Aus der
Tierklinik für Fortpflanzung
des Fachbereichs Veterinärmedizin
der Freien Universität Berlin

How to train dogs to detect cows in heat by smell – Lessons learned by training scent detection dogs

Inaugural-Dissertation zur Erlangung des Grades eines Doktors
der Veterinärmedizin an der Freien Universität Berlin

vorgelegt von
Dorothea Johnen
Tierärztin aus Berlin

Berlin 2016

Journal-Nr.: 3864

Gedruckt mit Genehmigung des Fachbereichs Veterinärmedizin
der Freien Universität Berlin

Dekan:	Univ.-Prof. Dr. Jürgen Zentek
Erster Gutachter:	Univ.-Prof. Dr. Wolfgang Heuwieser
Zweiter Gutachter:	Univ.-Prof. Dr. Christa Thöne-Reineke
Dritter Gutachter:	Univ.-Prof. Dr. Hansjoachim Hackbarth

Deskriptoren (nach CAB-Thesaurus):
dog, canine, scent detection, dog training, quality of literature, oestrus detection, dairy cow, best practice standard

Tag der Promotion: 17.06.2016

Bibliographische Information der *Deutschen Nationalbibliothek*
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.de abrufbar.

ISBN: 978-3-00-056594-6

Meinen Eltern und Großeltern

TABLE OF CONTENT

1 INTRODUCTION

Humans have used the dogs' olfactory capacities for different types of scent detection tasks, especially as assistance in hunting, for hundreds of years. In recent times the use of scent detection dogs beyond the typical tasks such as drugs and explosives detection is expanding and there are numerous scientific publications proving scent detection abilities of dogs in various fields of applications (Browne et al., 2006). Trained scent detection dogs are used for various biological purposes, as locating environmental pollution (Arner et al., 1986; Reynolds et al., 2008), conservation of endangered animal species (Cablk, 2008; Kerley, 2010), elimination of pests such as rodents (Gsell et al., 2010), screwworms (Welch, 1990) or bed bugs (Vaidyanathan and Feldlaufer, 2013; Cooper et al., 2014) and fecal contamination on raw produce (Partyka et al., 2014). In addition several applications in human medicine have been described and tested (Bijland et al., 2013).

It is known that trained dogs are able to discriminate estrus specific odour in dairy cows by smell in different types of samples (Kiddy et al., 1984; Fischer-Tenhagen et al., 2011), but to our knowledge there are no reports that dogs have worked at cows for estrus detection on dairy farms.

The overall objectives of this thesis were to develop a specific training protocol for training dogs to identify cows in estrus under farm conditions and to identify factors in the procedure of scent detection dog training and testing that could limit applicability or bias the study outcome. A total of three studies were conducted.

1.1 Systematic literature evaluation concerning canine scent detection training and testing and comparison of two different testing systems

Evaluating the scent detection capability of trained dogs is challenging. On one hand dogs are able to notice hidden clues (i.e. unconscious reactions of the trainer), known as the "Clever Hans" effect (Moll, 1904; Lit et al., 2011). On the other hand dogs can probably recognize and indicate individual samples instead of searching for a specific target scent (Elliker et al., 2014). Methodological differences in the design of studies concerning canine scent detection make it hard to directly compare and to evaluate their results. For determining accuracy or sensitivity and specificity of canine scent detection different testing strategies were used in previous studies, such as the free search for the target scent in a defined area (Paula et al., 2011; Savidge et al., 2011), differentiation tasks in which the dogs had to find a positive sample between negative samples (Jezierski et al., 2008; Richards et al., 2008), a special testing platform (Fischer-Tenhagen et al., 2011), a multiple-choice apparatus (Fjellanger et al., 2002) or a skinner box.

The objective of my first study was to systematically evaluate the quality and comparability of published literature concerning canine scent detection and to determine a possible influence of the testing system on the outcome of a scent detection task considering two different testing systems.

The results of this study were recently published in Applied Animal Behaviour Science (impact factor 2.255):

D. Johnen, W.Heuwieser, C. Fischer-Tenhagen. 2013. Canine scent detection - Fact or fiction? 148: 201-208.

1.2 Practical training of dogs to detect cows in heat by smell

Efficient and accurate estrus detection is a key management factor determining acceptable reproductive performance in dairy herds (Heersche Jr and Nebel, 1994; At-Taras and Spahr, 2001). Average estrus detection rates on commercial dairy farms, however, are below 60% (Senger, 1994; Becker et al., 2005). In several experiments, dogs were trained to distinguish between different body fluid samples (i.e. vaginal mucus, urine, milk, saliva) of cows in estrus and samples of cows in diestrus (Kiddy et al., 1978; Hawk et al., 1984; Kiddy et al., 1984; Jezierski, 1992; Fischer-Tenhagen et al., 2011; Fischer-Tenhagen et al., 2013). Kiddy et al. (1978) trained dogs to detect estrus by direct examination of only 12 cows placed side by side in groups of three in adjacent stalls. Dogs searched from a position behind the cows walking on a platform to raise the dogs to a convenient working height. Mean percentage of correct detections was 87.3%. Such it could be proven that dogs actually could identify cows in estrus by smell.

The objectives of this study were to develop a specific training protocol for training dogs to identify cows in estrus from the feed alley and to determine sensitivity and specificity of trained dogs to detect cows in estrus.

The results of this study were recently published in Applied Animal Behaviour Science (impact factor 2.255):

D. Johnen, W.Heuwieser, C. Fischer-Tenhagen. 2015. How to train a dog to detect cows in heat - Training and success. 171: 39-46

The two papers are presented in the format outlined in the guide for authors of the journal.

1.3 Identifying possible sources of bias in scent detection dog testing

When we first started to conduct scientific studies on scent detection with trained dogs we searched the scientific literature in hope to find accepted quality standards to perform scent dog studies and tests to minimize potential biases. Unfortunately, adequate guidelines were missing in the literature. In a systematic review we showed that precise information on test criteria such as reliability, validity and objectivity of the scent detection performance tests were not presented in most reviewed publications (Johnen et al., 2013), which agrees with Furton and Myers (2001). The information, however, is essential to evaluate the quality of the test results and compare results of similar studies. Potential biases influencing the results of a study on scent detection work can be caused by flaws of the training and testing procedure. Therefore the objective of this systematic review was to evaluate recent publications on scent dog studies and identify factors in the procedure that can eventually bias the study outcome. Finally we wanted to summarize this information to recommend a best practice standard for scent dog studies.

These data have not been published yet and are presented in section 4 “Additional unpublished work”.

2 PUBLICATION I

Canine Scent Detection – Fact or Fiction?

Dorothea Johnen*, Wolfgang Heuwieser*, Carola Fischer-Tenhagen*

*Clinic for Animal Reproduction, Faculty of Veterinary Medicine, Freie Universität Berlin, Berlin, Germany

Published in:
Applied Animal Behaviour Science, September 2013, Volume 148, Issue 3-4, Pages 201-208

Please find the original article via the following digital object identifier:
http://dx.doi.org/10.1016/j.applanim.2013.09.002

2.1 Abstract

Dogs have been used in a variety of scent detection tasks for hundreds of years. However, methodological differences in the design of studies concerning canine scent detection make it hard to directly compare and to evaluate their results. We set out to (1) evaluate the quality and comparability of published literature concerning canine scent detection according to criteria of evidence-based medicine and (2) to determine the influence of the testing system on the outcome of a scent detection task considering two different testing systems. For the literature evaluation we retrieved 31 studies. After applying specific exclusion criteria 14 studies were left for final evaluation. A check list detailing relevant information about the study design and the training and testing process was used. Our results demonstrate many differences in methodology and a high variability of the results of those studies leading to diversity in respect to relevant quality criteria. For the second part of our study seven dogs were trained by means of positive reinforcement to detect black tea as target scent in two different testing systems, a testing platform and a scent detection board. Our data show that using an optimized training strategy high sensitivity (92.1%) and specificity (97.4%) can be achieved in a short time. Sensitivity and specificity for the detection of black tea was similar for the two testing systems.

2.2 Keywords

Dog, Canine, Scent detection, Dog training, Testing system, Quality of literature

2.3 Introduction

Dogs have a highly sensitive olfactory system and therefore been used in a variety of scent detection tasks for hundreds of years. There are many reports mostly of anecdotal evidence about the “amazing scent detection abilities” of trained dogs. In most cases, however, objective data on test characteristics of the scent detection performance of those dogs were not presented. Therefore the quality and validity of those reports are questionable. Today the most important and frequent applications are detection of explosives and land mines for police and military (Gazit and Terkel, 2003). Recently, several applications in human medicine (e.g. cancer, diabetes) have been described and tested (Moser and McCulloch, 2010) and it is speculated that the importance will increase considerably. There are many other scent detection applications for which dogs were used (Browne et al., 2006) and where studies have been conducted to evaluate their reliability, e.g. indication of toxic contamination of the environment (Arner et al., 1986), illicit discharge polluting habits (Reynolds et al., 2008) and mold formation and other microbial growth in houses (Kauhanen et al., 2002). Trained dogs can also contribute to the protection of endangered animal species by detecting their feces and identify individual animals by scent matching (Kerley, 2010) or by detecting animals in their natural habitats (Cablk and Heaton, 2006). The dog’s olfactory capabilities were also used in elimination of pests, such as rodents (Gsell et al., 2010) or screwworms (Welch, 1990). In dairy research, dogs have been trained to identify estrus specific odors in different body fluids (Kiddy et al., 1978; Hawk et al., 1984; Kiddy et al., 1984; Fischer-Tenhagen et al., 2011).

Evaluating the scent detection capability of dogs is challenging. On one hand dogs are able to notice hidden clues (i.e. unconscious reactions of the trainer), known as the “Clever Hans” effect (Moll, 1904). On the other hand dogs can easily recognize individual samples instead of searching for the target scent. As it takes only 1 to 2 s for a dog to determine the direction of an odor trail made by the footsteps of the person hiding the target (Hepper and Wells, 2005) it is possible that a dog simply follows those traces instead of performing a free search for the target.

Methodological differences in the design of studies concerning canine scent detection make it hard to directly compare and to evaluate their results. For determining accuracy or sensitivity and specificity of canine scent detection different testing strategies were used in previous studies, such as the free search for the target scent in a defined area (Paula et al., 2011), a differentiation task in which the dogs had to find a positive sample between negative samples (Richards et al., 2008), a special testing platform (Fischer-Tenhagen et al., 2011), a multiple-choice apparatus (Fjellanger et al., 2002) or a skinner box (Göth, 2003). Usually, trainers and handlers judge their dogs’ ability to perform certain scent detection tasks as high and flawless. There is a dearth of science-based information, however, on test characteristics of dogs as a diagnostic test for the target scent eliminating hidden clues or other hints. Considering the wide variety of training and testing methods and the diversity of publications the overall objective of this study was twofold: First, to evaluate the quality and comparability of published literature concerning canine scent detection according to criteria of evidence-based veterinary medicine. Secondly, to determine the influence of the testing system on the outcome of a scent detection task considering two different testing systems.

2.4 Material and methods

2.4.1 Literature evaluation

A systematic literature research was conducted on 5th June 2012 using the databases Pubmed (www.pubmed.gov) and CAB (http://ovidsp.tx.ovid.com) to find studies concerning scent detection work in dogs. The subject headings “scent” AND “detection” AND “dogs” were used to find articles written in English or German language about training dogs for specific scent detection tasks. In a supplementary hand search additional publications were recruited. We excluded duplicates, systematic reviews without original data and papers that did not include canine scent detection training. The remaining articles were evaluated using a check list detailing relevant information about the study design and the training and testing process. Specifically, we recorded quality criteria of the analyzed studies, such as detailed description of dog training, number of dogs and percentage of dogs finishing the training process, training duration, blinding of the dog trainer and other personnel towards the sample position in testing, random placement of samples in testing, use of new samples in testing mode and type of task (differentiation vs. free search). Retrieval and management of references was performed with Endnote (Version X4 for Windows, Thomson Reuters, New York, USA).

2.4.2 Own study

2.4.2.1 Dogs

Seven privately owned pet dogs (two female Labrador retrievers, one and six years old, both spayed; a female Berger de Pyrénées, six years old; a female German Jagdterrier, six years old; a male Border collie, one year old; a male Bernese mountain dog, six months old and a female Shepherd cross-breed, nine years old) were enrolled.

All dogs had previously been trained to detect at least one target substance (e.g. dried chamomile) by scent on the training platform. Of the seven dogs, three had previously been trained to detect at least one target substance by scent on the scent detection board.

2.4.2.2 Training laboratory

Training and testing took place in an indoor laboratory at the Nordiska Hund training center, Kälarne, Sweden, using a training platform (Fig. 1) and a scent detection board (Fig. 2) as special training devices. The training platform (80 cm x 200 cm) was covered with a rubber mat for secure footing of the dogs and allowed to present different samples in a changing order and rapid succession to the dogs. This led to a high rewarding frequency as previously described by Fischer-Tenhagen et al. (2011). In brief, the training platform consisted of a movable sledge where seven stainless steel jars with perforated lids containing the test substrates were fixed each 20 cm apart, a cover panel with four holes and a vertical shield. The sledge could be moved to either side by the dog trainer with a handle to which a clicker was fixed. The jar holding the positive sample was always positioned directly opposite the handle. This assured that the trainer knew the position of the positive sample and could reward the dog for indicating the correct jar immediately and accurately. The sledge was running under the cover panel with four holes (diameter 6 cm). Through movement of the sledge the seven samples could be moved to four different positions under the cover panel. Consequently, the dogs had to identify one target sample out of four samples, i.e. three negatives and one positive. The vertical shield (80 cm x 60 cm) concealed the handle to assure that the dogs could not follow the trainer’s hand movements. A food treat (Educ, ROYAL CANIN Tiernahrung GmbH & Co. KG, Cologne, Germany) could be thrown into a ridge in front of the sledge to reward the dogs spatially close to its correct indication.

The scent detection board consisted of a wooden plank (250 cm x 40 cm) with eight holes (7 cm diameter, 30 cm apart) on the center line in which the jars containing the test substrates could be inserted. The dogs could walk over the plank and sniff on every jar in order to find the positive sample. The dogs began to search on one end of the board and were allowed to walk in both directions in order to check all eight samples. The board was divided into two sections with four holes each. Each section contained one positive and three negative jars.

To determine sensitivity and specificity a specific testing platform and the scent detection board were used. The testing platform was identical to the training platform but utilized a sledge with a total of eight stainless steel jars. The sledge could be moved to one of two positions. When the sledge was moved to position 1 the dog could check the first four samples. After the sledge had been moved to position 2 the dog could check the second set of four samples. In both positions the dog had to differentiate between one positive sample and three negative samples. The dog handler was blinded to the position of the positive samples.

The jars were loaded with the samples immediately before use and cleaned in a dish washer at 90 °C afterwards.

2.4.2.3 Dog training

According to the German and Swedish legislation no part of this study included any insult to any animal.

The dogs were trained with positive reinforcement using a clicker as conditioned secondary reinforcement and small pieces of food as reward. All the dogs were familiar with the sound of the clicker as a predictor for food. Scent detection training was adapted from previous studies that trained dogs for finding landmines (Göth, 2003) or identifying cancer in exhaled breath samples (McCulloch et al., 2006) and consisted of three modules, i.e. adaptation to the laboratory and training methods, imprinting to the target scent and actual discrimination training (Fischer-Tenhagen et al., 2011). In the beginning dogs were adapted to the laboratory and the devices (training platform, scent detection board, testing platform) by reinforcing them for standing on the platforms or walking on the scent detection board. When the dogs were familiarized with the training and testing devices, adaptation was finished. As the dogs had prior experience in scent detection, adaptation did not need more than 5 min per dog. The imprinting phase began on the training platform. The jar opposite to the handle was loaded with black tea as target substrate while the other jars were filled with tap water as negative samples. In the presentation mode the dogs were rewarded for every nose contact with the positive sample. The sledge was moved such that a rewarding frequency of 15 to 30 reinforced contacts in 2 min was reached. After a training time of 2 min all dogs indicated the jar containing black tea as the positive sample by pointing with their nose to it and ignoring the negative samples. The indication behavior was shaped by delaying the click stepwise until the dogs were able to stand still for at least 10 s pointing with the nose to the positive sample. Three to five replicates of 2 min each were conducted. To train discrimination negative jars were filled with chewing tobacco, coffee, cigarettes, cacao, milk powder and bread, respectively, as smelling distractions. These jars were used as negative samples instead of the tap water. One discrimination training session per dog was conducted on the testing platform in order to adapt them to the work on the testing platform. The whole training procedure was repeated on the scent detection board to familiarize the dogs with this device.

Overall, every dog passed five to eight training sessions on the training platform, three to five training sessions on the scent detection board and one adaptation session on the testing platform. The whole training process was completed within two days.

2.4.2.4 Evaluating sensitivity, specificity and the influence of the testing system on canine scent detection

To determine sensitivity and specificity of dogs for the detection of black tea and to evaluate the influence of the testing system (i.e. testing platform vs. scent detection board), two experiments were conducted. In the first experiment the dogs had to find black tea against tap water while in the second experiment the negative samples contained the smelling distractions. In both experiments all dogs had to pass five replicates on the testing platform and on the scent detection board each. In each replicate the dogs had to check eight jars divided in two sections. Each section contained one positive sample and three negative samples. The handlers were blinded regarding the position of the positive sample which was randomized with the random number function of Excel (Microsoft Office, Microsoft Cooperation, Redmont, USA). The experimenter was positioned 3 m apart from the testing platform or the scent detection board. He was not allowed to talk or make eye contact to the dog or the handler. If the dog handler identified an indication of his dog he announced it to the experimenter. In case of a correct positive indication the experimenter clicked, the dog was rewarded by the handler and the result documented. On the scent detection board the indicated positive sample was removed so the dog could continue searching for the second one. If the dog was not correct, the experimenter did not use the clicker and the dog was not rewarded. Such an event was documented as false positive indication. If the dog did not indicate a negative sample, it was documented as true negative indication but not rewarded. If the dog failed to indicate a positive sample because it made a false positive indication on another sample, a false negative indication was documented. In this case, the experimenter announced the position of the positive sample, so the handler could help the dog to find it and reward the dog.

Data were documented in Excel (Microsoft Office, Microsoft Cooperation, Redmont, USA). Mean differences between the different sample types (i.e. tea, tap water, smelling distractions) were evaluated with the Mann-Whitney-test. The level of significance was adjusted to reduce the type I error. For the comparison of three groups a P-value of 0.02 was assumed. Effect of individual dogs on the test result was evaluated using Kruskal-Wallis-test. Effect of the testing system on dog performance was evaluated using crosstabs function of descriptive statistics of SPSS (IBM Corporation, New York, USA).

2.5 Results

2.5.1 Literature evaluation

In total, 31 publications were retrieved (CAB 11, Pubmed 14, hand search six). After excluding doublets (n = 2), 29 publications remained. One publication was excluded because information from the title and abstract did not meet the inclusion criteria. Another 13 papers were excluded because they did not describe training and testing of scent detection work in dogs leaving 14 studies for final evaluation. Results are illustrated in tables 1and 2.

