internet]. 2004 Nov;89(1–2):107–16. Available from: https://linkinghub.elsevier.com/retrieve/pii/S016815910400111X
32. 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 Clin Appl Res [Internet]. 2012 Sep 1 [cited 2018 Oct 11];7(5):283–94. Available from: https://www.sciencedirect.com/science/article/pii/S1558787812000032

33. Horvath G, Andersson H, Nemes S. Cancer odor in the blood of ovarian cancer patients: a retrospective study of detection by dogs during treatment, 3 and 6 months afterward. 2013; Available from: http://www.biomedcentral.com/1471-2407/13/396
34. Amundsen T, Sundstrøm S, Buvik T, Gederaas OA, Haaverstad R. Can dogs smell lung cancer? First study using exhaled breath and urine screening in unselected patients with suspected lung cancer. Acta Oncol (Madr) [Internet]. 2014 Mar 19;53(3):307–15. Available from: http://www.tandfonline.com/doi/full/10.3109/0284186X.2013.819996
35. 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. Sensors Actuators B Chem [Internet]. 2014 Oct;202:615–21. Available from: http://dx.doi.org/10.1016/j.snb.2014.06.006
36. Elliker KR, Sommerville BA, Broom DM, Neal DE, Armstrong S, Williams HC. 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 [Internet]. 2014 Dec 27;14(1):22. Available from: BMC Urology
37. Rudnicka J, Walczak M, Jezierski T, Buszewski B. IS IT POSSIBLE TO DETECT LUNG CANCER BY TRAINED DOGS? Heal Probl Civiliz [Internet]. 2015;2(2):19–26. Available from: http://www.termedia.pl/doi/10.5114/hpc.2015.57108
38. Taverna G, Tidu L, Grizzi F, Torri V, Mandressi A, Sardella P, et al. Olfactory System of Highly Trained Dogs Detects Prostate Cancer in Urine Samples. J Urol [Internet]. 2015 Apr;193(4):1382–7. Available from: http://dx.doi.org/10.1016/j.juro.2014.09.099
39. Taverna G, Tidu L, Grizzi F, Stork B, Mandressi A, Seveso M, et al. The Ability of Dogs to Detect Human Prostate Cancer Before and After Radical Prostatectomy. EC Vet Sci [Internet]. 2015;1(2015):47–51. Available from: https://www.ecronicon.com/ecve/veterinary-science-ECVE-02-00007.php
40. Urbanová L, Vyhnánková V, Krisová Š, Pacík D, Nečas A. Intensive training technique utilizing the dog's olfactory abilities to diagnose prostate cancer in men. Acta Vet Brno [Internet]. 2015 [cited 2018 Oct 5];84(1):77–82. Available from: http://actavet.vfu.cz/
41. Yoel U, Gopas J, Ozer J, Peleg R, Shvartzman P. Canine scent detection of volatile elements, characteristic of malignant cells, in cell cultures. Isr Med Assoc J. 2015;17(9):567–70.
42. 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 [Internet]. 2006 Mar 25;5(1):30–9. Available from: http://journals.sagepub.com/doi/10.1177/1534735405285096
43. Hackner K, Errhalt P, Mueller MR, Speiser M, Marzluf BA, Schulheim A, et al. Canine scent detection for the diagnosis of lung cancer in a screening-like situation. J Breath Res [Internet]. 2016 Sep 27;10(4):046003. Available from: https://iopscience.iop.org/article/10.1088/1752-7155/10/4/046003
44. Willis CM, Britton LE, Swindells MA, Jones EM, Kemp AE, Muirhead NL, et al. 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. Br J Dermatol [Internet]. 2016 Nov;175(5):1020–9. Available from: http://doi.wiley.com/10.1111/bjd.14887

45. Guerrero-Flores H, Apresa-García T, Garay-Villar Ó, Sánchez-Pérez A, Flores-Villegas D, Bandera-Calderón A, et al. A non-invasive tool for detecting cervical cancer odor by trained scent dogs. BMC Cancer [Internet]. 2017 Dec 26;17(1):79. Available from: http://dx.doi.org/10.1186/s12885-016-2996-4
46. Kitiyakara T. The detection of Hepatocellular carcinoma (HCC) from patients' breath using canine scent detection: Proof of concept study. J Breath Res [Internet]. 2017;(December 2016). Available from: https://pubmed.ncbi.nlm.nih.gov/28649095/
47. Guirao Montes Á, Molins López-Rodó L, Ramón Rodríguez I, Sunyer Dequigiovanni G, Viñolas Segarra N, Marrades Sicart RM, et al. Lung cancer diagnosis by trained dogs†. Eur J Cardio-Thoracic Surg [Internet]. 2017 Dec 1 [cited 2018 Nov 14];52(6):1206–10. Available from: http://academic.oup.com/ejcts/article/52/6/1206/4030727
48. Seo IS, Lee HG, Koo B, Koh CS, Park HY, Im C, et al. Cross detection for odor of metabolic waste between breast and colorectal cancer using canine olfaction. PLoS One. 2018;13(2):1–9.
49. Thuleau A, Gilbert C, Bauër P, Alran S, Fourchotte V, Guillot E, et al. A New Transcutaneous Method for Breast Cancer Detection with Dogs. Oncology [Internet]. 2018 Oct 2 [cited 2018 Nov 5];1–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30278460
50. 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 [Internet]. 2018;5(March):52. Available from: http://journal.frontiersin.org/article/10.3389/fvets.2018.00052/full
51. Pacik D, Plevova M, Urbanova L, Lackova Z, Strmiska V, Necas A, et al. Identification of Sarcosine as a Target Molecule for the Canine Olfactory Detection of Prostate Carcinoma OPEN. Sci RepoRts | [Internet]. 2018 [cited 2018 Oct 5];8:4958. Available from: www.nature.com/scientificreports
52. Schoon GAA, De Jonge D, Hilverink P. How dogs learn to detect colon cancer—Optimizing the use of training aids. J Vet Behav [Internet]. 2020 Jan;35:38–44. Available from: https://doi.org/10.1016/j.jveb.2019.10.006
53. Gordon RT, Schatz CB, Myers LJ, Kosty M, Gonczy C, Kroener J, et al. The Use of Canines in the Detection of Human Cancers. J Altern Complement Med [Internet]. 2008 Jan [cited 2018 Sep 25];14(1):61–7. Available from: http://biodetectionk9s.org/wp-content/uploads/2017/01/The-Use-of-Canines-in-the-Detection-of-Human-Cancers.pdf
54. Biehl W, Hattesohl A, Jörres RA, Duell T, Althöhn U, Koczulla AR, et al. VOC pattern recognition of lung cancer: a comparative evaluation of different dog- and eNose-based strategies using different sampling materials. Acta Oncol (Madr) [Internet]. 2019 Sep 2;58(9):1216–24. Available from: https://doi.org/10.1080/0284186X.2019.1634284
55. Feil C, Stein T, Forster A, Schmidtmann I, Riemann-Seibert T, Berger M, et al. Diagnosis of lung cancer by canine olfactory detection in urine and breath samples. J Clin Oncol [Internet]. 2019 May 20;37(15_suppl):e13067–e13067. Available from: http://ascopubs.org/doi/10.1200/JCO.2019.37.15_suppl.e13067
56. Guirao A, Molins L, Ramón I, Sunyer G, Viñolas N, Marrades R, et al. Trained dogs can identify malignant solitary pulmonary nodules in exhaled gas. Lung Cancer [Internet]. 2019 Sep;135(June):230–3. Available from: https://doi.org/10.1016/j.lungcan.2019.06.008

57. Junqueira H, Quinn TA, Biringer R, Hussein M, Smeriglio C, Barrueto L, et al. Accuracy of Canine Scent Detection of Non–Small Cell Lung Cancer in Blood Serum. J Am Osteopath Assoc [Internet]. 2019 Jul 1 [cited 2021 Apr 26];119(7):413. Available from: https://pubmed.ncbi.nlm.nih.gov/31206136/
58. Murarka M, Vesley-Gross ZI, Essler JL, Smith PG, Hooda J, Drapkin R, et al. Testing ovarian cancer cell lines to train dogs to detect ovarian cancer from blood plasma: A pilot study. J Vet Behav [Internet]. 2019 Jul;32:42–8. Available from: https://doi.org/10.1016/j.jveb.2019.04.010
59. Protoshhak V V, Andreev EA, Karpushhenko EG, Slepcov A V, Ovchinnikov D V, Alentev SA, et al. [Prostate cancer and dogs sense of smell: opportunities of noninvasive diagnostics]. Urologiia [Internet]. 2019 Dec;(5):22–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31808627
60. Kure S, Yamada M. Breast Cancer Detection from a Urine Sample by Dog Sniffing. :1–12.
61. Yamamoto A, Kamoi S, Kurose K, Ito M, Takeshita T, Kure S, et al. The Trained Sniffer Dog Could Accurately Detect the Urine Samples from the Patients with Cervical Cancer, and Even Cervical Intraepithelial Neoplasia Grade 3: A Pilot Study. Cancers (Basel) [Internet]. 2020 Nov 6;12(11):3291. Available from: https://www.mdpi.com/2072-6694/12/11/3291
62. Mazzola SM, Pirrone F, Sedda G, Gasparri R, Romano R, Spaggiari L, et al. Two-step investigation of lung cancer detection by sniffer dogs. J Breath Res [Internet]. 2020 Mar 11;14(2):026011. Available from: https://iopscience.iop.org/article/10.1088/1752-7163/ab716e
63. Guest C, Harris R, Sfanos KS, Shrestha E, Partin AW, Trock B, et al. Feasibility of integrating canine olfaction with chemical and microbial profiling of urine to detect lethal prostate cancer. PLoS One [Internet]. 2021;16(2 February):1–23. Available from: http://dx.doi.org/10.1371/journal.pone.0245530
64. Horvath G, Järverud G af K, Järverud S, Horváth I. Human Ovarian Carcinomas Detected by Specific Odor. Integr Cancer Ther [Internet]. 2008 Jun 27;7(2):76–80. Available from: http://journals.sagepub.com/doi/10.1177/1534735408319058
65. Horvath G, Andersson H, Paulsson G. Characteristic odour in the blood reveals ovarian carcinoma. BMC Cancer [Internet]. 2010 Dec 24 [cited 2018 Sep 25];10(1):643. Available from: http://www.biomedcentral.com/1471-2407/10/643
66. Cornu J-N, Cancel-Tassin G, Ondet V, Girardet C, Cussenot O. Olfactory Detection of Prostate Cancer by Dogs Sniffing Urine: A Step Forward in Early Diagnosis. Eur Urol [Internet]. 2011 Feb [cited 2018 Sep 26];59(2):197–201. Available from: https://www.europeanurology.com/article/S0302-2838(10)00944-9/pdf
67. Willis CM, Britton LE, Harris R, Wallace J, Guest CM. Volatile organic compounds as biomarkers of bladder cancer: Sensitivity and specificity using trained sniffer dogs. Cancer Biomarkers [Internet]. 2011 Oct 12;8(3):145–53. Available from: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/CBM-2011-0208
68. Sonoda H, Kohnoe S, Yamazato T, Satoh Y, Morizono G, Shikata K, et al. Colorectal cancer screening with odour material by canine scent detection. Gut [Internet]. 2011 Jun 1;60(6):814–9. Available from: https://gut.bmj.com/lookup/doi/10.1136/gut.2010.218305
69. 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 [Internet]. 2012 Mar 1 [cited 2018 Sep 25];39(3):669–76. Available from: www.erj.ersjournals.com