Target substrates for scent detection work were urine, breath or stool samples of cancer patients (n = 6), living animals (n = 3), tumour tissue (n = 1), bird carcasses (n = 1), body fluids of cows in oestrus (n = 1), microbes (n = 1) and nematode eggs in sheep faeces (n = 1). For none of the substrates a specific target scent was defined biochemically. Only in two studies dog training was described in detail that it would be possible to reproduce the training according to the instructions. All dog trainers worked with reward-based training and in six studies the clicker training was used. In one study, the dogs were verbally rebuked for wrong indications whereas in the other studies there was no information about the trainer's reaction on a wrong indication. Average number of dogs per study was 4.6 ± 3.2 and varied between one dog (three studies) and 10 dogs (one study). One study had started training with two dogs but one of them was excluded because of insufficient reliability. In two studies dogs with prior experience in scent detection work were utilized. In eight and four studies dogs did not have prior experiences or information concerning prior training was not provided, respectively. Duration of training varied from seven days to 16 months. In five studies training duration was not specified. In six studies all personnel was blinded regarding the position of the target substrate. In four studies information about blinding of the dog trainer or other personnel was not provided. In one study only the dog trainer was blinded but not another person being present during testing. In three studies the dog trainer was blinded but information was not provided if other personnel was present and blinded or not. In eight studies new samples not used in training were used in testing while in the other studies information on a potential reuse was not available. Number of testing trials per dog varied considerably. Random placement of samples during the testing was described in nine of the studies. For the remaining five studies information about the placement of samples was not provided. In 10 studies dogs were trained to perform a differentiation task in which they had to indicate a positive sample between several negative samples. A free search for the target in a defined area was performed by the dogs in three of the studies and in one study the dog had to perform both scent detection tasks. Training and testing were conducted in an indoor laboratory in eight of the studies. In one study, training and testing took place at the trainer's home and in two other buildings. For another study, training and testing were conducted indoors and outdoors. In four studies each training and testing locations were not described but in two studies testing was conducted outdoors. In six studies, sensitivity and specificity were calculated and varied from 88% to 100% for sensitivity and from 91% to 99% for specificity, respectively. In the remaining eight studies, testing results were declared as number of correct positive indications or other undefined rates varying from 35 % to 98%.

2.5.2 Own study

2.5.2.1 Evaluating sensitivity and specificity

To determine sensitivity and specificity seven dogs were utilized. Dogs had to distinguish between tea and tap water and between tea and smelling distractions on both the testing platform and the scent detection board. In total every dog had to conduct 20 replicates consisting of a total of 160 samples. Of those 40 were positive (i.e. black tea) and 120 were negative (i.e. 60 with tap water, 60 with smelling distractions). In total, 1120 samples were checked by the seven dogs. Overall, the seven dogs correctly identified 258 out of 280 black tea samples (i.e. sensitivity of 92.1%) and falsely identified black tea samples as negative samples in 22 cases (7.9% type II errors). Out of 840 negative samples 818 were correctly identified as true negative samples (i.e. specificity of 97.4%) and 22 negative samples were falsely identified as positive samples (2.6% type I errors). Considering the type of the negative sample (i.e. tap water and smelling distraction) sensitivity was 89.3% and 95.0% and specificity was 96.4% and 98.3%, respectively. As expected significant differences existed between tea and tap water ($p = 0.013$) and tea and smelling distractions ($p < 0.001$). There was no difference, however, between the negative sample types, i.e. tap water vs. smelling distractions ($p = 0.084$). The individual dog had no effect on the test result ($p = 0.645$).

2.5.2.2 Evaluating the influence of the testing system

Sensitivity and specificity (mean; CI 95%) for the testing platform and the training and testing board was 90.7% (86.1 - 95.3%) and 96.9% (90.2 - 97.0%) and 93.6 % (95.4 - 98.4%) and 97.9% (96.7 - 98.9%) respectively. Results of the two testing systems were similar ($p = 0.282$).

Table 1: Research articles (n=14) studying canine scent detection training and testing considering targets of scent detection, searching task, number, background of dogs trained for scent detection, training method and duration.

Author and year of publication	Target of scent detection	Searching task	Number of trained dogs	Type of trained dogs	Previous scent experience	Training method	Training duration
Brooks et al. (2003)	Termites	Diff.	6	no inf.	no inf.	Play and food reward-based	3 ws – 3 mo
Cornu et al. (2011)	Prostate cancer	Diff.	1	professional working dog	no	Clicker training (food reward)	16 mo
Ehmann et al. (2012)	Lung cancer	Diff.	4	family dogs	no inf.	Reward-based	no inf.
Fischer-Tenhagen et al. (2011)	Oestrus detection in cows	Diff.	7	family dogs	no	Clicker training (food reward)	7 d
Gordon et al. (2008)	Breast and prostate cancer	Diff.	10	trainer's family dogs	no	Clicker training (food reward)	no inf.
Horvath et al. (2008)	Ovarian carcinomas	Diff.	1	no inf.	no inf.	Play reward-based	12 mo
Kauhanen et al. (2002)	Microbial growth	Search	2	no inf.	no	Food reward-based	3 mo
Lin et al. (2011)	Fire ants	Diff./ search	3	no inf.	yes	Food reward-based	no inf.
McCulloch et al. (2006)	Lung and breast cancer	Diff.	5	family/prof. working dogs	no	Clicker training (food reward)	2-3 ws
Paula et al. (2011)	Bird and bat carcasses	Search	1	no inf.	no inf.	Play reward-based	4 mo
Richards et al. (2008)	Nematode infections in sheep	Diff.	1 (2)*	family dogs	no	Clicker training (play reward)	6 mo

Author and year of publication	Target of scent detection	Searching task	Number of trained dogs	Type of trained dogs	Previous scent experience	Training method	Training duration
Savidge et al. (2011)	Brown treesnakes	Search	2	no inf.	no inf.	Play reward-based	3-4 mo
Sonoda et al. (2011)	Colorectal cancer	Diff.	1	professional working dog	yes	Play reward-based	no inf.
Willis et al. (2004)	Bladder cancer	Diff.	10	no inf.	no	Clicker training	no inf.

no inf. = no information available in the study

Diff. = Differentiation

* = A second dog was withdrawn because of insufficient reliability

Table 2: Quality criteria of 14 trials studying canine scent detection training.

Author and year of publication	Trainer/all persons blinded	Number of testing replicates/dog (positive samples/negative samples)	Different samples in training and testing	Test characteristic calculated	Results
Brooks et al. (2003)	no inf.	75 (1/4)	no inf.	Positive indications False positives	74.7% - 97.3% 1.7% - 25.3%
Cornu et al. (2011)	yes/yes	33 (1/ 5)	yes	Sensitivity Specificity	91% 91%
Ehmann et al. (2012)	yes/yes	25 (1/4)	yes	Sensitivity Specificity	72% 94%
Fischer-Tenhagen et al. (2011)	yes/no inf.	24 (1/3)	yes	Sensitivity Specificity	80.3% 97.0%
Gordon et al. (2008)	yes/no inf.	2x9[b], 3x11[p] (1/6)	yes	Raw results Sensitivity, Specificity	11% - 28%[b], 6% - 28%[p] Plotted, no data provided
Horvath et al. (2008)	yes/yes	no inf.	no inf.	Sensitivity Specificity	100% 97.5%
Kauhanen et al. (2002)	yes/no inf.	5 (10/10), 5 (15/5), 42 (1/1)	no inf.	Indication of positive samples False positives	60% - 79%, 56% - 72%* 13%, 12%*
Lin et al. (2011)	yes/yes	10 (15/45)	yes	Positive indication rate False positive indication rate	>98% <2%
McCulloch et al. (2006)	yes/yes	no inf.	yes	Sensitivity Specificity	99%[l], 88%[b] 99%[l], 98%[b]
Paula et al.	no inf.	1 (26/0)	no inf.	Found carcasses	96%

Author and year of publication	Trainer/all persons blinded	Number of testing replicates/dog (positive samples/negative samples)	Different samples in training and testing	Test characteristic calculated	Results
(2011)					
Richards et al. (2008)	no inf.	8 (1/9)	no inf.	Mean rate of successful detection	85% -97.5%
Savidge et al. (2011)	yes/no	42, 43* (1/0)	no inf.	Success rate	35 %
Sonoda et al. (2011)	yes/yes	no inf.	yes	Sensitivity Specificity	91%[Breath s.], 97% [Stool s.] 99% [Breath s.], 99% [Stool s.]
Willis et al. (2004)	no inf.	9 (1/6)	yes	Mean success rate	41%

no inf. = no information available in the study, (x/y) = number of positive samples/number of negative samples

* = Dog 1, Dog 2, [b] = Breast cancer, [p] = Prostate cancer, [l] = Lung cancer

Figure 1: German Jagdterrier indicating the positive sample on the training platform.

Figure 2: Labrador retriever indicating the positive sample on the scent detection board.

2.6 Discussion

2.6.1 Literature evaluation

A systematic analysis of 14 papers describing training and testing of canine scent detection showed obvious differences in methodology and quality. In all studies training methods were based on positive reinforcement. There was no information, however, about the trainers' reactions and possible use of punishment in case of wrong indications except for one study. According to Walker et al. (2006), training methods based on positive reinforcement appear to have significant advantages compared to those based on aversive methods. In their study the authors trained dogs to detect *n*-amyl acetate (*n*AA) by means of positive reinforcement. Their dogs were able to detect *n*AA up to a 20,000-fold lower threshold compared to a study that utilized electro shocks and water deprivation (Krestel et al., 1984). Duration of training varied widely from seven days to 16 months indicating discrepancies in the training efficiency. One might speculate that the level of difficulty of the scent detection task might have influenced the duration. Two studies, however, both investigating cancer detection (McCulloch et al., 2006; Cornu et al., 2011) utilized dogs without prior experiences in scent detection work and described extremely different training durations (i.e. two to three weeks, 16 months). This difference for a comparable scent detection task is huge and allows the assumption that training efficiency was the essential difference between both studies. Interestingly, comments about the reason for such a long training duration were not provided. Most studies did not provide information about experience and qualification of dog trainers, which made it difficult to evaluate the results. Further research on relationships between a trainer's skills, training time needed for a defined task and training results is warranted. We speculate that the trainer might be a plausible influence on the duration necessary for training and the success rate. Two similar studies trained dogs to perform a free search for their targets, both used a play reward-based training system and both had a training duration of four months (Paula et al., 2011; Savidge et al., 2011). Interestingly, the success rates were considerably different with 96% (Paula et al., 2011) and 35% (Savidge et al., 2011) indicating an influence of the dog trainer. Another possible bias could be the different smell of the targets: bird carcasses (Paula et al., 2011) vs. living tree snakes (Savidge et al., 2011). We speculate that the low success rate of 35% could be due to a limitation of the dog's olfactory capability or to an insufficient training progress. As previously demonstrated the handler beliefs of the location of the target substrate can influence the outcome of scent detection work (Lit et al., 2011). Therefore, testing without blinding the handler potentially confounds results and can be considered a flaw in the study design. Thus four of the studies evaluated should be regarded carefully.

Presentation of results describing diagnostic performance was heterogeneous among the studies. In all studies that trained dogs for cancer detection results were given as sensitivity and specificity. As a gold standard, presence of tumors had been demonstrated by histopathological examinations of biopsies. Gordon et al. (2008) mentioned sensitivity and specificity but they were not provided as numerical values. In the other studies, results were presented as undefined rates such as success rate (Savidge et al., 2011), positive indication rate (Lin et al., 2011) or mean rate of successful detection (Richards et al., 2008) making it impossible to directly compare the results. In only six (42.9%) of 14 studies all personnel was blinded towards the sample positions and only six studies provided data on sensitivity and specificity. Only five studies (35.7%) met both criteria. Of those five studies one did not provide information on reuse of training samples in testing. Thus only four studies out of 14 (28.6%) met important quality criteria as they were blinded, calculated sensitivity and specificity and used new samples for testing. But in those four studies, other important information as the training duration was not provided or was extremely different (two to three weeks, 16 months).These differences in methodology and the high variability of the results show that the considered studies are diverse in respect to relevant quality criteria (Arlt et al., 2010).

2.6.2 Own study

2.6.2.1 Training progress

Our results demonstrate that dogs can learn quickly to perform a scent detection task as the indication of black tea. Dogs familiar with the training platform did not need more than eight training sessions to learn the correct indication of positive samples. With every training session lasting between 2 and 3 min the overall training time on the platform took less than half an hour per dog and was completed in one day. During this time each dog had approximately 300 reinforced contacts with the target scent. This rewarding frequency in combination with the short training makes the training platform an efficient and practical training device to introduce dogs to a new target scent. To the best of our knowledge other training devices allowing a comparable rewarding frequency do not exist. Compared to the 14 studies analyzed the training duration in our study was extremely short. Sensitivity (92.1%) and specificity (97.4%) however were similar to those studies providing data on sensitivity and specificity (sensitivity 72% - 100%, specificity 91% - 99%). As Fischer-Tenhagen at al. (2011) described the main advantage of the training platform in comparison with other training methods was the rewarding strategy. The use of the clicker allowed rewarding the dog exactly at the time of the desired behavior. In addition, the construction of the training platform permitted to present the food treat within centimeters to the positive sample. This could have helped the dogs to strengthen the link between the target scent and the rewarding food. To train the dogs to indicate black tea samples on the scent detection board took an overall training time of less than half an hour per dog. This short training time could be explained with the high number of rewarded contacts with the target scent the dogs previously had encountered on the training platform and is even shorter than the training time of

Williams and Johnston (2002), who trained dogs to identify 10 different odors. After a basic training the authors needed only two days for training new odors.

2.6.2.2 Evaluating sensitivity and specificity

Our data showed that training with a special training device and with a frequent rewarding strategy allows training dogs to perform a scent detection task as the detection of black tea with a high sensitivity and specificity. There was no significant difference in the sensitivity and specificity for negative samples with or without smelling distractions. It is well known that dogs have a more sensitive olfactory system than humans and are able to detect odors at lower concentrations (Hepper and Wells, 2005). As the scent of black tea is perceptible even for humans, this scent detection task should be easy for dogs. In a previous study addressing ovarian cancer sensitivity and specificity was 100% and 97.5%, respectively (Horvath et al., 2008). We assumed that it was easier for a dog to detect the smell of tea than the smell of ovarian cancer and expected a sensitivity and specificity close to 100%. However, Horvath et al. (2008) spent 12 months on training in contrast to two days in our study. Our data show that using an optimized training strategy high sensitivity (92.1%) and specificity (97.4%) can be achieved in a short time.

2.6.2.3 Evaluating the influence of the testing system

Sensitivity and specificity for the detection of black tea was similar for the testing platform and the scent detection board. Obviously, the two testing systems chosen did not influence the results in these detection tasks. To our knowledge, this is the first study comparing scent detection capabilities in dogs considering two different testing systems with and without distracting odors. Different testing systems have been used to test scent detection ability of dogs. Due to a variety of other variables (e.g. study design, type of sample matrix, dog trainer) potentially influencing the results it is not possible to directly compare the systems described. Our data confirm a previous study on melanoma detection utilizing three different testing systems (Pickel et al., 2004). The authors used a linear scent line-up with one tumor sample and nine negative control samples in boxes, a test with healthy volunteers with artificially planted tumor samples and negative control samples, and a test with patients in which the dogs had to locate real melanomas between healthy skin areas as negative control samples, both areas covered by bandages. Their results also indicate that there is only little influence of the testing system even though a direct comparison was not conducted. Also only two dogs were utilized. Further research is warranted to confirm our results with more demanding scent detection tasks.

The use of dogs as a diagnostic tool for medical and biological applications is becoming increasingly important. Scent detection dogs promise to become adequate alternative to technical diagnostic tools. It is mandatory, however, to develop standardized test characteristics for scent detection dogs to allow valid comparisons of results.

2.7 Acknowledgements

We wish to thank ROYAL CANIN Tiernahrung GmbH & Co. KG, Cologne, Germany, for financial support and providing of EDUC food treats. Thanks to Kim Salomonsson and Robert Fagerberg from Nordiska Hund training center, Kälarne, Sweden, for supporting the dog training process and the experiment.

2.8 References

Arlt, S., Dicty, V., Heuwieser, W., 2010. Evidence-based medicine in canine reproduction: quality of current available literature. Reprod. Domest. Anim. 45, 1052-1058.

Arner, L.D., Johnson, G.R., Skovronek, H.S., 1986. Delineating toxic areas by canine olfaction. J. Hazard. Mater. 13, 375-381.

Brooks, S.E., Oi, F.M., Koehler, P.G., 2003. Ability of canine termite detectors to locate live termites and discriminate them from non-termite material. J. Econ. Entomol. 96, 1259-1266.

Browne, C., Stafford, K., Fordham, R., 2006. The use of scent-detection dogs. Ir. Vet. J. 59, 97-102.

Cablk, M.E., Heaton, J.S., 2006. Accuracy and reliability of dogs in surveying for desert tortoise (Gopherus agassizii). Ecol. Appl. 16, 1926-1935.

Cornu, J.N., Cancel-Tassin, G., Ondet, V., Girardet, C., Cussenot, O., 2011. Olfactory detection of prostate cancer by dogs sniffing urine: a step forward in early diagnosis. Eur.Urol. 59, 197-201.

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

Fischer-Tenhagen, C., Wetterholm, L., Tenhagen, B.-A., Heuwieser, W., 2011. Training dogs on a scent platform for oestrus detection in cows. Appl. Anim. Behav. Sci. 131, 63-70.

Fjellanger, R., Andersen, E.K., McLean, I.G., 2002. A Training Program for Filter-Search Mine Detection Dogs. Int. J. Comp. Psychol. 15, 278-287.

Gazit, I., Terkel, J., 2003. Explosives detection by sniffer dogs following strenuous physical activity. Appl. Anim. Behav. Sci. 81, 149-161.

Gordon, R.T., Schatz, C.B., Myers, L.J., Kosty, M., Gonczy, C., Kroener, J., Tran, M., Kurtzhals, P., Heath, S., Koziol, J.A., Arthur, N., Gabriel, M., Hemping, J., Hemping, G., Nesbitt, S., Tucker-Clark, L., Zaayer, J., 2008. The use of canines in the detection of human cancers. J. Altern. Complement. Med. 14, 61-67.

Göth, A., Mc Lean, I.G., Trevelyan, J., 2003. How do dogs detect landmines? Mine Detection Dogs 18, 195-208.

Gsell, A., Innes, J., Monchy, P.d., Brunton, D., 2010. The success of using trained dogs to locate sparse rodents in pest-free sanctuaries. Wildl. Res. 37, 39-46.

Hawk, H.W., Conley, H.H., Kiddy, C.A., 1984. Estrus-Related Odors in Milk Detected by Trained Dogs. J. Dairy Sci. 67, 392-397.

Hepper, P.G., Wells, D.L., 2005. How many footsteps do dogs need to determine the direction of an odour trail? Chem. Senses 30, 291-298.
Horvath, G., Jarverud, G.A., Jarverud, S., Horvath, I., 2008. Human ovarian carcinomas detected by specific odor. Integr. Cancer Ther. 7, 76-80.
Kauhanen, E., Harri, M., Nevalainen, A., Nevalainen, T., 2002. Validity of detection of microbial growth in buildings by trained dogs. Environ. Int. 28, 153-157.
Kerley, L.L., 2010. Using dogs for tiger conservation and research. Integr. Zool. 5, 390-396.
Kiddy, C.A., Mitchell, D.S., Bolt, D.J., Hawk, H.W., 1978. Detection of estrus-related odors in cows by trained dogs. Biol. Reprod. 19, 389-395.
Kiddy, C.A., Mitchell, D.S., Hawk, H.W., 1984. Estrus-Related Odors in Body Fluids of Dairy Cows. J. Dairy Sci. 67, 388-391.
Krestel, D., Passe, D., Smith, J.C., Jonsson, L., 1984. Behavioral determination of olfactory thresholds to amyl acetate in dogs. Neurosci. Biobehav. Rev. 8, 169-174.
Lin, H.M., Chi, W.L., Lin, C.C., Tseng, Y.C., Chen, W.T., Kung, Y.L., Lien, Y.Y., Chen, Y.Y., 2011. Fire ant-detecting canines: a complementary method in detecting red imported fire ants. J. Econ. Entomol. 104, 225-231.
Lit, L., Schweitzer, J.B., Oberbauer, A.M., 2011. Handler beliefs affect scent detection dog outcomes. Anim. Cogn. 14, 387-394.
McCulloch, M., Jezierski, T., Broffman, M., Hubbard, A., Turner, K., Janecki, T., 2006. Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. Integr. Cancer Ther. 5, 30-39.
Moll, A., 1904. Mitteilungen über den "klugen Hans". Z. Pädagog. Psychol. 3, 369.
Moser, E., McCulloch, M., 2010. Canine scent detection of human cancers: a review of methods and accuracy. J. Vet. Behav.: Clin. Appl. Res. 5, 145-152.
Paula, J., Leal, M.C., Silva, M.J., Mascarenhas, R., Costa, H., Mascarenhas, M., 2011. Dogs as a tool to improve bird-strike mortality estimates at wind farms. J. Nat. Conserv. 19, 202-208.
Pickel, D., Manucy, G.P., Walker, D.B., Hall, S.B., Walker, J.C., 2004. Evidence for canine olfactory detection of melanoma. Appl. Anim. Behav. Sci. 89, 107-116.
Reynolds, S.D., Christian, D.P., Reynolds, K.G., 2008. Illicit discharge detection is going to the dogs. American Society of Civil Engineers, Reston.
Richards, K.M., Cotton, S.J., Sandeman, Richard M., 2008. The use of detector dogs in the diagnosis of nematode infections in sheep feces. J. Vet. Behav.: Clin. Appl. Res. 3, 25-31.
Savidge, J.A., Stanford, J.W., Reed, R.N., Haddock, G.R., Adams, A.A.Y., 2011. Canine detection of free-ranging brown treesnakes on Guam. N Z. J. Ecol. 35, 174-181.
Sonoda, H., Kohnoe, S., Yamazato, T., Satoh, Y., Morizono, G., Shikata, K., Morita, M., Watanabe, A., Kakeji, Y., Inoue, F., Maehara, Y., 2011. Colorectal cancer screening with odour material by canine scent detection. Gut 60, 814-819.

Walker, D.B., Walker, J.C., Cavnar, P.J., Taylor, J.L., Pickel, D.H., Hall, S.B., Suarez, J.C., 2006. Naturalistic quantification of canine olfactory sensitivity. Appl. Anim. Behav. Sci. 97, 241-254.

Welch, J.B., 1990. A detector dog for screwworms (Diptera: Calliphoridae). J. Econ. Entomol. 83, 1932-1934.

Williams, M., Johnston, J.M., 2002. Training and maintaining the performance of dogs (Canis familiaris) on an increasing number of odor discriminations in a controlled setting. Appl. Anim. Behav. Sci. 78, 55-65.

Willis, C.M., Church, S.M., Guest, C.M., Cook, W.A., McCarthy, N., Bransbury, A.J., Church, M.R., Church, J.C., 2004. Olfactory detection of human bladder cancer by dogs: proof of principle study. Brit. Med. J. 329, 712.

3 PUBLICATION II

How to train a dog to detect cows in heat – Training and success

Dorothea Johnen*, Wolfgang Heuwieser*, Carola Fischer-Tenhagen*

*Clinic for Animal Reproduction, Faculty of Veterinary Medicine, Freie Universität Berlin, Berlin, Germany

Published in:
Applied Animal Behaviour Science, August 2015, Volume 171, Pages 39-46

Please find the original article via the following digital object identifier:
http://dx.doi.org/10.1016/j.applanim.2015.08.019

3.1 Abstract

Efficient and accurate estrus detection is a key management factor determining acceptable reproductive performance in dairy herds. Previous studies have shown that dogs can differentiate between vaginal mucus samples of cows in estrus and vaginal mucus samples of cows in diestrus with an accuracy between 40.3% and 97.0%. We set out 1) to develop a specific training protocol for training dogs to identify cows in estrus from the feed alley and 2) to determine sensitivity and specificity of trained dogs to detect cows in estrus. Six dogs were trained by means of positive reinforcement to detect cows in estrus from the feed alley following the training protocol. Four of those dogs participated in the final test after an average training time of 50 h per dog. Overall, they correctly identified positive cows as being positive in 23 out of 32 cases (i.e. sensitivity of 71.9%) and falsely classified positive cows as being negative in nine cases (28.1% type II errors). Out of 128 cases 119 were correctly classified as true negative cows (i.e. specificity of 93.0%) and in nine cases negative cows were falsely identified as positive cows (7.0% type I errors). Further research is warranted to develop an optimized training protocol that allows training estrus detection dogs for practical use within an appropriate time.

3.2 Keywords

Dog, Canine, Scent detection, Training, Oestrus detection, Dairy Cow

3.3 Introduction

For hundreds of years, dogs have been used for a variety of scent detection tasks. There are numerous scientific publications proofing scent detection abilities of dogs (Browne et al., 2006). Helton (2009) even designates scent detection dogs for a gold standard of detection technology. Dogs are used for multitudinous types of scent detection tasks, e.g. detection of explosives and land mines for police and military (Gazit and Terkel, 2003) or drugs (Nash, 2005). There are scent detection dogs used in wildlife conservation, e.g. by detecting desert tortoises in their natural habit (Cablk and Heaton, 2006), to indicate toxic contamination of the environment (Arner et al., 1986), or to support the elimination of rodents (Gsell et al., 2010). Recently, several applications in human medicine have been described and tested (Bijland et al., 2013). There are different types of tasks scent detection dogs can be trained to perform, i.e. a free search for a target in a defined area as in detection of explosives (Gazit and Terkel, 2003), a scent line-up, that is often used for cancer detection dogs (Moser and McCulloch, 2010) or matching the sample. In some countries, scent matching dogs are accepted to hold up as evidence in court of law (Tomaszewski and Girdwoyn, 2006). Target odours for scent detection dogs often are distinct and chemically known substances, such as explosives or drugs (Gazit and Terkel, 2003). In other cases the chemical composition of the target odour is unknown as in cancer detection (Willis et al., 2004; Johnen et al., 2013) or in identification of individual human scent (Schoon, 1996). In those examples dogs have to find a common denominator of all individual samples that are presented and have to generalize it as the target odour without recourse to the "pure" source of the odour (Willis et al., 2004).

Training dogs for scent detection usually comprises of three steps, i.e. adaptation to the training environment and the training methods, imprinting to the target odour, actual detection or discrimination training (Göth, 2003; Fischer-Tenhagen et al., 2011). During imprinting phase, dogs learn to associate a specific scent with reinforcement. In the actual discrimination training dogs are trained to identify this odour, to differentiate it from distracting odours (Fischer-Tenhagen et al., 2011), and to indicate the target odour by performing a trained indication behavior, often lying down or sitting at the target (Jezierski et al., 2008). Optimizing rewarding time and place has great influence on the training progress (Mackintosh, 1983).

Training methods differ a lot between different studies on canine scent detection. According to Walker et al. (2006), training methods based on positive reinforcement appear to have significant advantages compared to those based on aversive methods. In their study the authors trained dogs to detect n-amylacetate (nAA) by means of positive reinforcement. Their dogs were able to detect nAA up to a 20,000-fold lower threshold compared to a study that utilized electro shocks and water deprivation (Krestel et al., 1984). In a previous study, we showed that most studies concerning scent detection work with dogs did not provide sufficient information about training methods, which made it difficult to assess the evidence level of the studies and to compare the results (Johnen et al., 2013). This confirms Helton (2009), who criticized a lack of information regarding canine training provided in published research. We found that all dog trainers in the 14 studies on training scent detection dogs evaluated in our previous study worked with reward-based training (Johnen et al., 2013) such as play (Horvath et al., 2008; Sonoda et al., 2011), food (Gordon et al., 2008; Fischer-Tenhagen et al., 2011; Lin et al., 2011) or both (Brooks et al., 2003). A clicker was used as a conditioned positive reinforcement in six of the evaluated studies (Willis et al., 2004; McCulloch et al., 2006; Gordon et al., 2008; Richards et al., 2008; Cornu et al., 2011; Fischer-Tenhagen et al., 2011).

Although it is known that trained dogs are able to detect estrus in different types of samples from dairy cows (Kiddy et al., 1978), to our knowledge dogs are currently not used for estrus detection on farm. Efficient and accurate estrus detection is a key management factor determining acceptable reproductive performance in dairy herds (Heersche Jr and Nebel, 1994; At-Taras and Spahr, 2001). Visual observation for behavioral changes such as standing to be mounted, mounting, and increase in activity or physical signs such as clear and viscous vaginal mucus is a common method. Average estrus detection rates on commercial dairy farms, however, are below 60% (Senger, 1994; Becker et al., 2005). Estrus related odours constitute a major source of mammalian chemical communication and are present in vaginal mucus, urine, saliva, faeces, and milk of cows in estrus (Sankar and Archunan, 2004). In several experiments, dogs were trained to distinguish between vaginal mucus samples of cows in estrus and samples of cows in diestrus with an accuracy between 40.3% and 97.0% (Kiddy et al., 1978; Hawk et al., 1984; Jezierski, 1992; Fischer-Tenhagen et al., 2011). Also the ability of trained dogs to detect estrus-associated odours in other materials (i.e. urine, milk, blood plasma and saliva) has been demonstrated (Hawk et al., 1984; Kiddy et al., 1984; Fischer-Tenhagen et al., 2011; Fischer-Tenhagen et al., 2013). Of seven studies on dog training for estrus detection in cows only one study described the dog training procedure including the time needed in detail (Fischer-Tenhagen et al., 2011). Kiddy et al. (1978) trained dogs to detect estrus by direct examination of real cows placed side by side in groups of three in adjacent stalls. Cows were defined as being in estrus when they would stand firm when mounted by another cow. Cows were judged as being in diestrus at days 6 to 12 after estrus was detected. Mean percentage of correct detections was 87.3%. However, sensitivity and specificity were not calculated. Such it could be proven that dogs actually could identify cows in estrus by smell with only 36 cows. In 12 sessions with 16 to 26 replicates per session, the dogs had to detect one out of three cows (i.e. one in estrus and two in diestrus). Individual cows were changed only between sessions. Dogs searched from a position behind the cows. A platform of 46 cm height was situated behind the cows to raise the dogs to a convenient working height.

The objective of this study was to find out if and how it would be able to train dogs that could be used for heat detection under practical conditions on a dairy farm without laborious modifications of the installation. For a practical application it would be necessary to utilize detection dogs for screening eligible cows and to indicate those in estrus. Fischer-Tenhagen et al. (2013) showed that trained dogs were able to detect estrus-specific odours in saliva samples. Dogs passing behind the cows would include risk of injuries, hygienic problems, and a more stressful experience for the cows. Therefore, identifying cows in estrus from the feed alley with the cows fixed in the head locks would be advantageous (Fischer-Tenhagen et al., 2013). Specifically we set out 1) to develop a specific training protocol for training dogs to identify cows in estrus from the feed alley and 2) to determine sensitivity and specificity of trained dogs to detect cows in estrus.

3.4 Material and methods

3.4.1 Dogs

Six privately owned pet dogs were enrolled in this study (*table 1*). Selection of dogs was by convenience. All dogs had basic obedience education. Four of the six dogs had previously participated in a study on estrus detection in cows by odour in the laboratory (Fischer-Tenhagen et al., 2015). In this study, 5 dogs were trained to differentiate natural vaginal fluid from cows in estrus and diestrus, and 5 different dogs were trained to differentiate spray with or without synthetic estrus molecules. Dogs trained on natural fluid and on spray could detect the estrus odour they had been trained on with an overall accuracy of 69.0% and 82.4%, respectively. To validate the synthetic molecules, dogs trained with synthetic molecules had to detect estrus odour in natural vaginal fluid without further training and reached an accuracy of 37.6%. Dogs trained on natural fluid detected the synthetic molecules with an accuracy of 50.0% without further training.

Only dog 6 was used to cows before the training procedure started as it was raised on a farm while the other dogs had no experiences with cows before the training started.

3.4.2 Cows

Cows were kept on a commercial dairy farm in Brandenburg, Germany, milking 200 Holstein-Frisians and housed in a free-stall barn on deep bedded straw. Cows were fed twice a day with totally mixed ratio (TMR). For the training and the testing procedure, ultrasound examination was chosen as a gold standard to classify the cows as positive, i.e. being in estrus, or negative, i.e. being in diestrus or pregnant. For training the dogs, a total of 324 cows were selected as positive when displaying signs of estrus such as standing firm when mounted by another cow and clear and viscous vaginal mucus. Transrectal manual and ultrasound examination (BCF easy scan, BCF Technologies Ltd., Livingstone, Great Britain) showed a highly turgid uterus and a prominent follicle of 1.2 – 2.5 cm but no corpus luteum. In 57 cows estrus was induced by injection of 0.5 mg Cloprostenol (PGF Veyx forte, Veyx-Pharma GmbH, Schwarzenborn, Germany). A total of 641 cows served as negative controls. In these cows a transrectal ultrasound examination revealed a pregnancy (n=416) or a prominent corpus luteum ≥ 1.0 cm on at least one ovary but no prominent follicle (n=225). During the examinations and the training and testing process, cows were fixed in head locks.

3.4.3 Dog training

According to European legislation no part of this study included any insult to any animal. The dog trainer in this study was an experienced animal trainer who had conducted practical training with many species for more than four years. The trainer had contributed to three other scent detection studies with dogs (Johnen et al., 2013; Fischer-Tenhagen et al., 2014; Schallschmidt et al. 2015). Training was conducted on the farm between April 2013 and June 2014 two times per week. Training location was the feed alley of the dairy barn. The feed alley was 3 m wide and bordered by the head locks on one side and a solid wall on the other side. As some of the dogs were distracted by the TMR the entire cow feed was removed from the feed alley before training. Equipment used in training comprised the dogs' normal collars or harnesses and a leash of 5 m length. The leash was used for safety reasons e.g. to prevent dogs from being hit by agricultural machines or to stop any potential aggressive behaviour towards the cows, which never occurred. The leash, however, was never used to force the dogs to go in any direction or to prevent the dogs from leaving the training area.

Training methods based on positive reinforcement using a clicker as conditioned secondary reinforcement and small food treats (Educ, ROYAL CANIN Tiernahrung GmbH & Co. KG, Cologne, Germany) as reward. All dogs were familiar with the sound of the clicker as a predictor for food. The food was always delivered within 1 m of the positive cow during adaptation and within 30 cm during the following training and testing process. Dogs were trained individually and one after another. Position of the positive cow was changed after every training session. A given positive cow was never used for more than two consecutive training sessions. Individual negative cows were never used for more than three consecutive training sessions.

The specific training protocol is described in *table 2*. In the beginning, dogs were adapted to the training environment and the cows by reinforcing them for responding properly to easy, well trained cues, e.g. “sit” or “down”, and for approaching the cows. When the dogs were able to be on the feed alley close to the cows without trying to increase the distance to the cows or to leave the training area, adaptation was finished. For the imprinting phase, one positive cow was fixed in the head locks with at least eight head locks being empty on both sides. The dogs were reinforced for approaching the cow. This was repeated with several positive cows fixed in different locations in the head locks until the dogs approached the head of the cow voluntarily up to 30 cm without hesitation. In the next step, indication behavior was shaped by delaying the reward stepwise until the dogs were able to stand still for at least 5 s in front of the cow looking at it. When the dogs had learned to walk directly to a positive cow fixed in the head lock and perform the indication behavior for at least 5 s a combined verbal and optical cue was introduced. The cue was the word “Search” in combination with a hand gesture (i.e. pointing down the feed alley in direction of the cows). Afterwards, the actual discrimination training started. In the beginning of this phase, a positive and a negative cow were fixed in the feed locks with at least 2.4 m distance between both cows (i.e. three empty stalls). The dog was directed to the positive cow to perform the trained indication behavior from a position that was closer to the positive cow compared to the negative one, making it more likely that the dog indicated the positive cow. Gradually, the position was changed so that the dog had to pass the negative cow in order to reach the positive cow. When the dog was able to discriminate between one cow in estrus and one cow in diestrus, a second negative cow was introduced (*table 2*). Number of negative cows was raised gradually so that the dog really had to discriminate between positive and negative cows. If a dog made a mistake in any training phase (indicated a negative cow or did not indicate a positive cow), the unwanted behavior was ignored and training was interrupted for about 10 s.

3.4.4 Pretest

After the dogs had reached an accuracy of approximately 80% in training, a first test was conducted in February 2014. As results were not satisfying at all we performed a systematic analysis of error and adapted our training protocol accordingly. This first test was consequently classified as pretest.

Five replicates per dog were conducted on one day. In total, 40 cows were used: five positive ones (i.e. being in estrus) and 35 negative ones (i.e. 18 being pregnant and 17 being in diestrus). For every replicate eight cows (i.e. one positive cow and seven negative cows) were fixed in the head locks, with a distance between cows of at least 120 cm (i.e. one empty stall). All personnel in the barn (i.e. dog trainer and support personnel) was blinded regarding the position of the positive cow which was randomized with the random number function of

Excel (Microsoft Office, Microsoft Cooperation, Redmont, USA) for every replicate. The order of the dogs to perform every replicate was also randomized with the random number function of Excel. The dog trainer went with individual dogs into the barn and directed the dog to identify the positive cow. The starting position of the dog was 2 m away from the first cow. The dogs were allowed to pass all the cows before reaching a decision. It was led to the trainer's discretion to decide how often a dog could pass all cows before she made her final decision .When the dog indicated a cow the trainer rewarded the dog without knowing if the dog was correct or not. This approach was inevitable for the pretest and the final test. Otherwise the trainer would have been unblinded according to the position of the positive cow when repeating the search with the other dogs. Only during the two tests the dogs were reinforced without knowing if they were correct or not. Afterwards the trainer left the barn with the dog and announced the position of the indicated cow to the experimenter who documented the finding without giving feedback to the dog trainer. If the dog had indicated the positive cow correctly a correct positive indication for the positive cow and correct negative classifications for the negative cows were documented. If the indicated cow was negative a false positive indication for the indicated negative cow and a false negative classification for the positive cow were documented. Individual cows were changed every time all five dogs had passed a replicate. Every cow was only used in one replicate.

3.4.5 Final test

In June 2014 the final test was performed. Eight test replicates were performed on two consecutive days (four replicates per day). In total, 40 cows were used: 8 positive ones (i.e. being in estrus) and 32 negative ones (i.e. 17 being pregnant and 15 being in diestrus). In contrast to the pretest only five cows (i.e. one positive cow and four negative cows) for every replicate were fixed in the head locker with a distance in between cows of at least 120 cm (i.e. one empty stall). Testing procedure was performed as described above.

3.4.6 Statistical methods

Data were entered into Excel spread sheets (Version 2010, Microsoft, Redmond WA, United States) and statistical analyses were performed with IBM SPSS Statistics for Windows (Version 20.0, IBM Deutschland GmbH, Ehningen, Germany). Level of significance was set at $p \leq 0.05$ for all statistical tests.

Effect of individual dogs on the results of the final test and effect of pregnancy status of negative cows (i.e. being pregnant or in diestrus) were evaluated using Kruskal–Wallis-test.

Table 1: Description of scent detection dogs (n=6) used considering breed, age, sex, previous estrus detection training, completion of training program and diagnostic performance.

Dog	Breed	Sex	Reproductive status	Age in years	Previous estrus detection training	Completion of training program	Diagnostic performance
Dog 1	Labrador Retriever	Female	Spayed	3	Yes	Yes	Sensitivity 87.5% Specificity 96.9%
Dog 2	Irish Terrier cross-breed	Female	Spayed	6	Yes	Yes	Sensitivity 62.5% Specificity 90.6%
Dog 3	Labrador cross-breed	Female	Spayed	5	Yes	Yes	Sensitivity 75% Specificity 93.8%
Dog 4	Hound cross-breed	Male	Neutered	3	Yes	No	- -
Dog 5	Golden Retriever	Male	Neutered	6	No	Yes	Sensitivity 62.5% Specificity 90.6%
Dog 6	Entlebuch mountain dog cross-breed	Female	Intact	4	No	No	- -

Table 2: Specific training protocol for training dogs to detect dairy cows in estrus from the feed alley.