70. Koivusalo M, Vermeiren C, Yuen J, Reeve C, Gadbois S, Katz K. Canine scent detection as a tool to distinguish meticillin-resistant Staphylococcus aureus. J Hosp Infect [Internet]. 2017;96(1):93–5. Available from: http://dx.doi.org/10.1016/j.jhin.2017.03.005
71. Guest C, Pinder M, Doggett M, Squires C, Affara M, Kandeh B, et al. Trained dogs identify people with malaria parasites by their odour. Lancet Infect Dis [Internet]. 2019 Jun;19(6):578–80. Available from: http://dx.doi.org/10.1016/S1473-3099(19)30220-8
72. Maurer M, McCulloch M, Willey AM, Hirsch W, Dewey D. Detection of Bacteriuria by Canine Olfaction. Open Forum Infect Dis [Internet]. 2016 Apr 1 [cited 2020 Dec 11];3(2). Available from: /pmc/articles/PMC4866566/?report=abstract
73. Bomers MK, van Agtmael MA, Luik H, van Veen MC, Vandenbroucke-Grauls CMJE, Smulders YM. Using a dog's superior olfactory sensitivity to identify Clostridium difficile in stools and patients: proof of principle study. BMJ [Internet]. 2012 Dec 13;345(dec13 8):e7396–e7396. Available from: https://www.bmj.com/lookup/doi/10.1136/bmj.e7396
74. Bomers MK, van Agtmael MA, Luik H, Vandenbroucke-Grauls CMJE, Smulders YM. A detection dog to identify patients with Clostridium difficile infection during a hospital outbreak. J Infect [Internet]. 2014 Nov;69(5):456–61. Available from: http://dx.doi.org/10.1016/j.jinf.2014.05.017
75. Bryce E, Zurberg T, Zurberg M, Shajari S, Roscoe D. Identifying environmental reservoirs of Clostridium difficile with a scent detection dog: preliminary evaluation. J Hosp Infect [Internet]. 2017 Oct;97(2):140–5. Available from: http://dx.doi.org/10.1016/j.jhin.2017.05.023
76. Taylor MT, McCready J, Broukhanski G, Kirpalaney S, Lutz H, Powis J. Using Dog Scent Detection as a Point-of-Care Tool to Identify Toxigenic Clostridium difficile in Stool. Open Forum Infect Dis [Internet]. 2018 Aug 1;5(8):1–4. Available from: https://academic.oup.com/ofid/article/doi/10.1093/ofid/ofy179/5056931
77. Li C, Zurberg T, Kinna J, Acharya K, Warren J, Shajari S, et al. Using scent detection dogs to identify environmental reservoirs of Clostridium difficile: Lessons from the field. Can J Infect Control [Internet]. 2019;34(2):93–5. Available from: https://ipac-canada.org/photos/custom/CJIC/CJIC_Summer2019_Li.pdf
78. Grandjean D, Sarkis R, Lecoq-Julien C, Benard A, Roger V, Levesque E, et al. Can the detection dog alert on COVID-19 positive persons by sniffing axillary sweat samples? A proof-of-concept study. PLoS One. 2020;15(12 December):1–19.
79. Jendrny P, Schulz C, Twele F, Meller S, von Köckritz-Blickwede M, Osterhaus ADME, et al. Scent dog identification of samples from COVID-19 patients – a pilot study. BMC Infect Dis [Internet]. 2020 Dec 23;20(1):536. Available from: https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-020-05281-3
80. Vesga O, Valencia AF, Mira A, Ossa F, Ocampo E, Agudelo M, et al. Dog Savior: Immediate Scent-Detection of SARS-COV-2 by Trained Dogs. bioRxiv [Internet]. 2020 Jan 1;2020.06.17.158105. Available from: http://biorxiv.org/content/early/2020/06/19/2020.06.17.158105.abstract
81. Eskandari E, Ahmadi Marzaleh M, Roudgari H, Hamidi Farahani R, Nezami-Asl A, Laripour R, et al. Sniffer dogs as a screening/diagnostic tool for COVID-19: a proof of concept study. BMC Infect Dis [Internet]. 2021 Dec 5 [cited 2021 Jun 8];21(1):243. Available from: https://pubmed.ncbi.nlm.nih.gov/33673823/
82. Essler JL, Kane SA, Nolan P, Akaho EH, Berna AZ, DeAngelo A, et al. Discrimination of SARS-CoV-2 infected patient samples by detection dogs: A proof of concept study. Leal WS, editor. PLoS One [Internet]. 2021 Apr

14;16(4):e0250158. Available from: http://dx.doi.org/10.1371/journal.pone.0250158