Step	Activity
Phase 1: Adaptation to the training environment and the cows	
1	Start working in a distance to the cows where the dog is able to perform easy, well trained behaviours on cue (e.g. “sit” or “down”) without trying to increase the distance. Reinforce the dog for performing those behaviours. Proceed with step 2 when the dog does not show signs of stress or discomfort.
2	Diminish the distance to the cows in small steps. Proceed with phase 2 when the dog does not show signs of stress or discomfort and approaches the cows up to 1 m distance.
Phase 2: Imprinting to estrus specific cow scents and training of an indication behavior	
3	Fix a single positive (estrus) cow in the feed alley with at least eight head locks being empty on both sides. Reinforce the dog for approaching the cow. Repeat with different cows on different feed places. Proceed with step 4 when the dog approaches the cow voluntarily up to 30 cm without hesitation.
4	Train the dog to stand still in front of the cow. Proceed with step 5 when the dog is able to stand still for at least 5 s looking at it. Deliver the food treat within 30 cm to the positive cow.
5	Introduce a signal like “Search”.
Phase 3: Actual discrimination training	
6	Add a negative cow (i.e. diestrus or pregnant) at a distance of at least 2.4 m (i.e. three empty stalls) from the positive one. Make sure that the negative cow is further away from the dog than the positive cow, so that the dog is more likely to reach the positive one first. Click as soon as the dog stops in front of the positive cow. Deliver the food treat within 30 cm to the positive cow. Repeat 10 times with different cows before proceeding with step 7. Switch randomly from left to right.
7	Same as step 6, however the dog should hold the indication for 5 seconds. Proceed with step 8 when the dog is successful 8 times out of 10.
8	Present both cows side by side in random order with at least 120 cm space (i.e. one empty stall) between them. Click quickly as soon as the dog stops in front of the positive cow. Proceed with step 9 when the dog is successful 8 times out of 10. Use different cows.
9	Same as step 8, however, the dog should be able to extend the indication to 5 seconds. Use the cue introduced in step 5. Proceed with step 10 when the dog is successful 8 times out of 10.
10	Same as step 9, but the dog handler is blinded for the position of the positive cow. An assistant confirms the location of the positive cow before reinforcement. Proceed with step 11 when the dog is successful 8 times out of 10.

11	Add a second negative cow. Present all three cows side by side in random order with at least 120 cm space (i.e. one empty stall) between them. Repeat steps 8-10, but with three cows. Use different cows in each replicate. Proceed with step 12 when the dog is successful 8 times out of 10.
12	Increase the number of cows up to five (one positive and four negative cows) repeating steps 8-10 for every new quantity of cows.

3.5. Results

3.5.1 Dog training

Only four of the six dogs completed training and participated in the pretest and final test. Dog 4 was excluded from training because it showed severe signs of distress and discomfort in the presence of the cows. It refused to approach the cows closer than 2 m and did not finish the adaptation phase. Dog 6 was excluded because of insufficient reliability, probably due to high distractibility by environmental noises.

Training was conducted over a period of 15 months with two training days per week. On every training day, every dog passed between two to three training sessions of 10 min each. An average number of 300 training sessions per dog were conducted, totaling an average training time of 50 h per dog.

3.5.2 Pretest

Three of the four dogs finishing the total training procedure participated in the pretest procedure: dog 1, dog 2 and dog 3. Dog 5 unfortunately was not available at the day of the pretest due to organizational reasons of its owner. The dogs had to distinguish between one positive and seven negative cows in each replicate. Every dog conducted five replicates with a total of 40 cows resulting in 120 distinctions (three dogs, 40 distinctions). Of those, five cows were positive (i.e. in estrus) and 35 were negative, (i.e. in diestrus or pregnant). Overall, the three dogs correctly identified positive cows as being positive in five out of 15 cases (i.e. sensitivity of 33.3%) and falsely classified positive cows as being negative in 10 out of 15 cases (66.7% type II errors). Out of 105 cases 92 were correctly classified as true negative cows (i.e. specificity of 87.6%) and in 13 cases negative cows were identified as false positive cows (12.4% type I errors).

3.5.3 Final test

Four dogs participated in the final test (*table 1*). The dogs had to distinguish between one positive and four negative cows in every replicate. In total, every dog had to conduct eight replicates checking a total of 40 cows (i.e. total of 160 decisions). Of those, eight were positive (i.e. in estrus) and 32 negative (i.e. in diestrus or pregnant). Overall, the four dogs correctly identified positive cows as being positive in 23 out of 32 cases (i.e. sensitivity of 71.9%) and falsely classified positive cows as being negative in nine cases (28.1% type II errors). Out of 128 cases 119 were correctly classified as true negative cows (i.e. specificity of 93.0%) and in nine cases negative cows were falsely identified as positive cows (7.0% type I errors). The individual dog had no effect on the test result ($p = 0.742$) nor had the pregnancy status of negative cows, i.e. being pregnant or in diestrus ($p = 0.880$).

3.6 Discussion

3.6.1. Dog training

Overall training time for the dogs in our study was 15 months, with 300 sessions of 10 min each per dog. Kiddy et al. (1978) described the training time needed for training dogs to discriminate between vaginal mucus samples of cows in estrus and cows in diestrus with several weeks. There was no specification of eventual extra training time for the on farm work. In previous studies, training time for dogs to perform a specific scent discrimination task varied widely from two to three weeks (McCulloch et al., 2006), six months (Richards et al., 2008) to 16 months (Cornu et al., 2011). In some studies, training time was not specified at all (Willis et al., 2004; Gordon et al., 2008; Ehmann et al., 2012). In a previous study we speculated that the level of difficulty of the scent detection task may influence the training time (Johnen et al., 2013). We only found few studies on canine scent detection that specified the training time, i.e. frequency, mean duration of training sessions, total training sessions and total training time per dog (Fischer-Tenhagen et al., 2011). Due to the lacking information it is difficult to directly compare our training time with those of other studies. Nevertheless, the time we needed for training the dogs for our study was longer than the training time specified in most other studies. Adaptation of the dogs to the training environment and the cows was one of the biggest challenges in our training process and was highly time-consuming (see below).

The decision when to end the training phase and to start testing was also not defined in other studies on canine scent detection training. Several studies on canine scent discrimination did not describe training protocols in detail or lacked important information about the procedure implemented (Johnen et al., 2013).

3.6.2 Challenges in training

In the training process it became obvious, that the training concept had to be adapted in various parts. Working on the farm and approaching the cows' heads was a challenge for the dogs. All dogs showed signs of acute stress when they first approached the cows, such as body shaking, crouching, oral behaviours (i.e. brief extension of the tongue tip, snout licking, swallowing, smacking), yawning, restlessness and a low posture (Beerda et al., 1998). This confirms the findings of Kiddy et al. (1978), who reported that two of six dogs were reluctant to approach cattle and therefore had to be excluded. In our study those signs of stress diminished during the adaptation phase, but did not completely disappear. All dogs except one were able to approach the cows up to 30 cm distance when entering phase 2, but they still showed subtle signs of stress, i.e. oral behaviours and a low posture. The dogs tended to increase the distance to the cows when those made sudden movements or noises, which was particularly true for dog 3. In two cases (e.g. one dog refused to enter the barn) we had to go re-enter adaptation for two and three training sessions, respectively.

The feed alley was only equipped with head locks for 40 cows, the rest of the bunk space was bordered by a neck rail. Cows were fixed in the head locks only for procedures like artificial inseminations, rectal or vaginal examinations or injections presumably being aversive. Therefore, locking up cows was sometimes difficult and changing or re-arranging cows became a challenge. In addition we wanted to avoid being training and testing a stressful experience for the cows by chasing them or using aversive methods. So we often had to accept the order chosen by the cows and were not able to re-arrange them according to a random plan. This confirms the experiences of Kiddy et al. (1978) who described changing positions of cows as laborious and time consuming. The authors therefore changed positions only after every sixth to eighth trial. As it was not possible for us to change position of cows after every training session, the dogs sometimes started to learn the position of the positive cow and indicated a specific position rather than a specific cow. To reduce this problem we changed at least the position of the positive cow after every training session. We never used an individual positive cow for more than two consecutive training sessions on a given day. As we only conducted two to three training sessions per day and used different individual cows on a given day we never performed more than two to three training sessions with the same individual negative cows. Planning training days was limited by the availability of cows in estrus and sometimes training had to be cancelled altogether or restricted to one training session per dog. .

Dogs are able to notice hidden clues (i.e. unconscious reactions of the trainer), known as the “Clever Hans” effect (Moll, 1904). It has been demonstrated that handler beliefs of the location of the target substrate can influence the outcome of scent detection work (Lit et al., 2011). Therefore, testing without blinding the handler potentially confounds results (Johnen et al., 2013). To assure that the dogs do not respond to unconscious reactions of the trainer during the training protocol but really focus on the scent detection task under ideal conditions the trainer should be blinded as early as possible in the training procedure, as specified in *table 2*. Unfortunately, in our study, the trainer only occasionally had a support person. Therefore it was not possible to consistently train in a blinded way. This can be considered as a flaw in our experimental design and possibly led to confusion of the dogs during the pretest and the final test, as those procedures were conducted in a double-blind manner. We speculate the lacking adaptation of the dogs to the work with the trainer being blinded as a possible reason for the poor results of the pretest. But as usually number of employees in agriculture is limited, too, this shortage of support persons is reflecting real life conditions.

To optimize training progress in scent discrimination, minimizing the environmental distraction is advantageous (Fischer-Tenhagen et al., 2011). A certain level of distraction is unavoidable on dairy farms such as employees passing the training area, machinery, feed and other animals. Therefore the dogs had to learn to deal these distractions. For the adaptation process it would have been beneficial to start adaptation to the cows without other external distractions. Exclusion of distractions was not possible as the daily routine of the farm could not be interrupted. We tried, however, to minimize the TMR mediated distraction by removing the TMR from the feed bunk. This limited our training time as dairy cows need constant access to food in order to assure a certain dry matter intake. Small TMR particles, however, remained in the bunk and on the concrete of the feeding alley and challenged the dogs' concentration on the scent detection task. We noticed that two dogs (i.e. dogs 1 and 5) were more interested in the remaining feed particles the more signs of stress they showed during training. Therefore we interpreted this as a displacement behaviour. In addition, there was always distraction by noises generated by farm machinery and the cows themselves. Especially dog 3 was sensitive towards sudden noises and sometimes reacted fearful. It is well known that more severe or prolonged stress has deleterious effects upon broad aspects of cognition (McEwen and Sapolsky, 1995) and can slow down learning performance.

Another limitation of our study was the selection process of the dogs. We did not have the opportunity to select dogs according to their reaction towards cows or their motivation to work. From the six dogs that started training procedure, two had to be excluded as it turned out that they were not suitable for this kind of task. A selection process during training with following exclusion of unqualified dogs is described in various studies on scent detection dog training (Kiddy et al., 1978; Kiddy and Mitchell, 1981; Richards et al., 2008; Walczak et al., 2012). In some studies concerning scent detection training play was used as reward instead of food (Horvath et al., 2008; Sonoda et al., 2011). We speculate that this kind of reinforcement would not be suitable for estrus detection dogs, as fast movements of playing dogs could scare the cows. We conclude that training success does not only depend on the chosen training procedure and the trainer but also on the individual dog.

Odour molecules released from objects are distributed through the environment by flowing fluid such as water or air (Willis et al., 2008). Most flows are turbulent and a combination of turbulent flow and evaporating chemicals generates an intermittent odour plume (Vickers, 2000; Willis et al., 2008). As the feed alley was open at both sides there was always a certain air movement in front of the cows. We speculate that this air movement led to a turbulent flow of released odours making it harder for the dogs to correctly locate the odour source. In consequence, we decided to always leave one stall in the head locks between two cows unoccupied to make it easier for the dogs to detect the positive cow. Furthermore, we cannot rule out that in some instances there were other positive cows standing next to the training area emitting additional odour that possibly complicated the task for the dogs.

3.6.3 Pretest

After the dogs had reached an accuracy of approximately 80% in training, the pretest was performed. Results of this test were poor. Probably this finding can be attributed to differences of the testing and training setting. The dog trainer was not only completely blinded regarding the position of the positive cow but also had to reinforce the dogs without knowing whether the indication was correct or not. It has been reported that the dog handlers' intrinsic state (e.g. stress) may influence dogs' performance (Jezierski et al., 2014). When the handlers knew that trials were certification trials, the dogs made more false alerts and fewer correct indications and searching time was longer. As observation agreed with our experiences from previous studies, we decided to classify this first testing as a pretest exposing potential flaws of previous training. As results of the pretest were not satisfying we spent additional four months of training before the final test was conducted. In this extra training time we concentrated on training in a double-blind manner. Most importantly the testing situation has to be an integral part of the training routine in order to familiarize both the trainer and the dogs to the testing procedure. In addition we lowered the number of cows for the final test to only five cows (i.e. one positive and four negatives) instead of eight (i.e. one positive and seven negatives) per replicate. In the pretest we had the impression that dogs were overburdened by checking a number of eight cows per replicate, which was very time consuming and obviously challenged the dogs' concentrativeness. In previous studies on estrus detection dogs had to check either two or three (Kiddy et al., 1978) or four samples (Fischer-Tenhagen et al., 2011, Fischer-Tenhagen et al., 2013, Fischer-Tenhagen et al., 2015). As every new quantity of negative cows had to be trained separately as described in *table 2*, training the dogs to perform the final test with eight cows reaching comparable results would have occupied much more time. With a number of five cows per replicate we were above the sample size of the mentioned studies. Therefore we assumed that a quantity of five cows per replicate with a total number of eight replicates per dog was adequate to confirm our hypotheses. With a lower quantity of cows per replicate the probability for the dogs to find the positive one by just guessing increased. To compensate this we decided to perform more replicates per dog (i.e. eight instead of four) maintaining the total number of 40 cows to check for every dog.

3.6.4 Final test

Our data show that dogs are able to detect estrus cow side under practical conditions with an overall sensitivity and specificity of 71.9% and 93.0%, respectively.

Our results are lower than those of Kiddy et al. (1978), whose dogs averaged 87.3% correct detection of cows in estrus. Dog 1 reached the best results in our study with a sensitivity of 87.5% and a specificity of 96.9%, which corresponds to the previous study (1978). In the study of Kiddy et al. (1978), with every group of three cows (i.e. one in estrus, two in diestrus) 16 to 26 replicates were performed. Position of cows was randomly changed after sixth to eighth replicates. Therefore the dogs did not have to adapt to new individuals of cows in every replicate as in our study. In the study of Kiddy et al. (1978) dogs were searching from a position behind the cows walking on a platform, leading to a completely different experimental design than in our study where dogs searched from the feed alley in a more realistic scenario. Unfortunately Kiddy et al. (1978) did only describe that techniques of differential operant reward conditioning were used and that dogs were reinforced with praise and food for correctly detecting and responding to the estrus sample. A detailed reporting of a training plan or protocol including specific rewarding strategies is missing. Thus, we cannot compare the training approaches of the two studies.

A limitation in our study design was that only one trainer had to work with all four dogs one after another in every replicate. Therefore, it can be argued that the trainer was only completely unprejudiced when working with the first dog in every replicate. When the first dog had indicated a cow as being positive the trainer could be considered as biased for the next dog. We minimized this bias by randomizing the order in which the dogs worked. In addition, the trainer did not receive any feedback until the end of the testing and she therefore did not know if the indications were correct or not. Schoon (1996) mentioned the motivation of the dog handler as potential source of biases. When the handlers saw their dogs performing well in the first trial they felt more confident and relaxed in the next trial. We assume that this interaction also works the other way round and that a poor performance of the dogs can result in less confidence and more stress. This was another reason why the experimenter did not provide any feedback during testing.

Estrus detection efficiency in dairy cattle by visual observation varies from 90% to less than 50% (Roelofs et al., 2010). Utilizing trained dogs for estrus detection that have an overall sensitivity and specificity of 71.9% and 93.0%, respectively, could be an effective approach to improve reproductive performance in dairy farms and would be a good alternative to established estrus detection systems.

Further research, however, is warranted to determine if it is possible to use dogs for estrus detection on farms on a long term basis. This would require that dogs have the endurance and motivation to check cows on a daily basis for years without any loss of performance. Based on our experience and those of others, it remains highly speculative if such a task can be accomplished without high expenditure of time for constant re-training. For using trained dogs as a heat detection tool on dairy farms, those dogs would have to be able to identify individual positive cows among negatives even if cows would be close together. In addition, those dogs would have to ignore the cow feed as distraction because a removal of feed before the commitment of the dog would be obviously impractical. When used under practical conditions it would be necessary to either reinforce the dogs for indications of positive cows without knowing if the indications are correct or no reinforcement at all. The first possibility holds the risk of reinforcing wrong behaviour which could lead to a decline in reliability over time. A lack of reinforcement can lead to a loss of confidence in the dogs and deterioration in their performance (Willis et al., 2010). A training time of 15 months would be too labour and time consuming for an efficient use. Further research is warranted to develop an optimized training plan that allows training estrus detection dogs for practical use within an appropriate time and to assure a certain sensitivity and specificity in working estrus detection dogs over time. As many farms keep dogs for guarding, implementing a workshop system in order to allow interested farmers to train their own dogs to support on farm estrus detection might be a feasible solution.

3.7 Acknowledgement

We are grateful to ROYAL CANIN Tiernahrung GmbH &Co. KG (Cologne, Germany) for financial support and providing of EDUC food treats. Many thanks to the owner and the staff of the participating farm, without whom this study would not have been possible. We acknowledge the financial support given to Dorothea Johnen by Tiergyn Berlin e.V. (Berlin, Germany).

3.8 References

Arner, L.D., Johnson, G.R., Skovronek, H.S., 1986. Delineating toxic areas by canine olfaction. J. Hazard. Mater. 13, 375-381.

At-Taras, E.E., Spahr, S.L., 2001. Detection and Characterization of Estrus in Dairy Cattle with an Electronic Heatmount Detector and an Electronic Activity Tag. J. Dairy Sci. 84, 792-798.

Becker, F., Kanitz, W., Heuwieser, W., 2005. Advantages and disadvantages of different methods of heat detection in cattle. Züchtungskunde 77, 140-150.

Beerda, B., Schilder, M.B.H., van Hooff, J.A.R.A.M., de Vries, H.W., Mol, J.A., 1998. Behavioural, saliva cortisol and heart rate responses to different types of stimuli in dogs. Appl. Anim. Behav. Sci. 58, 365-381.

Bijland, L.R., Bomers, M.K., Smulders, Y.M., 2013. Smelling the diagnosis: a review on the use of scent in diagnosing disease. Neth. J. Med. 71, 300-307.

Brooks, S.E., Oi, F.M., Koehler, P.G., 2003. Ability of canine termite detectors to locate live termites and discriminate them from non-termite material. J. Econ. Entomol. 96, 1259-1266.

Browne, C., Stafford, K., Fordham, R., 2006. The use of scent-detection dogs. Ir. Vet. J. 59, 97-102.

Cablk, M.E., Heaton, J.S., 2006. Accuracy and reliability of dogs in surveying for desert tortoise (Gopherus agassizii). Ecol. Appl. 16, 1926-1935.

Cornu, J.N., Cancel-Tassin, G., Ondet, V., Girardet, C., Cussenot, O., 2011. Olfactory detection of prostate cancer by dogs sniffing urine: a step forward in early diagnosis. Eur. Urol. 59, 197-201.