83. Grandjean D, Marzooqi DH Al, Lecoq-Julien C, Hammadi HK Al, Alvergnat G, Blooshi KM Al, et al. Use Of Canine Olfactory Detection For COVID-19 Testing Study On U.A.E. Trained Detection Dog Sensitivity. bioRxiv [Internet]. 2021;1–29. Available from: https://biorxiv.org/cgi/content/short/2021.01.20.427105
84. Guest C, Dewhirst SY, Allen DJ, Aziz S, Baerenbold O, Chabildas U, et al. Using trained dogs and organic semi-conducting sensors to identify asymptomatic and mild SARS-CoV-2 infections . Available from: https://www.lshtm.ac.uk/media/49791
85. Jendrny P, Twele F, Meller S, Schulz C, von Koeckritz-Blickwede M, Osterhaus A, et al. Scent dog identification of SARS-CoV-2 infections, similar across different body fluids. bioRxiv [Internet]. 2021;2021.03.05.434038. Available from: http://biorxiv.org/content/early/2021/03/05/2021.03.05.434038.abstract
86. Petri AL, Høgdall C, Christensen IJ, Simonsen AH, T'Jampens D, Hellmann M-L, et al. Sample handling for mass spectrometric proteomic investigations of human urine. PROTEOMICS - Clin Appl. 2008 Sep;2(9):1184–93.
87. Shen Y-S, Ku Y. Treatment of gas-phase volatile organic compounds (VOCs) by the UVO3 process. Chemosphere [Internet]. 1999 Apr;38(8):1855–66. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0045653598004007
88. Sato T, Katsuoka Y, Yoneda K, Nonomura M, Uchimoto S, Kobayakawa R, et al. Sniffer mice discriminate urine odours of patients with bladder cancer: A proof-of-principle study for non-invasive diagnosis of cancer-induced odours. Sci Rep [Internet]. 2017 Dec 7;7(1):14628. Available from: http://dx.doi.org/10.1038/s41598-017-15355-z
89. Edwards TL, Browne CM, Schoon A, Cox C, Poling A. Animal olfactory detection of human diseases: Guidelines and systematic review. J Vet Behav Clin Appl Res [Internet]. 2017;20(May):59–73. Available from: http://dx.doi.org/10.1016/j.jveb.2017.05.002
90. Lit L, Schweitzer JB, Oberbauer AM. Handler beliefs affect scent detection dog outcomes. Anim Cogn [Internet]. 2011 May 12 [cited 2018 Oct 2];14(3):387–94. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3078300/pdf/10071_2010_Article_373.pdf
91. Lazarowski L, Krichbaum S, DeGreeff LE, Simon A, Singletary M, Angle C, et al. Methodological Considerations in Canine Olfactory Detection Research. Front Vet Sci [Internet]. 2020 Jul 17;7(July):1–17. Available from: https://www.frontiersin.org/article/10.3389/fvets.2020.00408/full
92. Pirrone F, Albertini M. Olfactory detection of cancer by trained sniffer dogs: A systematic review of the literature. J Vet Behav Clin Appl Res [Internet]. 2017;19:105–17. Available from: http://dx.doi.org/10.1016/j.jveb.2017.03.004
93. Weber CM, Cauchi M, Patel M, Bessant C, Turner C, Britton LE, et al. Evaluation of a gas sensor array and pattern recognition for the identification of bladder cancer from urine headspace. Analyst. 2011;136(2):359–64.
94. Holz O, Mücke M, Zarza P, Loppow D, Jörres RA, Magnussen H. Freezing of homogenized sputum samples for intermittent storage. Clin Exp Allergy [Internet]. 2001 Aug;31(8):1328–31. Available from: http://doi.wiley.com/10.1046/j.1365-2222.2001.01136.x