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

Fischer-Tenhagen, C., Johnen, D., Le Danvic, C., Gatien, J., Salvetti, P., Tenhagen, B.A., Heuwieser, W., 2015. Validation of Bovine Oestrous-Specific Synthetic Molecules with Trained Scent Dogs; Similarities Between Natural and Synthetic Oestrous Smell. Reprod. Domest. Anim. 50, 7-12.

Fischer-Tenhagen, C., Tenhagen, B.A., Heuwieser, W., 2013. Short communication: Ability of dogs to detect cows in estrus from sniffing saliva samples. J. Dairy Sci. 96, 1081-1084.
Fischer-Tenhagen, C., Wetterholm, L., Tenhagen, B.-A., Heuwieser, W., 2011. Training dogs on a scent platform for oestrus detection in cows. Appl. Anim. Behav. Sci. 131, 63-70.
Gazit, I., Terkel, J., 2003. Explosives detection by sniffer dogs following strenuous physical activity. Appl. Anim. Behav. Sci. 81, 149-161.
Gordon, R.T., Schatz, C.B., Myers, L.J., Kosty, M., Gonczy, C., Kroener, J., Tran, M., Kurtzhals, P., Heath, S., Koziol, J.A., Arthur, N., Gabriel, M., Hemping, J., Hemping, G., Nesbitt, S., Tucker-Clark, L., Zaayer, J., 2008. The use of canines in the detection of human cancers. J. Altern. Complement. Med. 14, 61-67.
Göth, A., Mc Lean, I.G., Trevelyan, J., 2003. How do dogs detect landmines? Mine Detection Dogs 18, 195-208.
Hawk, H.W., Conley, H.H., Kiddy, C.A., 1984. Estrus-Related Odors in Milk Detected by Trained Dogs. J. Dairy Sci. 67, 392-397.
Heersche Jr, G., Nebel, R.L., 1994. Measuring Efficiency and Accuracy of Detection of Estrus. J. Dairy Sci. 77, 2754-2761.
Helton, W.S., 2009. Overview of scent detection work. Issues and opportunities. CRC Press, Boca Raton, FL, 83–97.
Horvath, G., Jarverud, G.A., Jarverud, S., Horvath, I., 2008. Human ovarian carcinomas detected by specific odor. Integr. Cancer Ther. 7, 76-80.
Jezierski, T., 1992. The Effectiveness of Estrus Detection in Cows by a Trained Dog. Anim. Sci. Papers Rep. 10, 57-65.
Jezierski, T., Walczak, M., Górecka, A., 2008. Information-seeking behaviour of sniffer dogs during match-to-sample training in the scent lineup. Polish Psychological Bulletin 39, 71-80.
Jezierski, T., Adamkiewicz, E., Walczak, M., Sobczynska, M., Gorecka-Bruzda, A., Ensminger, J., Papet, E., 2014. Efficacy of drug detection by fully-trained police dogs varies by breed, training level, type of drug and search environment. Forensic Sci. Int. 237, 112-118.
Johnen, D., Heuwieser, W., Fischer-Tenhagen, C., 2013. Canine scent detection—Fact or fiction? Appl. Anim. Behav. Sci. 148, 201-208.
Kiddy, C.A., Mitchell, D.S., 1981. Estrus-Related Odors in Cows: Time of Occurrence. J. Dairy Sci. 64, 267-271.
Kiddy, C.A., Mitchell, D.S., Bolt, D.J., Hawk, H.W., 1978. Detection of estrus-related odors in cows by trained dogs. Biol. Reprod. 19, 389–395.
Kiddy, C.A., Mitchell, D.S., Hawk, H.W., 1984. Estrus-Related Odors in Body Fluids of Dairy Cows. J. Dairy Sci. 67, 388-391.
Krestel, D., Passe, D., Smith, J.C., Jonsson, L., 1984. Behavioral determination of olfactory thresholds to amyl acetate in dogs. Neurosci. Biobehav. Rev. 8, 169-174.

Lin, H.M., Chi, W.L., Lin, C.C., Tseng, Y.C., Chen, W.T., Kung, Y.L., Lien, Y.Y., Chen, Y.Y., 2011. Fire ant-detecting canines: a complementary method in detecting red imported fire ants. J. Econ. Entomol. 104, 225-231.
Lit, L., Schweitzer, J.B., Oberbauer, A.M., 2011. Handler beliefs affect scent detection dog outcomes. Anim. Cogn. 14, 387-394.
Mackintosh, N.J., 1983. Conditioning and Associative Learning. Claredon Press, Oxford.
McCulloch, M., Jezierski, T., Broffman, M., Hubbard, A., Turner, K., Janecki, T., 2006. Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. Integr. Cancer Ther. 5, 30-39.
McEwen, B.S., Sapolsky, R.M., 1995. Stress and cognitive function. Curr. Opin. Neurobiol. 5, 205-216.
Moll, A., 1904. Mitteilungen über den "klugen Hans". Z. für Pädagog. Psychol. 3, 369.
Moser, E., McCulloch, M., 2010. Canine scent detection of human cancers: a review of methods and accuracy. J. Vet. Behav.: Clin. Appl. Res. 5, 145-152.
Nash, M., 2005. Who let the dogs in? The use of Drug Sniffer Dogs in mental health settings. J. Psychiatr. Ment. Health. Nurs. 12, 745-749.
Peralta, O.A., Pearson, R.E., Nebel, R.L., 2005. Comparison of three estrus detection systems during summer in a large commercial dairy herd. Anim. Reprod. Sci. 87, 59-72.
Richards, K.M., Cotton, S.J., Sandeman, Richard M., 2008. The use of detector dogs in the diagnosis of nematode infections in sheep feces. J. Vet. Behav.: Clin. Appl. Res. 3, 25-31.
Roelofs, J., Lopez-Gatius, F., Hunter, R.H., van Eerdenburg, F.J., Hanzen, C., 2010. When is a cow in estrus? Clinical and practical aspects. Theriogenology 74, 327-344.
Sankar, R., Archunan, G., 2004. Flehmen response in bull: role of vaginal mucus and other body fluids of bovine with special reference to estrus. Behav. Process. 67, 81-86.
Schallschmidt, K., Becker, R., Zwaka, H., Menzel, R., Johnen, D., Fischer-Tenhagen, C., Rolff, J., Nehls, I., 2015. In vitro cultured lung cancer cells are not suitable for animal-based breath biomarker detection. J. Breath Res. 9, 027103.
Schoon, G.A.A., 1996. Scent identification lineups by dogs (Canis familiaris): experimental design and forensic application. Appl. Anim. Behav. Sci. 49, 257-267.
Schoon, G.A.A., De Bruin, J.C., 1994. The ability of dogs to recognize and cross-match human odours. Forensic. Sci. Int. 69, 111-118.
Senger, P.L., 1994. The estrus detection problem: new concepts, technologies, and possibilities. J. Dairy Sci. 77, 2745-2753.
Sonoda, H., Kohnoe, S., Yamazato, T., Satoh, Y., Morizono, G., Shikata, K., Morita, M., Watanabe, A., Kakeji, Y., Inoue, F., Maehara, Y., 2011. Colorectal cancer screening with odour material by canine scent detection. Gut 60, 814-819.
Tomaszewski, T., Girdwoyn, P., 2006. Scent identification evidence in jurisdiction (drawing on the example of judicial practice in Poland). Forensic Sci Int 162, 191-195.

Vickers, N., 2000. Mechanisms of animal navigation in odor plumes. Biol. Bull. 198, 203-212.
Walczak, M., Jezierski, T., Górecka-Bruzda, A., Sobczyńska, M., Ensminger, J., 2012. Impact of individual training parameters and manner of taking breath odor samples on the reliability of canines as cancer screeners. J. Vet. Behav.: Clin. Appl. Res. 7, 283-294.
Walker, D.B., Walker, J.C., Cavnar, P.J., Taylor, J.L., Pickel, D.H., Hall, S.B., Suarez, J.C., 2006. Naturalistic quantification of canine olfactory sensitivity. Appl. Anim. Behav. Sci. 97, 241-254.
Wasser, S.K., Smith, H., Madden, L., Marks, N., Vynne, C., 2009. Scent-Matching Dogs Determine Number of Unique Individuals From Scat. J. Wildl. Manag. 73, 1233-1240.
Williams, M., Johnston, J.M., 2002. Training and maintaining the performance of dogs (Canis familiaris) on an increasing number of odor discriminations in a controlled setting. Appl. Anim. Behav. Sci. 78, 55-65.
Willis, C.M., Britton, L.E., Harris, R., Wallace, J., Guest, C.M., 2010. Volatile organic compounds as biomarkers of bladder cancer: Sensitivity and specificity using trained sniffer dogs. Cancer Biomark. 8, 145-153.
Willis, C.M., Church, S.M., Guest, C.M., Cook, W.A., McCarthy, N., Bransbury, A.J., Church, M.R., Church, J.C., 2004. Olfactory detection of human bladder cancer by dogs: proof of principle study. Brit. Med. J. 329, 712.
Willis, M.A., Avondet, J.L., Finnell, A.S., 2008. Effects of altering flow and odor information on plume tracking behavior in walking cockroaches, Periplaneta americana (L.). J. Exp. Biol. 211, 2317-2326.

4 ADDITIONAL UNPUBLISHED WORK

An approach to identify bias in scent detection dog testing

Dorothea Johnen*, Wolfgang Heuwieser*, Carola Fischer-Tenhagen*

*Clinic for Animal Reproduction, Faculty of Veterinary Medicine, Freie Universität Berlin, Berlin, Germany

4.1 Abstract

Dogs are used in a variety of scent detection tasks since hundreds of years. Nevertheless, no accepted quality standards or adequate guidelines for performing scent studies and testing dogs to minimize potential biases have been published yet. Therefore the objective of this systematic review was to evaluate publications on scent dog studies including testing procedures and identify methodological issues that can potentially bias or confound the study outcome. We then critically analyzed how the authors discussed potential flaws. Finally, we were able to condense the information into a best practice standard for scent dog studies.

4.2 Keywords

Dog, Canine scent detection, Dog training, Quality of literature, best practice standard

4.3 Introduction

Dogs are used in a variety of scent detection tasks since hundreds of years. Besides numerous reports of anecdotal evidence on the "amazing scent detection abilities" of trained dogs, there are also numerous scientific publications proofing scent detection abilities of dogs (Browne et al., 2006). Helton (2009) designated scent detection dogs as the gold standard of detection technology. Before we started to conduct scientific studies on scent detection with trained dogs (Fischer-Tenhagen et al., 2011; Fischer-Tenhagen et al., 2013; Johnen et al., 2013; Fischer-Tenhagen et al., 2015; Johnen et al., 2015; Schallschmidt et al., 2015), we searched and reviewed the scientific literature. We hoped to find accepted quality standards for performing scent studies and testing dogs to minimize potential biases. Unfortunately, adequate guidelines have not been published. In a systematic review we showed that precise information on test criteria such as reliability, validity and objectivity of the scent detection performance tests were not reported in most publications (Johnen et al., 2013), which agrees with Furton and Myers (2001). This information is, however, essential to evaluate the quality of the test results and to compare findings of similar studies. Overall, the quality and validity of most of the reports are questionable. Potential biases due to different flaws of training and testing procedures can influence the results of a study and eventually lead to wrong study outcomes. Possible scenarios for misleading study outcomes could be that a study suggests that dogs are not able to detect a substance by its smell because of an insufficient training procedure or, in contrast, that contaminants are detected by the dogs instead of the real target odour.

Therefore the objective of this systematic review was to evaluate publications on scent dog studies including testing procedures and identify methodological issues that can potentially bias or confound the study outcome. Specifically, we first evaluated the reviewed publications using a modified check list for evaluating evidence of clinical studies in veterinary medicine (Haimerl and Heuwieser, 2014). To identify and describe potential confounders such as too small numbers of dogs tested, use of different samples in training and testing or blinding of dog trainer and other personnel we analyzed and compared the methods used in the evaluated publications. We then critically analyzed how the authors discussed potential flaws. Finally, we wanted to condense the information into a best practice standard for scent dog studies.

4.4 Methods of identifying the quality of published literature concerning canine scent detection by considering relevant methodological factors

4.4.1 Literature evaluation

A systematic literature research was conducted on 23th July 2015 using the databases Pubmed (www.pubmed.gov) and CAB (http://ovidsp.tx.ovid.com) to find studies on scent detection work in dogs. The subject headings "scent" AND "detection" AND "dogs" were used to find articles written in English language about training dogs for specific scent detection tasks. In a supplementary hand search additional publications were recruited. We excluded duplicates, systematic reviews without original data and papers that did not include systematic testing of canine scent detection performance. The remaining articles were evaluated using a check list detailing relevant information about the study design and the training and testing procedures. Specifically, we recorded quality criteria of the analyzed studies, such as number of dogs, detailed description of dog training, use of new samples in testing, random placement of samples, blinding of the dog trainer and other personnel, presentation of results and whether or not obtained results were critically discussed considering potential sources of biases. Retrieval and management of references was performed with Endnote (Version X4 for Windows, Thomson Reuters, New York, USA).

4.4.2 Relevant quality criteria assessed for further evaluation

All publications included were rated on a 10-Point scale. Studies with a sum of 9-10 points were evaluated as excellent. Studies with a sum of 6-8 points were evaluated as sufficient. Studies with a sum of less than 6 points were evaluated as poor. Specifically, the following quality criteria were assessed.

4.4.2.1 Number of dogs tested

In a previous report based on 14 studies we showed that average number of dogs per study was 4.6 varying between 1 (n=5) and 10 (n=2) dogs (Johnen et al., 2013). Therefore we used a 3-point scale to rate number of dogs utilized (1 = 1 dog; 2 = 2 – 4 dogs; 3 = five or more dogs).

4.4.2.2 Relevant information on training given

For this parameter we assessed whether the training procedure was described in a way that enables the reader to potentially reproduce it. We assigned 1 point for studies that described the training protocol including the rewarding strategy, i.e. use of food, toy or praise. If that information was not available no point was assigned.

4.4.2.3 Use of different samples in training and testing

One point was given if samples used during the testing procedure were completely new to the dogs. A point was not given if the same samples were used in training and testing or no information was available about a potential re-use of samples. For some studies it was obvious that a re-use of samples could be excluded, i.e. trials which employed external, professional dog-handler teams (i.e. already working for law enforcements or commercially) to perform a certain task without previous training. For those studies the point was given even without an explicit reporting of use of different samples in training and testing.

4.4.2.4 Randomization of sample positions

If position of target samples was destined according to a randomization procedure 1 point was given. If there was no information available about decision of sample positions no point was given. For free searching tasks under practical conditions (e.g. for animals in their natural habits or bed bugs in an apartment), the point for randomization was given as well.

4.4.2.5 Blinding of dog trainer and other personnel

Studies conducted in a double-blind way (i.e. dog trainer and all other personnel present during the testing procedure were blinded regarding the position of the target samples) received 2 points. 1 point was assigned for studies with a single-blind design (i.e. only the dog handler blinded, but at least one person present during the test that knew about position of positive samples) or if information about blinding of other personnel was missing. Studies without blinding or without information about blinding received no points.

4.4.2.6 Presentation of results

For discrimination tasks that reported sensitivity and specificity and for free searching tasks that presented accuracy or success rate we assigned 1 point. If results were presented in less clearly defined terms no point was assigned.

4.4.2.7 Critical and objective discussion of results

If results of a given study were critically discussed considering potential sources of biases 1 point was given.

4.5 Results

4.5.1 Literature evaluation

In total, 87 publications were retrieved (CAB 33, Pubmed 32, hand search 22). After excluding doublets (n = 12) and reviews without original data (n = 8), 67 publications remained. Another 10 papers were excluded because they did not describe testing of scent detection work in dogs. As three papers were not obtainable via internet 54 studies were left for final evaluation (Table 1).

Level of quality was assessed as being excellent, sufficient or poor for 11, 29 and 13 papers, respectively (median score was 7). For one study, an evaluation of quality according to the score was not possible. Number of dogs varied between one dog (n = 6) and 164 dogs (n = 1). Average number of dogs per study was 9.2. Relevant information on dog training was given in 31.5% of the studies (n = 17). In 79.6% of the studies (n = 43) different samples were used for training and testing. In the remaining studies (n = 11) samples were re-used or information about a possible re-use was not provided. Sample position was specified as being randomized in 71.7% of the studies (n = 38). For one study an evaluation in this section was not possible. Only 31.2% of all studies (n = 16) were conducted in a double-blind way. In 49.1% of studies (n = 26) only the dog trainer was blinded or information about blinding of other personnel was not available. In 20.8% of the studies, the study design was not blinded or no information was given about blinding (n = 11). For one study evaluation was not possible regarding blinding. In 56.6% of the evaluated studies (n = 30) results were presented as sensitivity and specificity for differentiation tasks or as accuracy or success rate for free search tasks. The remaining studies presented their results in less clearly defined terms (n = 23) whilst for one study no evaluation was possible. In 88.9% of the studies (n = 48) obtained results were critically discussed considering potential sources of biases.

4.5.2 Factors potentially influencing the outcome of scent detection studies with trained dogs

4.5.2.1 Target odour

The target odour is the odour the dogs are trained to find (search for), identify or discriminate. A target odour for scent detection dogs can be a distinct, chemically known substance, such as explosives or drugs (Gazit and Terkel, 2003b). Any odorous object releases molecules according to their vapor pressure. It can vary in composition and concentration. Many odours consist of hundreds of individual chemical components (Hatt, 2004). It is unknown whether dogs search for the complete odour signature of a target substance, or if only one or some components serve as target odor. Lorenzo et al. (2003) assumed that canines often alert to a scent associated with forensic specimens rather than the specimen itself. It has been hypothesized that dogs trained to detect explosives possibly include in their perception of the target odours also the odour of many of the contamination, degradation and other products naturally occuring in TNT (Göth, 2003). It cannot be predicted which components of an odour any particular dog uses to identify the object (Göth, 2003). Dogs that were trained to detect pure potassium chlorate failed to detect potassium chlorate-based explosive mixtures reliably (Lazarowski and Dorman, 2014). This finding highlighted the potential limitations of dogs trained to detect a specific substance by smell to detect and indicate this substance when mixed with other substances. According to Derby et al. (1985) natural chemical stimuli are usually mixtures of chemicals, and responses to mixtures often cannot be predicted even when the response to each component of the mixture is known. Detection of individual substances within a mixture may depend on chemical interactions of different components within the mixture. This can lead to a change of chemical integrity of those substances before the odour contacts receptor cells (Derby et al., 1985; Derby et al., 1996). Cooper et al. (2014) suggested that dogs even can change their picture of the target odour within several trials. In their study a reliable bed bug detection dog appeared to change its target scent profile from live bugs to that of the sachet, the latex gloves worn when placing out the hides, or both.

There are other scent detection tasks where the target odour is varying such as identification of missing people (Fenton, 1992) or individual animals of one species (Cablk, 2008; Savidge et al., 2011). Obviously dogs can learn to identify a discrete set of individuals within a target class and then detect a new individual they have not encountered before (Cablk, 2008) indicating that they are able to generalize. For some scent detection tasks the composition of the target odour is not known at all or it is un-known if there is a typical odour that dogs can detect e.g. training dogs for cancer detection (Willis et al., 2004; Johnen et al., 2013). Trying to train dogs to detect a target substance without known chemical composition of the odour always holds the risk of failing because of a lack of perceptible target odour. Lung cancer cells growing on a culture medium are not suitable for canine olfactory detection as they do not emit any odour perceivable for trained dogs (Schallschmidt et al., 2015). It is unclear which human scent components dogs use for their discrimination and if each dog uses the same components (Schoon, 1996). In addition it remains unexplained if the dogs' olfactory acuity differs for these different components or why some human scent samples seem to have a special attraction for some dogs (Schoon, 1996). In those cases the dogs have to find a common ground of all individual samples that are presented and have to generalize it as the target scent without recourse to the "pure" source of the odour (Willis et al., 2004). Another important issue is the amount of the target substance. As the weight of explosive devices can vary considerably (i.e. 100 g to more than 100 kg), dogs have to be trained to indicate variable amounts of the target substance (Furton and Myers, 2001). Odours change over time and depending on external influences. In a study with only two dogs fingerprints on clean glass surfaces were to be detected. The dogs failed to find human odour traces on glass slides weathered outdoors for three weeks and were beginning to make unreliable responses with one and two weeks old samples. Similar slide material kept indoors was detected up to an age of six weeks (King et al., 1964). As human odour traces change over time and under the influence of weather, dogs should be trained and tested with traces of different age.

Another issue is potential degradation of the target substance (e.g. land mines). These often have been buried under ground for months or years before the dogs have to detect them. This can lead to changes of the odour signature. The rate of this so-called biotransformation depends on humidity, temperature and the presence of micro-organisms (Göth, 2003).

The major chemical components in explosive mixtures have low vapor pressures or limited olfactory receptor responses making them challenging as target odours. Dogs probably detect one of the numerous volatile compounds occurring naturally in the substance or being a current contamination (Furton and Myers, 2001). If training and testing would be conducted with pure substances it could be possible that the dogs would not be able to detect the substance in its natural occurrence under practical conditions.

4.5.2.2 Type of scent detection task and experimental set-up

There are different types of scent detection tasks a dog can be trained to perform such as free search for a target in a defined area, the search along a gradient, a scent line-up and matching the sample. In all types of scent detection tasks the dogs have to indicate the target odour by performing an indication behavior. This is a learned behavior, often lying down or sitting at the target (Jezierski et al., 2008). The requirements for dogs that are trained to perform a free search for practical field deployment include that the dog must be able to independently seek the target odour, cover great distances and differentiate between odours that may be similar to the target odour or share common odour signature elements (Cablk, 2008). The dog has to lead the handler to and then indicate the source of the odour within a large unconstrained search area, identify the target odour irrespective of its concentration, continuously learn and refine its picture of the target odour, and locate targets that are not surface visible (Cablk, 2008). Moreover, dogs search for those substances in highly variable concentrations or under extreme conditions like high temperatures or after strenuous activities (Gazit and Terkel, 2003b). In some cases, free searches are conducted indoors in sometimes constricted rooms with distracting scents such as for drug detection dogs working in psychiatric units of hospitals (Nash, 2005). Testing dogs in free search scenarios is challenging and difficult to standardize. If the same hiding place of the target is chosen for several dogs, some dogs could easily follow the footsteps of the previous dog and handler team instead of searching for the target scent. In a study with 2 dogs brown tree snakes were equipped with radio trackers and released in the search area (Savidge et al., 2011). By implementing this procedure the snakes got the opportunity to hide themselves and a real search for wild snakes was imitated. This procedure approximated searching for wild snakes under practical conditions. However, a direct comparison of the performance of the dogs is not possible as there is a potential bias of different levels of difficulty caused by the variation of hiding places. Also, only 2 dogs were employed not allowing any sound conclusions.

Similarly challenging is the search along an odour gradient. This type of scent detection task is used in the detection of leakages in gas pipelines (Mandal, 2014). To perform this kind of task, the dog has to follow an odour trace to its source as in the free search. The difference is that the dog has to locate the odour source, i.e. the leak, at the place of the biggest odour concentration accurately. To our knowledge there are no published studies on dogs searching along a gradient. Therefore we were not able to identify potential bias for this kind of task.

If dogs are trained to work under laboratory conditions, they usually have to check different samples and differentiate between target and non-target samples. In those scent line-ups, samples with different odours are placed next to each other and the dog is asked to identify the target sample and ignore the non-target samples (Schoon and De Bruin, 1994). Those discrimination tasks are typically performed by cancer detection dogs (Pickel et al., 2004; Willis et al., 2004; McCulloch et al., 2006; Cornu et al., 2011; Sonoda et al., 2011) or as an intermediate step in training dogs to identify a target odour in a free search (Brooks et al., 2003; Lin et al., 2011). Another scent detection task performed under laboratory conditions is matching the sample. This kind of task is usually used in wildlife conservation programs, in which dogs are trained to assign individual identities to a scat sample by matching samples from the same individuals (Wasser et al., 2009; Kerley, 2010). In some countries, results of scent matching dogs are accepted as evidence in court of law (Tomaszewski and Girdwoyn, 2006), where they have to match a scent sample taken at the crime scene in a scent line-up containing a scent sample of the suspicious and scent samples of noninvolved persons (Schoon and De Bruin, 1994; Schoon, 1996; Jezierski et al., 2008).

When dogs are tested under laboratory conditions it is easier to standardize the test procedure. The experimental set-up for discrimination tasks can be arranged in different ways. For discrimination or matching tasks samples can be positioned in a row, in a circle line, or randomly distributed. The different ways of arranging the samples require different statistical approaches. If the number of positive samples in the scent line-ups is defined and identical for every trial, it is important to define whether the dog is allowed or obligated to check every sample before it indicates the target sample or if it has to perform the indication behavior immediately at the first contact with the target sample. If the dog is committed to indicate every target sample at the first contact and must not check samples several times before making a decision, there is a statistical challenge. If the total number of samples is x and the number of target samples is y, the probability of the first sample to be a target sample is y/x. If the first sample is not the target sample, the probability of the second and third sample to be a target sample changes to y/x-1 and y/x-2, respectively. The probability of the next sample to be a target sample increases with every sample the dog classifies as non-target sample and consequently the likelihood of the dog to perform an indication behavior increases with every sample the dog has classified as non-target sample. Such an experimental set-up can lead to incorrect indications of non-target samples at the end of the line if the dog has falsely classified a target sample as non-target sample before. This statistical problem can be eliminated if the number of positive samples in every trial is randomized and/or if there are trials without any positive sample. By such an approach the probability of every sample being a positive one is not related to its position in the scent line-up and the dogs can learn to regard each sample as independent instead of searching for a fixed number of target samples (Willis et al., 2010). If the number of positive samples is set and the dog is allowed to re-check samples, the probability of every sample to be a target sample is constant. For such an approach samples should be arranged circularly as there is no defined beginning or end of the samples. A potential disadvantage of this set-up could be that the dogs start to compare the samples instead of making a decision for every single sample.

The simplest statistical approach of testing scent detection dogs is to train them to check only one sample at a given time which gives every sample a 50% chance of being a target or a non-target sample. This can be done by presenting only one sample and asking for a yes/no decision, meaning one indication behaviour for a target sample (e.g. sitting down in front of the sample) and a different indication behaviour for a non-target sample (e.g. turning away from the sample and back to the handler). It seems to be important to train a distinct indication behavior for non-target samples leading to a reward of equal value as an indication of a target sample. Experiments with rats showed that the animals quickly learned to perform an indication behavior as response to matching scents but had significantly more difficulty not indicating non-matching scents (Lu et al., 1993). Doing nothing might simply be more difficult than performing a learned indication behavior (Schoon, 1996). This confirms the findings of Partyka et al. (2014). In their study dogs showed 10% false-positive responses of negative samples. The authors reasoned the need to train dogs both to indicate the presence of the target odour and to not give an indication in absence of the target odour.
To the best of our knowledge studies have not been published yet in which dogs had been trained to check only one sample at a given time making and indicating a yes/no decision.

Another test design is to have a set of known reference samples (e.g. one positive and three negative samples) and one sample to be tested. A major advantage of such a set-up is the possibility to test the actual working ability of the dog with the reference samples. If the dog indicates the reference samples correctly, the probability that it will decide correctly on the test samples is high. Schoon (1996) described a similar method to ensure the dog's will to work by performing a control trial before the actual trial. This control trial consisted of a match that was considered 'easy' for the dog by his handler. If the dog failed, it was not used for a more difficult match. It was not described, however, how easy and more difficult was defined making it impossible to reproduce the study.

Because of the discrepancies between the different types of scent detection tasks it is not possible to directly compare the outcomes (i.e. sensitivity and specificity or accuracy/success rate) of the different types. Only if the experimental set-up of two scent detection tasks is similar it is possible to directly compare sensitivity and specificity or accuracy/success rate.

4.5.2.3 Samples

The samples used for training and testing of scent detection dogs are one of the core components being able to bias or confound the results of a given task and therefore have to be chosen and handled with uppermost care. It is important to have a sufficient quantity of samples that come from a variety of sources. If training is conducted with a small number of samples, there is the risk that the dog does not generalize the smell of the target odour but performs a search for a specific sample rather than a target search, especially if testing is done with the same samples as in the training procedure (Willis et al., 2010; Elliker et al., 2014). Furthermore, the possibility of a cross contamination of the samples has been described

(Furton and Myers, 2001). Samples used in training and testing can become contaminated by human scent from any person handling the samples, by any substance stored nearby or by the storage container itself. Negative samples can become cross contaminated by positive ones and vice versa. In this case, the dog may start to search for the contaminants rather than the target scent that the handler believes the dog is detecting (Furton and Myers, 2001) biasing the results. Storage of samples can have an enormous influence on their volatile elements (Petri et al., 2008). If positive and negative samples are not handled identically, dogs may pick up these differences instead of the differences of target odour.

Weber et al. (2011) showed that classification accuracy of an electronic nose device to differentiate urine samples of individuals with and without bladder cancer decreased if more individuals with other diseases than cancer were added to the healthy control group. Bodily fluids, tissues and emissions from young, healthy individuals seemed to differ in composition from those of older cancer patients to a greater extent than did samples from age-matched individuals with non-cancerous disease of the same organ (Willis et al., 2010). If sample donors are not matched in respect to age, sex and other malignancies, dogs could detect these differences instead of detecting the target odour. Breath samples collected in hospitals may contain components of "hospital odours" which may be a confounding factor (Walczak et al., 2012). Taking breath samples of cancer hospitalized patients and control samples of healthy individuals outside of hospitals could therefore cause serious confounding. Elliker et al. (2014) hypothesized that finding sufficient case and control samples to allow closer age matching, and also present enough samples to dogs to ensure that they generalize on a disease signature odour, may therefore always be a major limitation of studies of this type. Cornu et al. (2011) evaluated the efficacy of prostate cancer detection by a dog trained on human urine samples. The dog correctly designated the cancer samples in 30 of 33 cases. Of the three negative classifications one patient was re-biopsied and prostate cancer was diagnosed. This shows the importance of using only samples with confirmed accuracy of being target or non-target samples (i.e. use of a gold standard). Use of samples with wrong classification (i.e. using target samples as non-target samples or vice versa) in training can produce bias by confusing the dog about the intended target odour. Use of samples with wrong classification in the testing process can lead to misclassifications of correct indications as false. This shows that wrong classification of samples in training and testing can produce wrong study outcomes either by training the dog to detect the wrong odour or by misclassifying a correct indication on a pretended negative sample as false.

4.5.2.4 Test design

According to Helton (2009) a lack of uniformity in the evaluation of scent detection dog performance and testing conditions leads to problems in assessing the quality of published research. In a systematic review we showed that there are many studies on canine scent detection that do not provide information on blinding of the dog trainer and other personnel concerning the location of the positive sample (Johnen et al., 2013). Blinding of the dog trainer and all other personnel present is highly important. Dogs are able to understand fine changes of the human body language (Cooper et al., 2003; Hare and Tomasello, 2005) and therefore can detect unconsciously given reactions of humans as clues. This phenomenon is known as the “Clever Hans” effect (Moll, 1904) and can lead to serious bias. It has been demonstrated that the dog handler beliefs of the location of the target substrate can affect outcomes of scent detection dog deployments (Lit et al., 2011). In their study 18 handlers of drug and/or explosive detection dogs were falsely told that locations of target samples were marked in a scent line-up. In fact, there was no target sample at all in the experimental setting, meaning that any alerting response was incorrect. In only 15% of the runs the dogs did not show any alerting response whereas in the remaining 85% of the runs there were one or more false positive alerts.

As the influence of the handler has been identified as a possible source of bias (Schoon, 1996), efforts have been made to eliminate the handler from the process. Such an approach did not lead to consistent results, as social reward seemed to be the most effective and the bond between handler and dog therefore inevitable (Schoon, 1996). In a study investigating sniffing behavior of 10 dogs trained to detect pentylacetate the handler stood behind a screen with a one way mirror window at a height which made it possible for the handler to observe the dog without being seen by it (Concha et al., 2014). After every correct indication the dog returned to the handler and was reinforced with food. This experimental design prevented unconscious clues by the handler but allowed the social component of the reward to happen.

A double blind approach, however, can be problematic as well (Willis et al., 2010). As neither the trainer nor support personnel present during testing knew the identity of the samples, dogs were not rewarded for correct indications immediately after each trial. This can cause a loss of confidence in the dogs and deterioration in their performance. Alternatively, dogs could be reinforced for every indication not knowing if it was correct or not. Again this could lead to a decrease in performances as dogs would eventually be reinforced for false responses (Johnen et al., 2015). In most studies that were conducted in a double-blind fashion an experimenter not being present during scent detection confirmed the dogs’ indications before the reward (Cornu et al., 2011; Johnen et al., 2013).

4.5.2.5 Dog trainer and training procedure

Dog training requires special expertise (Bijland et al., 2013) and the implementation of a specific training method. According to Walker et al. (2006) training methods based on positive reinforcement have significant advantages compared to those based on aversive methods. In their study the authors trained dogs to detect n-amylacetate (nAA) by means of positive reinforcement. Their dogs were able to detect nAA up to a 20,000-fold lower threshold compared to a study that utilized electro shocks and water deprivation (Krestel et al., 1984). In previous studies, it has been shown that most studies concerning scent detection work with dogs did not provide sufficient information about training methods, experience, and qualification of dog trainers, which made it difficult to evaluate and compare the results (Helton, 2009; Johnen et al., 2013).

We speculate that the trainer might influence the duration necessary for training and the success rate because of the methods used and his practical experience in training dogs for a given scent detection work (Johnen et al., 2013). Schoon (1996) mentioned changes in motivation of the handler as potential source of biases. When dogs were performing well in a first trial, handlers felt more confident and relaxed in the next trial. We assume that this mechanism also works the other way round. Cooper et al. (2014) highlighted the importance of a dog handler's capability to interpret the dog's behavior to allow high accuracy.

4.5.2.6 Dog breeds

There are various types and breeds of dogs that have been selectively bred for different tasks. Concerning the selection of dogs to train, it seems to be obvious that some types of dogs are more appropriate for scent detection training than others i.e. the working races compared to toy or brachiocephalic races. But in a recent study Hall et al. (2015) showed that contrary to expectations, Pugs significantly outperformed German Shepherds in acquiring an odour discrimination task and maintaining performance when concentration of the target odour was decreased. The authors suggested that the difference found between both breeds may rather be related to behavioral differences, such as differences in sniff patterns, than to differences in olfactory capacities. Special breeds classified as scent hounds were bred for their ability to hunt and track using olfactory cues (Issel-Tarver and Rine, 1996). It was hypothesized that selection on the basis of olfactory ability has primarily had an effect on the total number of olfactory neurons. Thus, scent hounds would appear to be the most suitable breeds for those purposes (Gazit and Terkel, 2003a). However, for many published scent detection studies, a variety of breeds, especially shepherds and other working races were trained (Schoon and De Bruin, 1994; Cornu et al., 2011; Sonoda et al., 2011). Wasser et al. (2009) selected dogs with a highly focused, excessive play-drive because of the assumption that only those dogs were able to sustain the focus and motivation required to make many comparisons needed for the scent detection line-up. In a study of Jezierski et al. (2014) German shepherds proved to be significantly superior to Labrador retrievers and Terriers, whereas no significant difference to Cocker spaniels could be shown. Until now, there is generally a lack of scientific studies comparing the suitability of particular breeds for detection tasks (Jezierski et al., 2014). Further research is warranted as number of dogs is too small in most studies on canine scent detection to give sound scientific evidence concerning selection of different breeds (Johnen et al., 2013).

Table 1: Research articles on training and testing of scent detection dogs eligible for systematic evaluation based on a literature search on date using the search terms "scent" AND "detection" AND "dogs".

Author and year of publication	Score (number of dogs tested)[a]	Relevant information on training given[b]	Use of different samples in training and testing [c]	Randomization of sample positions[d]	Blinding[e]	Presentation of results[f]	Critical discussion of results[g]	Sum (X of 10 points)[h]
Angle et al. (2014)	3 (17)	0	1	1	1	0	1	7
Bomers et al. (2012)	1 (1)	0	1	0 (no inf.)	1	1	1	5
Brooks et al. (2003)	3 (6)	1	0 (no inf.)	0 (no inf.)	0 (no inf.)	0	0	5
Cablk et al. (2008)	2 (2)	0	1	1	1	0	1	6
Concha et al. (2014)	3 (10)	0	1	1	2	0	1	8
Cooper et al. (2014)	3 (11)	0	1	1	1	0	1	7
Cornu et al. (2011)	1 (1)	0	1	0 (no inf.)	2	1	1	6
Curran et al. (2010)	3 (13)	0	1	1	2	1	1	9
Dominguez-Ortega et al. (2013)	2 (2)	0	0	0 (no inf.)	1	1	0	4

Author and year of publication	Score (number of dogs tested)[a]	Relevant information on training given[b]	Use of different samples in training and testing [c]	Randomization of sample positions[d]	Blinding[e]	Presentation of results[f]	Critical discussion of results[g]	Sum (X of 10 points)[h]
Ehmann et al. (2012)	2 (4)	0	1	1	2	1	1	8
Elliker et al. (2014)	3 (10)	1	1	1	2	1	1	10
Fischer-Tenhagen et al. (2015)	3 (9)	1	1	1	2	0	1	9
Fischer-Tenhagen et al. (2013)	3 (12)	1	1	1	2	0	1	9
Fischer-Tenhagen et al. (2011)	3 (7)	1	1	1	1	1	1	9
Garner (2000)	2 (4)	0	1	0 (no inf.)	0 (no inf.)	1	1	5
Gazit et al. (2005)	3 (7)	0	1	0	1	0	1	6
Gazit and Terkel (2003a)	3 (6)	0	0	0	1	0	1	5
Gazit and	3 (6)	0	1	0 (no inf.)	0 (no inf.)	1	1	6

Author and year of publication	Score (number of dogs tested)[a]	Relevant information on training given[b]	Use of different samples in training and testing [c]	Randomization of sample positions[d]	Blinding[e]	Presentation of results[f]	Critical discussion of results[g]	Sum (X of 10 points)[h]
Terkel (2003b)								
Goodwin et al. (2010)	2 (3)	1	1	1	1	1	1	8
Gordon et al. (2008)	3 (10)	0	1	1	1	1	1	8
Gsell et al. (2010)	2 (2)	0	1	1	0 (no inf.)	1	1	6
Hall et al. (2015)	3 (21)	1	0 (no inf.)	1	1	0	1	7
Horvath et al. (2008)	1 (1)	0	1	0 (no inf.)	2	1	1	6
Jezierski et al. (2014)	3 (164)	0	1	0	1	1	1	7
Jezierski et al. (2010)	3 (6)	1	1	1	1	1	1	9
Jezierski (2008)	3 (6)	1	1	1	2	0	1	9
Johnen et al. (2015)	2 (4)	1	1	1	2	1	1	9
Johnen et al.	3 (7)	1	1	1	1	1	1	9