95. Wilson AD, Baietto M. Advances in electronic-nose technologies developed for biomedical applications [Internet]. Vol. 11, Sensors. Molecular Diversity Preservation International; 2011 [cited 2021 Jun 24]. p. 1105–76. Available from: www.mdpi.com/journal/sensors
96. Lippi G, Cervellin G. Canine olfactory detection of cancer versus laboratory testing: Myth or opportunity. Clin Chem Lab Med. 2012;50(3):435–9.
97. Slabbert J., Odendaal JS. Early prediction of adult police dog efficiency—a longitudinal study. Appl Anim Behav Sci [Internet]. 1999 Aug;64(4):269–88. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0168159199000386
98. Weiss E. Selecting Shelter Dogs for Service Dog Training. J Appl Anim Welf Sci [Internet]. 2002 Jan;5(1):43–62. Available from: http://www.tandfonline.com/doi/abs/10.1207/S15327604JAWS0501_4
99. Turcsán B, Kubinyi E, Miklósi Á. Trainability and boldness traits differ between dog breed clusters based on conventional breed categories and genetic relatedness. Appl Anim Behav Sci [Internet]. 2011 Jun;132(1–2):61–70. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0168159111000864
100. Fischer-Tenhagen C, Johnen D, Heuwieser W, Becker R, Schallschmidt K, Nehls I. Odor Perception by Dogs: Evaluating Two Training Approaches for Odor Learning of Sniffer Dogs. Chem Senses [Internet]. 2017 Jun 1;42(5):435–41. Available from: https://academic.oup.com/chemse/article-lookup/doi/10.1093/chemse/bjx020
101. Weetjens BJC, Mgode GF, Machang'u RS, Kazwala R, Mfinanga G, Lwilla F, et al. African pouched rats for the detection of pulmonary tuberculosis in sputum samples. Int J Tuberc Lung Dis. 2009;13(6):737–43.
102. Mahoney A, Weetjens B, Cox C, Beyene N, Mgode G, Jubitana M, et al. Using giant African pouched rats to detect tuberculosis in human sputum samples: 2010 findings. Pan Afr Med J [Internet]. 2011 Oct 26;9(1):367–71. Available from: http://www.ajol.info/index.php/pamj/article/view/71204
103. Kaulfuß P, Mills DS. Neophilia in domestic dogs (Canis familiaris) and its implication for studies of dog cognition. Anim Cogn. 2008 Jul;11(3):553–6.
104. Hawken MP, Muhindi DW, Chakaya JM, Bhatt SM, Ng'ang'a LW, Porter JDH. Under-diagnosis of smear-positive pulmonary tuberculosis in Nairobi, Kenya. Int J Tuberc Lung Dis [Internet]. 2001 Apr;5(4):360–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11334255
105. Fend R, Kolk AHJ, Bessant C, Buijtels P, Klatser PR, Woodman AC. Prospects for Clinical Application of Electronic-Nose Technology to Early Detection of Mycobacterium tuberculosis in Culture and Sputum. J Clin Microbiol [Internet]. 2006 Jun;44(6):2039–45. Available from: https://journals.asm.org/doi/10.1128/JCM.01591-05
106. Hackner K, Pleil J. Canine olfaction as an alternative to analytical instruments for disease diagnosis: understanding "dog personality" to achieve reproducible results. J Breath Res [Internet]. 2017 [cited 2018 Sep 28];11:12001. Available from: http://iopscience.iop.org/article/10.1088/1752-7163/aa5524/pdf
107. Abergavenny RD, London LE. BMA warns against unnecessary screening tests in private sector Plan for genetic testing of German civil servants stirs controversy Global Fund pulls grants to Myanmar. 2005;2005. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1199058/

Tables

Due to technical limitations, tables 1-7 xlsx is only available as a download in the Supplemental Files section.

Figures

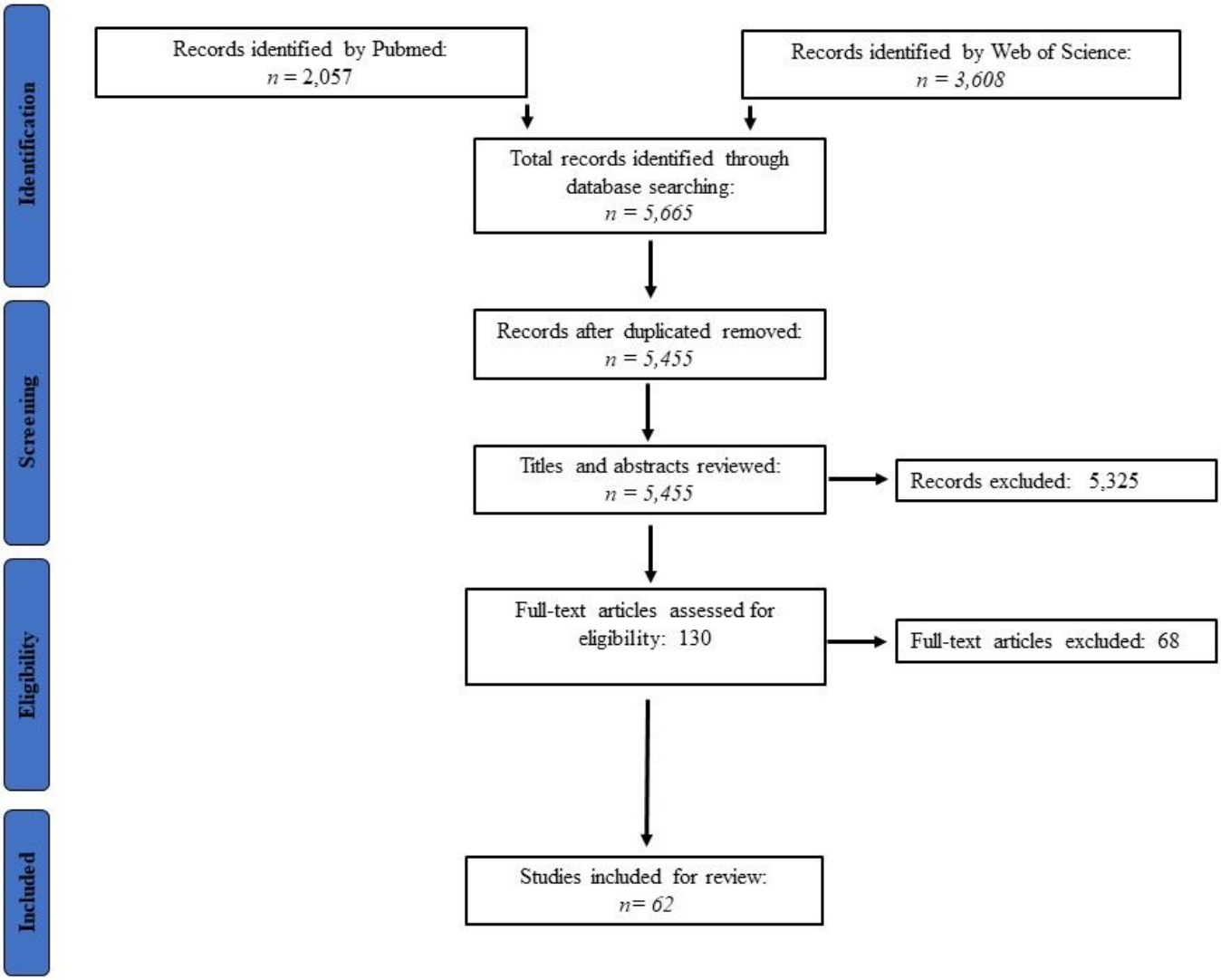

Figure 1

Flow chart

Figure 2

Evolution of the number of peer reviewed publications per year among selection

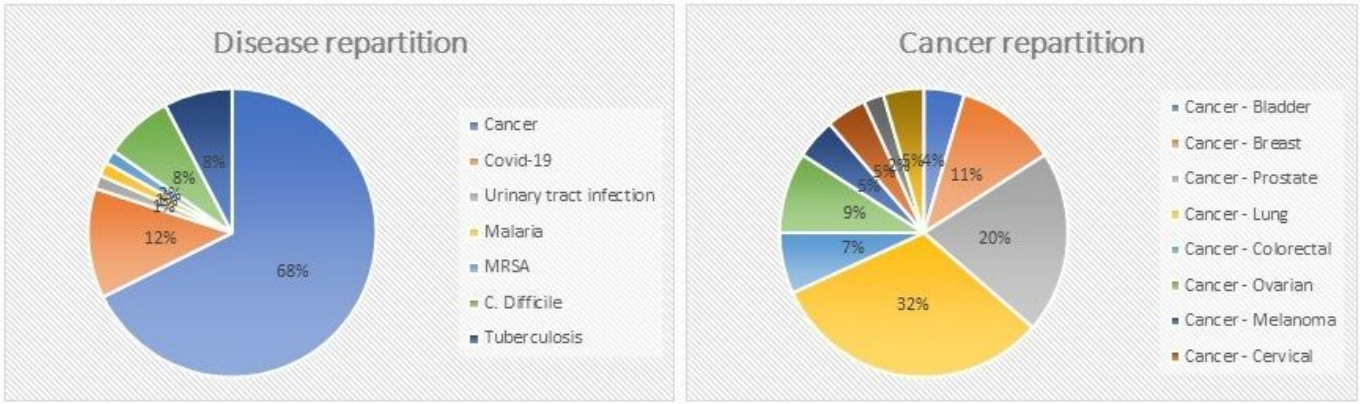

Figure 3

Targeted disease répartition among selected articles