Author and year of publication	Score (number of dogs tested)[a]	Relevant information on training given[b]	Use of different samples in training and testing [c]	Randomization of sample positions[d]	Blinding[e]	Presentation of results[f]	Critical discussion of results[g]	Sum (X of 10 points)[h]
(2013)								
Kauhanen et al. (2002)	2 (2)	0	0 (no inf.)	0 (no inf.)	1	1	1	5
King et al. (1964)	2 (2)	0	0 (no inf.)	1	0 (no inf.)	0	1	4
Lazarowsk et Dorman (2014)	3 (16)	1	0	1	0	0	1	6
Lin et al. (2011)	2 (3)	0	1	1	0 (no inf.)	0	0	5
Lit et al. (2011)	3 (18)	0	1	Eval. not poss.	Eval. not poss.	Eval. not poss.	1	Eval. not poss.
Lit et al. (2006)	3 (23)	0	1	0	2	1	1	8
Lorenzo et al. (2003)	3 (23)	0	1	1	1	0	1	7
McCulloch et al. (2006)	3 (5)	1	1	1	2	1	1	10
Oesterhelweg et al. (2008)	2 (3)	0	1	0 (no inf.)	1	1	0	5
Partyka et al.	2 (2)	0	1	1	2	1	1	8

Author and year of publication	Score (number of dogs tested)[a]	Relevant information on training given[b]	Use of different samples in training and testing [c]	Randomization of sample positions[d]	Blinding[e]	Presentation of results[f]	Critical discussion of results[g]	Sum (X of 10 points)[h]
(2014)								
Paula et al. (2011)	1 (1)	1	1	1	0 (no inf.)	1	1	6
Pickel et al. (2004)	2 (2)	0	0	0	1	0	1	4
Reed et al. (2011)	2 (2)	0	0 (no inf.)	1	0 (no inf.)	1	1	5
Reindl-Thompson et al. (2006)	2 (2)	0	1	1	1	1	1	7
Richards et al. (2008)	1 (1)	0	0 (no inf.)	1	0 (no inf.)	0	0	2
Savidge et al. (2011)	2 (2)	0	1	1	1	1	1	7
Schallschmidt et al. (2015)	2 (2)	0	1	1	1	1	1	7
Schoon (1996)	3 (8)	0	1	1	1	0	1	7
Schoon and De Bruin (1994)	3 (6)	0	1	1	1	0	1	7

Author and year of publication	Score (number of dogs tested)[a]	Relevant information on training given[b]	Use of different samples in training and testing [c]	Randomization of sample positions[d]	Blinding[e]	Presentation of results[f]	Critical discussion of results[g]	Sum (X of 10 points)[h]
Schoon et al. (2014)	3 (5)	1	1	1	2	1	1	10
Sonoda et al. (2011)	1 (1)	0	1	1	1	1	1	6
Vaidyanathan et Feldlaufer (2013)	2 (2)	0	1	0 (no inf.)	1	0	0	4
Walczak et al. (2012)	2 (3)	1	1	1	2	0	1	8
Wasser et al. (2009)	2 (3)	1	0	1	1	0	1	6
Willis et al. (2010)	2 (4)	0	1	1	2	1	1	8
Willis et al. (2004)	3 (6)	0	1	1	0 (no inf.)	0	1	6

No inf. = no information available in the study, evaluation not poss. = evaluation not possible

[a] Number of dogs: 1 dog: 1, 2-4 dogs: 2, ≥5 dogs: 3

[b] Relevant information on training given: yes: 1, no: 0

[c] Use of different samples in training and testing: yes: 1, no/ no information: 0

[d] Sample positions randomized: yes: 1, no/ no information: 0

[e] Blinding: yes: 2 only handler blinded/no information about blinding of other personnel: 1, no blinding/no information: 0

[f] Presentation of results: Sensitivity/specificity (for differentiation tasks) or accuracy/success rate (for free search tasks) calculated: yes: 1, no: 0

[g] Critical and objective discussion of results: yes: 1, no: 0

[h] Level of evidence: excellent: 9-10, sufficient: 6-8, poor: <6

4.6 Discussion and conclusions

4.6.1 Literature evaluation

A systematic analysis of 54 papers on canine scent detection showed considerable differences in methodology and quality. We did not include three papers in our evaluation because they were not obtainable via internet. Probably it would have been possible to get these publications by contacting the authors or using the library lending system. But implementing the concepts of evidence-based veterinary medicine requires information sources providing relevant, valid material that can be accessed quickly and with minimal effort (Smith, 1996; Haimerl and Heuwieser, 2014). Haimerl and Heuwieser (2014) postulated an online availability of every published paper as accessibility or use is not guaranteed otherwise. Therefore we decided to exclude papers not obtainable via internet.

In contrast to our previous literature evaluation using the same subject headings and the same databases (Johnen et al., 2013), an increase in number of studies found with the same methodology (i.e. initial search and hand search) was seen (31 in 2013 versus 87 in 2015). This clearly shows the rising importance for research and in practical applications. Mean number of dogs was 4.6 and 9.2 in 2013 and in the present study, respectively. In 68.5% of the publications no or only minimal information were given regarding the method of dog training. This makes it impossible to comprehensively understand or even reproduce the studies to identify potential flaws or biases in the training protocol. In 20.4% of the publications training samples were potentially re-used for testing. This includes the risk of dogs recognizing individual samples as being positive rather than searching for the target odour which is a potential source of bias. Elliker et al. (2014) suggested that dogs may memorize the individual odours of large numbers of training samples rather than generalize on a common odour.

In 28.3% of the studies position of the positive samples was not randomized. In a previous study we learned that dogs were easily capable of remembering the position of the positive sample and sometimes indicated a specific position rather than a specific sample (Johnen et al., 2015). In 11 of the evaluated studies information about blinding was missing at all or dog handlers were unblinded during testing. This approach holds the risk of unintended cueing (Moll, 1904; Lit et al., 2011). Elliker et al. (2014) demanded rigorous double-blind methods to be used for future studies to avoid confounding. Presentation of results describing diagnostic performance was heterogeneous, as only 27 studies gave results as sensitivity and specificity for differentiation tasks and accuracy or success rate for free searching tasks. Using less clearly defined terms complicates comparison of different studies' results. Cooper et al. (2014) attempted using the total correctness (TC) as an alternate method to measure a dog-handler-team's correctness. TC is the total of correct positive indications plus correct negative indications or non-indications of negative samples divided by the total number of samples inspected. The authors showed, however, that using TC can be misleading and concluded that TC should not be used as a measure of scent detection dog performance.

Most studies discussed their results critically considering potential sources of biases. Interestingly, all studies that did not receive a point for critically discussing the results were evaluated as poor (n= 6). Overall 24.5% (n = 13) publications were evaluated as being poor. Date of publication was between 1964 and 2015. Papers classified as poor had been published between 1964 and 2013. This implicates that suboptimal study design in testing scent detection dogs is still a current topic that needs to be discussed critical, and more emphasis needs to be placed on optimizing methodology. The majority of evaluated papers (n = 40) was evaluated as being either excellent or sufficient regarding relevant quality criteria. Thirty-two of the 54 papers evaluated served as references for the results and the discussion and conclusions sections of this publication. Twenty and 5 were evaluated sufficient and excellent, respectively. We assume that by choosing high quality references biases in results and conclusions drawn in this systematic review were minimized.

4.6.2 Best practice for testing scent detection dogs

Regarding the identified factors potentially biasing results of studies we deduced minimum requirements that should be followed when scent detection dogs are tested. In addition we envisioned perfect conditions for testing scent detection dogs eliminating as much sources of potential bias as possible. Results are summarized in Table 2.

4.6.2.1 Target odour

Exactly defining the target odour is probably one of the most important issues in canine scent detection training and testing. It is important to define whether the target odour is a distinct and stable, chemically known odour or if the target odour is varying. The underlying specific research hypothesis has to be defined as exactly as possible. If the aim of a test is to show the success of a training program for dogs working under practical field conditions, the target odour should vary between the different testing trials within the expected range, considering possible contaminations (Furton and Myers, 2001), aging processes (King et al., 1964), or individual odour components (Cablk, 2008). If the aim of a test is to show that dogs are able to detect a certain substance or compound by smell or to determine the minimum concentration of a certain substance a dog is able to detect by smell, any contamination must be avoided to prevent the dog from detecting anything else than the odour in question (Furton and Myers, 2001). There is always the possibility that dogs are not able to detect a particular odour (Schallschmidt et al., 2015) or a particular concentration. This is a valuable result which should be published the same way as a positive result.

4.6.2.2 Type of scent detection task

Free search testing trials should approximate conditions that are prevalent under practical conditions. As external influences as temperature, humidity, wind speed and terrain have an impact on the performance of a scent detection dog (Garner, 2000), those factors should be adapted to the real conditions and should be standardized between different trials and dogs as far as possible. In a perfect test scenario selection of hiding places is randomly chosen, e.g. by randomly selecting GPS-coordinates within the search area as hiding places or, if animals are targets, by allowing them to hide themselves. If this is not possible, selection of hiding places has to be done on the basis of careful analysis of hiding places under practical field conditions. The procedure of selecting the hiding places has to be described in scientific publications in a way that the study can be reproduced. A search along a gradient should fulfill similar requirements. The lack of scientific based information for this topic does not allow it to formulate sound recommendations.

In scent line-ups and matching the sample tasks, number of samples and ratio of target to non-target samples has to be described. It has to be defined if the dogs are allowed to re-check samples or not and if the dogs are obliged to check every single sample before an indication. Position of target samples has to be randomized. Furton and Myers (2001) demand that every test of scent detection dogs should include negative controls in order to ensure that the dog does not indicate on a non-target sample to get its reward. A study by Gazit et al. (2005) with explosives detection dogs showed that re-searching an area where the dogs had previously searched without finding explosives resulted in a reduction in motivation for the dog leading to fewer detections in subsequent searches in the same area. This shows that negative control searches have to be trained well and should lead to reinforcement as well in order to maintain the dog's motivation.

In a perfect test scenario the dogs check every sample only once making a yes/no decision for every sample. This can be realized by presenting only the questionable sample or by presenting the questionable sample between known reference samples as positive and negative controls. Sequence of target and non-target samples has to be randomized. This approach leads to a consistent probability for every sample to be a target sample or a non-target sample. For such an experimental design more samples have to be tested than in a scent line-up with several unknown samples, as the probability for every single sample being a target sample is 50% and the dog could be successful by simply guessing.

4.6.2.3 Samples

If the dogs have to recognize the target odour out of different other odours (e.g. individual odours of sample donors) it is important to have a sufficient number of target and non-target samples to enable the dogs to generalize the target odour in training without re-using individual samples. To our knowledge there is no statistical procedure to predict in advance how many samples are needed to train a specific scent detection task. Numbers of target and non-target samples used for different kinds of cancer detection training varied considerably between different studies. Cornu et al. (2011) and McCulloch et al. (2006) used a total of 26 and 86 target samples and 16 and 83 non-target samples for training, respectively, whereas

Willis et al. (2010) used several hundred target samples and about 500 non-target samples for training. Based on the studies evaluated and the limited evidence we propose to use at least 100 target and non-target samples, respectively, for training complex scent detection tasks such as cancer detection.

Collection, storage and handling of samples should be conducted to exclude the possibility of a cross-contamination of non-target samples with target odour or of both target and non-target samples with foreign odours (Furton and Myers, 2001). Again, there is no science based description how to collect and handle biological samples for scent detection dogs best. Petri et al. (2008) examined the influence of different urine collection methods, storage temperatures and times, and repetitive freeze-thaw procedures on retention of chemical substances for mass spectrometric proteomic investigations of human urine. Storing urine samples for a maximum of 6 hours at 4°C until freezing at -80°C prior to analysis provided the most stable profiles. No significant differences were demonstrated between native samples and frozen-thawed samples (up to eight freeze-thaw cycles). Heating of urine samples to 37°C might improve the relative proportions of the more volatile target molecules of interest (Willis et al., 2010). Further research concerning how to optimize collection and handling of biological samples for scent detection in dogs is warranted.

In order to prevent adding olfactory clues other than the target odour, target and non-target samples have to be prepared by the same person within a single training or testing session. Preparation should always start with the non-target samples to avoid cross-contamination. Handling all samples with tongs or gloves might help to prevent contamination (Schoon, 1996).

For cases in which dogs have to detect small differences in the odour signature of target and non-target samples (e.g. cancer detection) it is important to have samples confirmed by the best available gold standard (e.g. biopsy or cystoscopy in case of cancer) (Willis et al., 2010).

Testing scent detection dogs should always be done with samples and odour carriers the dog had never had contact to before. It has to be guaranteed that there are no other systematic odour differences between target and non-target samples than the target odour.

4.6.2.4 Test design

Testing scent detection dogs without blinding the handler potentially confounds results and can be considered a fatal flaw in the study design (Johnen et al., 2013). Therefore, testing scent detection dogs should always be conducted in a double-blind fashion (Furton and Myers, 2001). In order to prevent the trainer as well as the dog from being able to guess the position of the target samples, sample position should always be randomized. Jeziersky et al. (2014) reported that in their study the dog handlers' intrinsic state (stress) may have influenced dogs' performance. To reduce this kind of bias the trainer should be positioned in a way that the dog is not able to see him while searching. This can be achieved by either the trainer only bringing the dog into the testing room, send it to search, leave the room when the dog starts searching and observe the dog through a window or via video camera. When the dog performs the trained indication behavior, the location of the sample could be confirmed by the experimenter who should also not be in the testing room and the trainer could re-enter the room and reinforce the dog. As an alternative the trainer can be positioned behind a screen with a one way mirror to enable the handler to see the dog but not to be seen by the dog as long as the dog searches (Concha et al., 2014). With such an experimental design that has to be trained before the testing procedure, the potential bias of the intrinsic state of the dog trainer could be minimized. As an alternative, testing scenarios have to be trained in order to reduce the trainer's stress. For certain free searching tasks, it is obvious that the handler will be stressed under field conditions e.g. in explosive detection or search and rescue. For those tasks, it is impossible to remove the handler from the scenario. Potential bias can only be reduced through constant training of testing scenarios, but not be completely excluded.

4.6.2.5 Dog trainer and training procedure

In a previous study, we speculated that the trainer's skills and previous training experience might influence the duration necessary for training a certain task and the training outcome and proposed further research (Johnen et al., 2013). As a minimum standard, the used training method should be reported in a way that the reader would be able to comprehend and reproduce all procedures. For the perfect scenario, precise information concerning the trainer's skills and previous training experience should be provided. Time needed for dog training should be specified i.e. number of training sessions per dog, mean duration of a session, over-all training time per dog in hours and/or minutes.

4.6.2.6 Type of dogs

Until now, there is a lack of scientific evidence on suitability of particular breeds for detection tasks. Therefore recommendations on a task specific breed cannot be given. Information about breed, age, sex, reproductive status and previous scent detection experience should be provided in every study concerning canine scent detection.

Table 2: Recommendations for reporting studies on canine scent detection tasks.

Identified factors	Minimum requirement	Perfect testing scenario
Target odour	Describe target odour, i.e. distinct, chemically known odour or varying target odour Clear research hypothesis	-
Type of scent detection task	Describe number of samples, ratio of target to non-target samples Define if dogs are allowed to re-check samples Position of target samples has to be randomized Include negative controls Free search: approximate as close as possible to all conditions given in practical application	Yes/no decision for every sample, i.e. presenting only questionable sample or with known reference samples
Samples	Use of new samples and new odour carriers the dog had never had contact to before Storage and handling of samples such that the possibility of a cross-contamination is minimized No other systematic odour differences between target and non-target samples than the target odour Existing gold standard to differentiate between target and non-target samples	-

Identified factors	Minimum requirement	Perfect testing scenario
Blinding	Double-blind	Trainer leaves the testing room or is placed behind a screen during search or Constant training of testing scenarios to reduce potential bias as far as possible
Dog trainer and training procedure	Training method is reported in detail in a way that the reader is able to reproduce it Time needed for dog training provided as accurately as possible	Information concerning the trainer’s skills and previous training experience should be provided
Type of dog	Information about breed, age, sex and previous scent detection experience of trained dogs is given	-

4.7 Acknowledgements

We acknowledge the financial support given to Dorothea Johnen by Tiergyn Berlin e.V. (Berlin, Germany).

4.8 References

Angle, C.T., Wakshlag, J.J., Gillette, R.L., Steury, T., Haney, P., Barrett, J., Fisher, T., 2014. The effects of exercise and diet on olfactory capability in detection dogs. J. Nutr. Sci. 3, e44.

Bijland, L.R., Bomers, M.K., Smulders, Y.M., 2013. Smelling the diagnosis: a review on the use of scent in diagnosing disease. Neth. J. Med. 71, 300-307.

Bomers, M.K., van Agtmael, M.A., Luik, H., van Veen, M.C., Vandenbroucke-Grauls, C.M.J.E., Smulders, Y.M., 2012. Using a dog's superior olfactory sensitivity to identify Clostridium difficile in stools and patients: proof of principle study. Brit. Med J. 345, e7396

Brooks, S.E., Oi, F.M., Koehler, P.G., 2003. Ability of canine termite detectors to locate live termites and discriminate them from non-termite material. J Econ Entomol 96, 1259-1266.

Browne, C., Stafford, K., Fordham, R., 2006. The use of scent-detection dogs. Ir. Vet. J. 59, 97-102.

Cablk, M.E., Sagebiel, J. C., Heaton, J. S., Valentin, C., 2008. Olfaction-based Detection Distance: A Quantitative Analysis of How Far Away Dogs Recognize Tortoise Odor and Follow It to Source. Sensors 8, 2208-2222.

Concha, A., Mills, D.S., Feugier, A., Zulch, H., Guest, C., Harris, R., Pike, T.W., 2014. Using sniffing behavior to differentiate true negative from false negative responses in trained scent-detection dogs. Chem. Senses 39, 749-754.

Cooper, J.J., Ashton, C., Bishop, S., West, R., Mills, D.S., Young, R.J., 2003. Clever hounds: social cognition in the domestic dog (Canis familiaris). Appl. Anim. Behav. Sci. 81, 229-244.

Cooper, R., Wang, C., Singh, N., 2014. Accuracy of Trained Canines for Detecting Bed Bugs (Hemiptera: Cimicidae). J. Econ. Entomol. 6, 2171-2181

Cornu, J.N., Cancel-Tassin, G., Ondet, V., Girardet, C., Cussenot, O., 2011. Olfactory detection of prostate cancer by dogs sniffing urine: a step forward in early diagnosis. Eur. Urol. 59, 197-201.

Curran, A.M., Prada, P.A., Furton, K.G., 2010. Canine human scent identifications with post-blast debris collected from improvised explosive devices. Forensic Sci. Int. 199, 103-108.

Derby, C.D., Ache, B.W., Kennel, E.W., 1985. Mixture suppression in olfaction: electrophysiological evaluation of the contribution of peripheral and central neural components. Chem. Senses 10, 301-316.

Derby, C.D., Hutson, M., Andrew Livermore, B., Lynn, W.H., 1996. Generalization among related complex odorant mixtures and their components: Analysis of olfactory perception in the spiny lobster. Physiol. Behav. 60, 87-95.

Dominguez-Ortega, L., Diaz-Gallego, E., Pozo, F., Cabrera Garcia-Armenter, S., Serrano Comino, M., Dominguez-Sanchez, E., 2013. Narcolepsy and odor: preliminary report. Semergen 39, e41-46, 348-353.