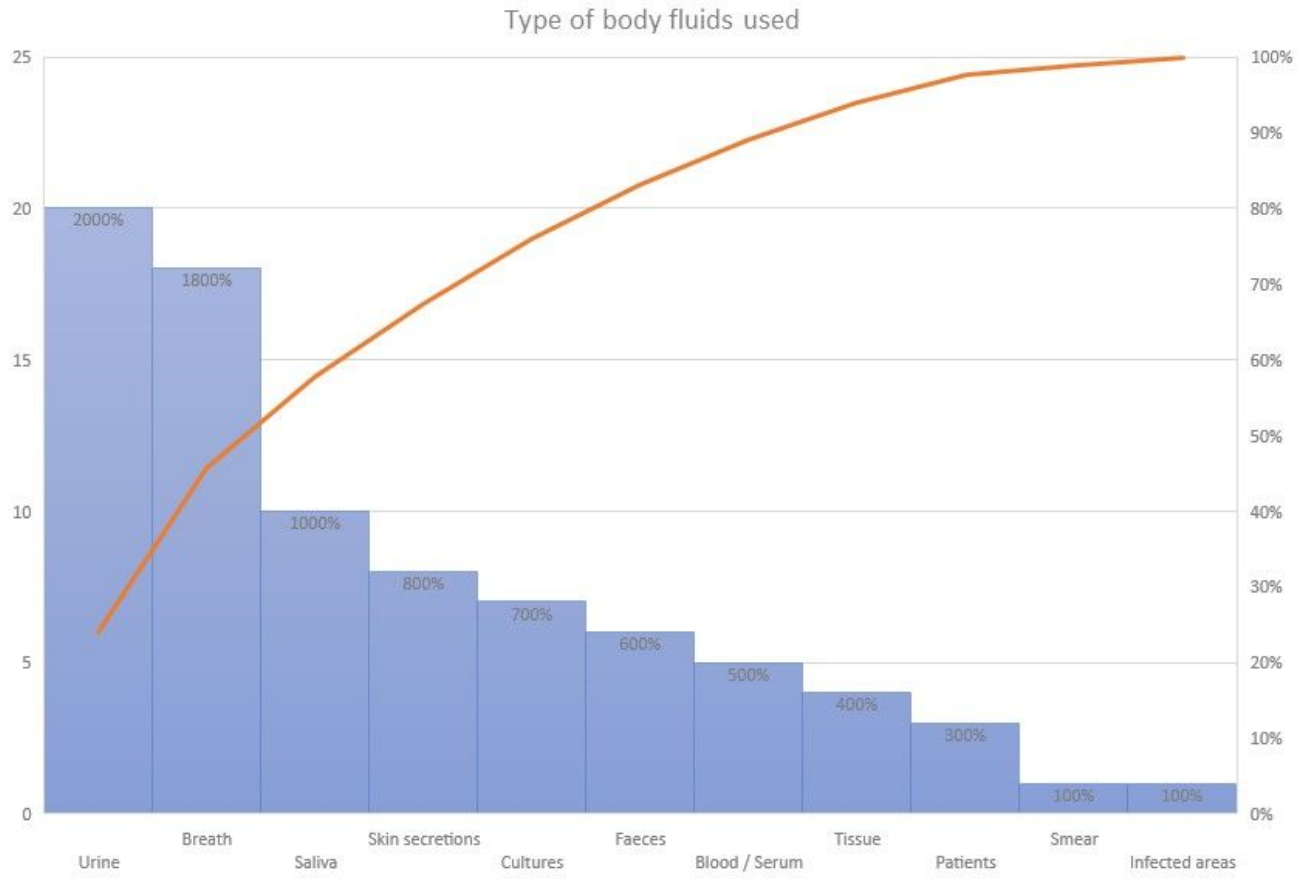

Figure 4

Type of body fluids used

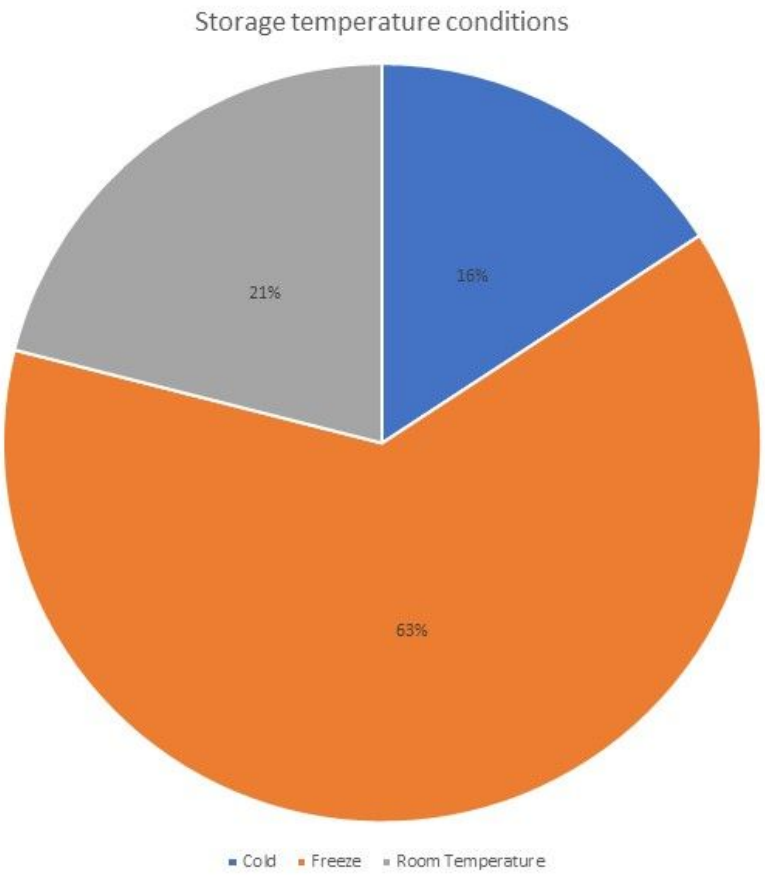

Figure 5

Samples’ storage temperatures

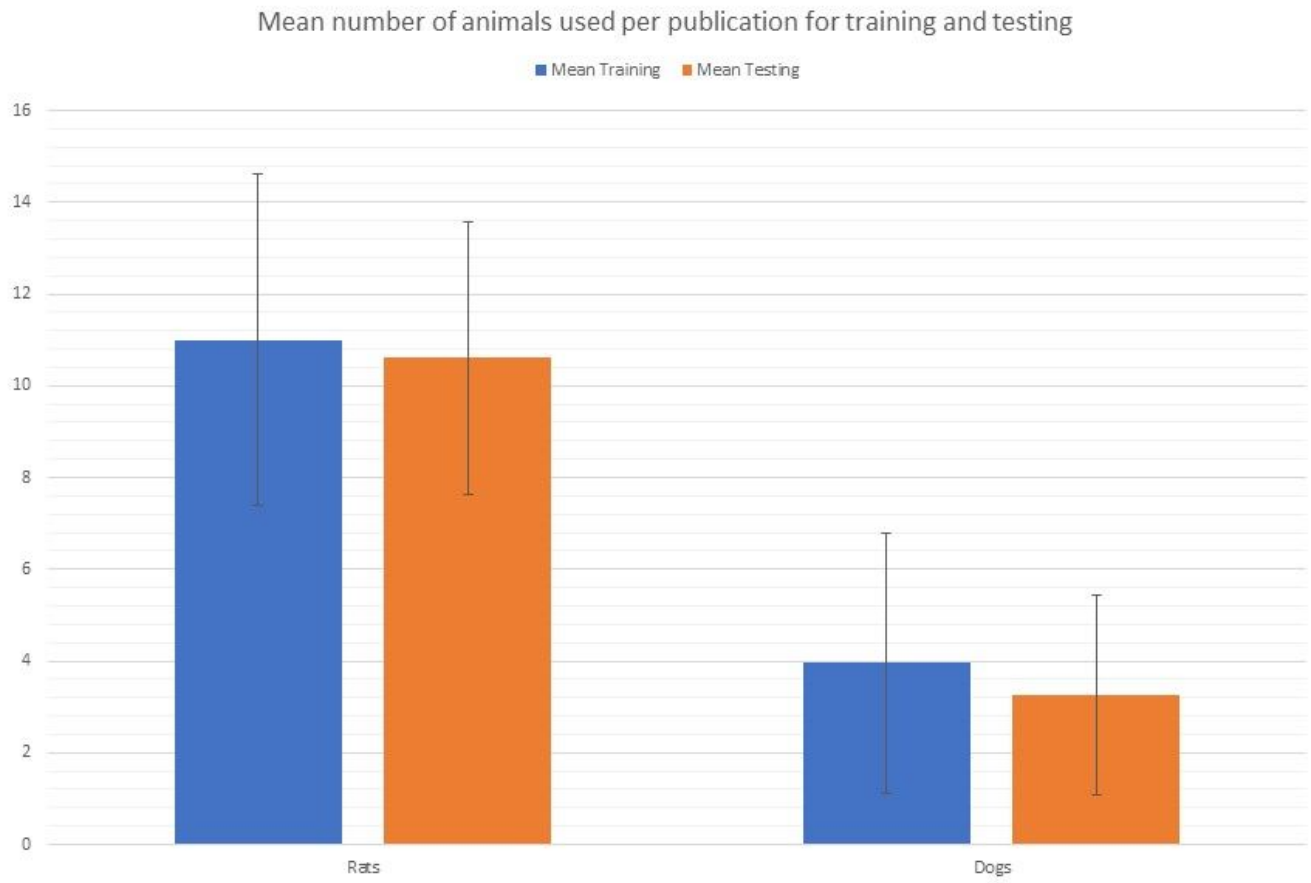

Figure 6

Mean number of animals used per publication for training and testing

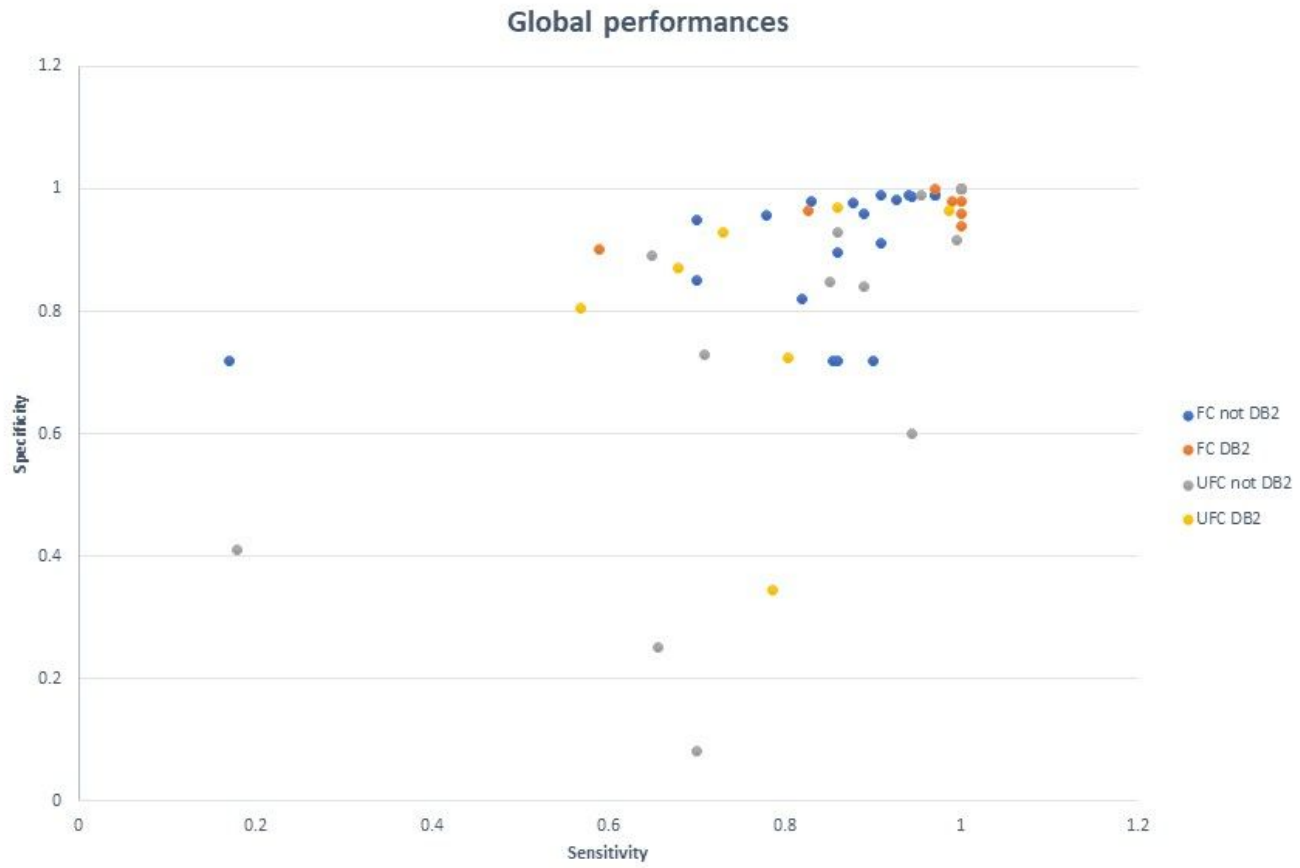

Figure 7

Global performances (data shown only for studies providing both Sensitivity and Specificity results)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- KDOGReviewTablesforBMCv20210723.xlsx