Ehmann, R., Boedeker, E., Friedrich, U., Sagert, J., Dippon, J., Friedel, G., Walles, T., 2012. Canine scent detection in the diagnosis of lung cancer: revisiting a puzzling phenomenon. Eur. Respir. J 39, 669-676.
Elliker, K.R., Sommerville, B.A., Broom, D.M., Neal, D.E., Armstrong, S., Williams, H.C., 2014. Key considerations for the experimental training and evaluation of cancer odour detection dogs: lessons learnt from a double-blind, controlled trial of prostate cancer detection. BMC Urol. 14, 22.
Fenton, V., 1992. The use of dogs in search, rescue and recovery. J. Wilderness Med. 3, 292-300.
Fischer-Tenhagen, C., Johnen, D., Le Danvic, C., Gatien, J., Salvetti, P., Tenhagen, B.A., Heuwieser, W., 2015. Validation of Bovine Oestrous-Specific Synthetic Molecules with Trained Scent Dogs; Similarities Between Natural and Synthetic Oestrous Smell. Rep. Domest. Anim. 50, 7-12.
Fischer-Tenhagen, C., Tenhagen, B.A., Heuwieser, W., 2013. Short communication: Ability of dogs to detect cows in estrus from sniffing saliva samples. J. Dairy Sci. 96, 1081-1084.
Fischer-Tenhagen, C., Wetterholm, L., Tenhagen, B.-A., Heuwieser, W., 2011. Training dogs on a scent platform for oestrus detection in cows. Appl. Anim. Behav. Sci. 131, 63-70.
Furton, K.G., Myers, L.J., 2001. The scientific foundation and efficacy of the use of canines as chemical detectors for explosives. Talanta 54, 487-500.
Garner, K.J., Busbee L., Cornwell, P., Edmonds J., Mullins K., Rader K., Johnston J.M., Williams J.M. , 2000. Duty cycle of the detection dog: A baseline study. Final Report, Federal Aviation Administration, Washington D.C.
Gazit, I., Goldblatt, A., Terkel, J., 2005. The role of context specificity in learning: the effects of training context on explosives detection in dogs. Anim. Cogn. 8, 143-150.
Gazit, I., Terkel, J., 2003a. Domination of olfaction over vision in explosives detection by dogs. Appl. Anim. Behav. Sci. 82, 65-73.
Gazit, I., Terkel, J., 2003b. Explosives detection by sniffer dogs following strenuous physical activity. Appl. Anim. Behav. Sci. 81, 149-161.
Goodwin, K.M., Engel Rick, E., Weaver, D.K., 2010. Trained Dogs Outperform Human Surveyors in the Detection of Rare Spotted Knapweed (Centaurea stoebe). Invasive Plant Sci. Manag. 3, 113-121.
Gordon, R.T., Schatz, C.B., Myers, L.J., Kosty, M., Gonczy, C., Kroener, J., Tran, M., Kurtzhals, P., Heath, S., Koziol, J.A., Arthur, N., Gabriel, M., Hemping, J., Hemping, G., Nesbitt, S., Tucker-Clark, L., Zaayer, J., 2008. The use of canines in the detection of human cancers. J. Altern. Complement. Med. 14, 61-67.
Göth, A., Mc Lean, I.G., Trevelyan, J., 2003. How do dogs detect landmines? Mine Detection Dogs 18, 195-208.
Gsell, A., Innes, J., Monchy, P.d., Brunton, D., 2010. The success of using trained dogs to locate sparse rodents in pest-free sanctuaries. Wildlife Res. 37, 39-46.

Haimerl, P., Heuwieser, W., 2014. Invited review: Antibiotic treatment of metritis in dairy cows: A systematic approach. J. Dairy Sci. 97, 6649-6661.
Hall, N.J., Glenn, K., Smith, D.W., Wynne, C.D., 2015. Performance of Pugs, German Shepherds, and Greyhounds (Canis lupus familiaris) on an Odor-Discrimination Task. J. Comp. Psychol. 25, 25.
Hare, B., Tomasello, M., 2005. Human-like social skills in dogs? Trends Cogn. Sci. 9, 439-444.
Hatt, H., 2004. Molecular and Cellular Basis of Human Olfaction. Chem. Biodivers. 1, 1857-1869.
Helton, W.S., 2009. Overview of scent detection work. Issues and opportunities. CRC Press, Boca Raton, FL, 83–97.
Horvath, G., Jarverud, G.A., Jarverud, S., Horvath, I., 2008. Human ovarian carcinomas detected by specific odor. Integr. Cancer Ther. 7, 76-80.
Issel-Tarver, L., Rine, J., 1996. Organization and expression of canine olfactory receptor genes. Proc. Natl. Acad. Sci. 93, 10897-10902.
Jezierski, T., Adamkiewicz, E., Walczak, M., Sobczynska, M., Gorecka-Bruzda, A., Ensminger, J., Papet, E., 2014. Efficacy of drug detection by fully-trained police dogs varies by breed, training level, type of drug and search environment. Forensic Sci. Int. 237, 112-118.
Jezierski, T., Gorecka-Bruzda, A., Walczak, M., Swiergiel, A.H., Chruszczewski, M.H., Pearson, B.L., 2010. Operant conditioning of dogs (Canis familiaris) for identification of humans using scent lineup. Anim. Sci. Pap. Rep. 28, 81-93.
Jezierski, T., Walczak, M., Górecka, A., 2008. Information-seeking behaviour of sniffer dogs during match-to-sample training in the scent lineup. Polish Psychol. Bull. 39, 71-80.
Johnen, D., Heuwieser, W., Fischer-Tenhagen, C., 2013. Canine scent detection—Fact or fiction? Appl. Anim. Behav. Sci. 148, 201-208.
Johnen, D., Heuwieser, W., Fischer-Tenhagen, C., 2015. How to train a dog to detect cows in heat—Training and success. Appl. Anim. Behav. Sci. 171, 39-46
Kauhanen, E., Harri, M., Nevalainen, A., Nevalainen, T., 2002. Validity of detection of microbial growth in buildings by trained dogs. Environm. Intern. 28, 153-157.
Kerley, L.L., 2010. Using dogs for tiger conservation and research. Integr Zool 5, 390-396.
King, J.E., Becker, R.F., Markee, J.E., 1964. Studies on olfactory discrimination in dogs: (3) ability to detect human odour trace. Anim. Behav. 12, 311-315.
Krestel, D., Passe, D., Smith, J.C., Jonsson, L., 1984. Behavioral determination of olfactory thresholds to amyl acetate in dogs. Neurosci. Biobehav. Rev. 8, 169-174.
Lazarowski, L., Dorman, D.C., 2014. Explosives detection by military working dogs: olfactory generalization from components to mixtures. Appl. Anim. Behav. Sci. 151, 84-93.
Lin, H.M., Chi, W.L., Lin, C.C., Tseng, Y.C., Chen, W.T., Kung, Y.L., Lien, Y.Y., Chen, Y.Y., 2011. Fire ant-detecting canines: a complementary method in detecting red imported fire ants. J. Econ. Entomol. 104, 225-231.

Lit, L., Crawford, C.A., 2006. Effects of training paradigms on search dog performance. Appl. Anim. Behav. Sci. 98, 277-292.

Lit, L., Schweitzer, J.B., Oberbauer, A.M., 2011. Handler beliefs affect scent detection dog outcomes. Anim. Cogn. 14, 387-394.

Lorenzo, N., Wan, T., Harper, R., Hsu, Y.-L., Chow, M., Rose, S., Furton, K., 2003. Laboratory and field experiments used to identify Canis lupus var. familiaris active odor signature chemicals from drugs, explosives, and humans. Anal. Bioanal. Chem. 376, 1212-1224.

Lu, X.C., Slotnick, B.M., Silberberg, A.M., 1993. Odor matching and odor memory in the rat. Physiol. Behav. 53, 795-804.

Mandal, P.C., 2014. Gas Leak Detection in Pipelines & Repairing System of Titas Gas. 2014. J. Appl. Eng. 2, 23-34.

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

Moll, A., 1904. Mitteilungen über den "klugen Hans". Z. für Pädagog. Psychol. 3, 369.

Nash, M., 2005. Who let the dogs in? The use of Drug Sniffer Dogs in mental health settings. J. Psychiatr. Ment. Health Nurs. 12, 745-749.

Oesterhelweg, L., Krober, S., Rottmann, K., Willhoft, J., Braun, C., Thies, N., Puschel, K., Silkenath, J., Gehl, A., 2008. Cadaver dogs--a study on detection of contaminated carpet squares. Forensic Sci. Int. 174, 35-39.

Partyka, M.L., Bond, R.F., Farrar, J., Falco, A., Cassens, B., Cruse, A., Atwill, E.R., 2014. Quantifying the sensitivity of scent detection dogs to identify fecal contamination on raw produce. J. Food Prot. 77, 6-14.

Paula, J., Leal, M.C., Silva, M.J., Mascarenhas, R., Costa, H., Mascarenhas, M., 2011. Dogs as a tool to improve bird-strike mortality estimates at wind farms. J. Nat. Conserv. 19, 202-208.

Petri, A.L., Høgdall, C., Christensen, I.J., Simonsen, A.H., T'Jampens, D., Hellmann, M.-L., Kjaer, S.K., Fung, E.T., Høgdall, E., 2008. Sample handling for mass spectrometric proteomic investigations of human urine. Proteomics Clin. Appl. 2, 1184-1193.

Pickel, D., Manucy, G.P., Walker, D.B., Hall, S.B., Walker, J.C., 2004. Evidence for canine olfactory detection of melanoma. Appl. Anim. Behav. Sci. 89, 107-116.

Reed, S.E., Bidlack, A.L., Hurt, A., Getz, W.M., 2011. Detection distance and environmental factors in conservation detection dog surveys. J. Wildl. Manag. 75, 243-251.

Reindl-Thompson, S.A., Shivik, J.A., Whitelaw, A., Hurt, A., Higgins, K.F., 2006. Efficacy of scent dogs in detecting black-footed ferrets at a reintroduction site in South Dakota. Wildl. Soc. Bull. 34, 1435-1439.

Richards, K.M., Cotton, S.J., Sandeman, Richard M., 2008. The use of detector dogs in the diagnosis of nematode infections in sheep feces. J. Vet. Behav.: Clin. Appl. Res. 3, 25-31.

Savidge, J.A., Stanford, J.W., Reed, R.N., Haddock, G.R., Adams, A.A.Y., 2011. Canine detection of free-ranging brown treesnakes on Guam. N. Z. J. Ecol. 35, 174-181.

Schallschmidt, K., Becker, R., Zwaka, H., Menzel, R., Johnen, D., Fischer-Tenhagen, C., Rolff, J., Nehls, I., 2015. In vitro cultured lung cancer cells are not suitable for animal-based breath biomarker detection. J. Breath Res. 9, 027103.

Schoon, A., Fjellanger, R., Kjeldsen, M., Goss, K.-U., 2014. Using dogs to detect hidden corrosion. Appl. Anim. Behav. Sci. 153, 43-52.

Schoon, G.A.A., 1996. Scent identification lineups by dogs (Canis familiaris): experimental design and forensic application. Appl. Anim. Behav. Sci. 49, 257-267.

Schoon, G.A.A., De Bruin, J.C., 1994. The ability of dogs to recognize and cross-match human odours. Forensic Sci. Int. 69, 111-118.

Smith, R., 1996. What clinical information do doctors need? BMJ 313, 1062.

Sonoda, H., Kohnoe, S., Yamazato, T., Satoh, Y., Morizono, G., Shikata, K., Morita, M., Watanabe, A., Kakeji, Y., Inoue, F., Maehara, Y., 2011. Colorectal cancer screening with odour material by canine scent detection. Gut 60, 814-819.

Tomaszewski, T., Girdwoyn, P., 2006. Scent identification evidence in jurisdiction (drawing on the example of judicial practice in Poland). Forensic Sci. Int. 162, 191-195.

Vaidyanathan, R., Feldlaufer, M.F., 2013. Bed bug detection: current technologies and future directions. Am. J. Trop. Med. Hyg. 88, 619-625.

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

Walker, D.B., Walker, J.C., Cavnar, P.J., Taylor, J.L., Pickel, D.H., Hall, S.B., Suarez, J.C., 2006. Naturalistic quantification of canine olfactory sensitivity. Appl. Anim. Behav. Sci. 97, 241-254.

Wasser, S.K., Smith, H., Madden, L., Marks, N., Vynne, C., 2009. Scent-Matching Dogs Determine Number of Unique Individuals From Scat. J. Wildlife Manag. 73, 1233-1240.

Weber, C.M., Cauchi, M., Patel, M., Bessant, C., Turner, C., Britton, L.E., Willis, C.M., 2011. Evaluation of a gas sensor array and pattern recognition for the identification of bladder cancer from urine headspace. Analyst. 136, 359-364.

Willis, C.M., Britton, L.E., Harris, R., Wallace, J., Guest, C.M., 2010. Volatile organic compounds as biomarkers of bladder cancer: Sensitivity and specificity using trained sniffer dogs. Cancer Biomark. 8, 145-153.

Willis, C.M., Church, S.M., Guest, C.M., Cook, W.A., McCarthy, N., Bransbury, A.J., Church, M.R., Church, J.C., 2004. Olfactory detection of human bladder cancer by dogs: proof of principle study. Br. Med. J. 329, 712.

5 DISCUSSION

The main objective of this thesis was to study how to train dogs to search and indicate cows in heat by smell under practical conditions. For the development of an employable training system studies on training and testing scent detection dogs for comparable tasks were consulted. In those studies a wide variety of study designs and implementation of scent dog training and testing was obvious and provoked us to focus on evidence of published literature.

The first study was performed to evaluate the quality and comparability of published literature concerning canine scent detection according to criteria of evidence-based medicine. Furthermore, an experiment was conducted to determine the influence of the testing system on the outcome of a scent detection task considering two different testing systems. For the literature evaluation a systematic analysis of 14 papers showed obvious differences in methodology and quality. In all studies training methods were based on positive reinforcement what has significant advantages compared to those based on aversive methods (Walker et al., 2006). Duration of training varied widely from seven days (Fischer-Tenhagen et al., 2011) to 16 months (Cornu et al., 2011) indicating discrepancies in the training efficiency. Most studies did not provide information about experience and qualification of dog trainers. We speculate that the trainer might be a plausible influence on the duration necessary for training and the success rate. As previously demonstrated the handler beliefs of the location of the target substrate can influence the outcome of scent detection work (Lit et al., 2011). Therefore, testing without blinding the handler potentially confounds results and can be considered a major flaw in the study design. Presentation of results describing diagnostic performance was also heterogeneous among the studies. These differences in methodology and the high variability of the results show that the considered studies are diverse in respect to relevant quality criteria. Only four studies out of 14 (28.6%) met important quality criteria for scent detection studies as they were double blinded, used test samples unknown to the dogs and calculated sensitivity and specificity or accuracy. For evaluating the possible influence of the testing system when performing the same scent detection task in two different testing systems seven dogs were trained to detect black tea as target scent in two different testing systems, a testing platform and a scent detection board. As the scent of black tea is perceptible even for humans, this scent detection task should be easy for dogs. Our data show high sensitivity (92.1%) and specificity (97.4%). In a study addressing ovarian cancer sensitivity and specificity was 100% and 97.5%, respectively (Horvath et al., 2008). We were surprised that our dogs performed less perfect in the easy task. However, Horvath et al. (2008) spent 12 months on training in contrast to two days in our study. Sensitivity and specificity for the detection of black tea was similar for the testing platform and the scent detection board and the detection against tap water and smelling distractions. Our data confirm a previous study on melanoma detection utilizing three different testing systems (Pickel et al., 2004). It seems that the testing system has no essential influence on the test outcome.

The objective of the second study was to train dogs for heat detection in cows on farm. We developed a training protocol and trained six dogs on a dairy farm in Brandenburg (Germany). Four dogs completed the training procedure. Their performance was tested in a double blind randomized trial. Overall training time for the dogs in our study was 15 months, with 300 sessions of 10 min each per dog. In the training process it became obvious that our training protocol had to be adapted in various parts. All dogs showed signs of distress when first confronted with the farm environment. This confirms the findings of Kiddy et al. (1978), who reported that two of six dogs were reluctant to approach cattle and therefore had to be excluded. We overcame the problem with intensive adaptation training. Nevertheless, one dog had to be excluded because of lacking adaptation. A second dog had to be excluded because of insufficient reliability, probably due to high distractibility by environmental noises. A selection process during training with following exclusion of unqualified dogs is described in various studies on scent detection dog training (Kiddy et al., 1978; Kiddy and Mitchell, 1981; Richards et al., 2008; Walczak et al., 2012). Locking up cows in head locks was difficult and changing or re-arranging cows became a challenge, which confirmed the experiences of Kiddy et al. (1978). The authors therefore changed positions only after every sixth to eighth trial. This holds the risk that dogs did not indicate the estrus specific smell but the individual cow. To reduce this problem we changed at least the position of the positive cow after every training session. Dogs are able to notice hidden clues (i.e. unconscious reactions of the trainer), known as the “Clever Hans” effect (Moll, 1904). It has been demonstrated that handler beliefs of the location of the target substrate can influence the outcome of scent detection work (Lit et al., 2011). Therefore, testing without blinding the handler potentially confounds results (Johnen et al., 2013). Unfortunately, in our study, the trainer only occasionally had a support person. Therefore it was not possible to consistently train in a blinded way. This can be considered as a limitation in our experimental design and possibly led to confusion of the dogs during the pretest and the final test, as those procedures were conducted in a double-blind manner. The data of the final test proved that dogs are able to detect estrus cow-side under practical conditions with an overall sensitivity and specificity of 71.9% and 93.0%, respectively. Our results are lower than those of Kiddy et al. (1978), whose dogs averaged 87.3% correct detection rate of cows in estrus. Study design of Kiddy et al. (1978) was not transferable to farm conditions, because dogs walked on a platform behind the cows to be able to sniff the urogenital region. In our study dogs walked on the feed alley which did not require extra equipment and was more hygienic and safer for the dogs. But with this setting the dogs had to pick up the estrus specific smell from saliva or breath, which is more challenging than vaginal fluid (Fischer-Tenhagen et al., 2013). Estrus detection efficiency in dairy cattle by visual observation varies from 90% to less than 50% (Roelofs et al., 2010). Trained dogs for estrus detection with a potential sensitivity and specificity of

71.9% and 93.0%, respectively, could be an effective approach to improve reproductive performance in dairy farms. Further research is warranted to develop an optimized training plan that allows training estrus detection dogs for practical use within an appropriate time and to assure a certain sensitivity and specificity in working estrus detection dogs over time.

In the context of our study we experienced a lack of accepted recommendations for testing scent detection dogs in the recent literature. Thus the objective of the third study was to evaluate recent publications on scent dog studies and identify factors in the procedure that can eventually bias the study outcome. Finally we wanted to summarize this information to recommend a best practice standard for scent dog studies. A systematic analysis and assessment of 54 papers describing testing of canine scent detection revealed obvious differences in methodology and quality. In contrast to my initial literature search using the same keywords and databases (Johnen et al., 2013), there was a major increase in number of studies found (31 in 2013 versus 87 in 2015). This demonstrates the increasing relevance of scent dogs for medicine and practical application. Level of quality was assessed using a check list. Studies were rated as being excellent, sufficient or poor for 11, 29 and 13 papers, respectively. For one study, an evaluation of quality according to the score was not possible. Date of publication was between 1964 and 2015. Papers classified as poor had a date of appearance between 1964 and 2013. This indicates that bias in testing scent detection dogs is still a current topic that needs to be discussed critically. Regarding the identified factors potentially biasing results of studies we deduced best practice standards that should be followed when scent detection dogs are tested to ensure best practice. Exactly defining the target odour is probably one of the most important issues in canine scent detection training and testing. It is not possible to directly compare the results of a scent detection task with a distinct, chemically known odour and a varying odour where the dogs have to generalize the target odour, as both kinds of scent detection tasks place different demands on the dogs i.e. recognition of the target odour vs. generalization (Cablk, 2008). Therefore the underlying question has to be defined (i.e. specific research hypothesis). Free search testing trials should approximate conditions that are prevalent under practical conditions as close as possible. In scent line-ups and matching the sample tasks, number of samples and ratio of target to non-target samples has to be described. It has to be defined if the dogs are allowed to re-check samples or not and if the dogs obligatory have to check every single sample before an indication. Position of target samples has to be randomized. Furton and Myers (2001) demand that every test of scent detection dogs should include negative controls in order to ensure that the dog does not indicate on a non-target sample to get its reward. In a perfect test scenario the dogs check every sample only once making a yes/no decision for every sample. Collection, storage and handling of the samples has to be conducted in a way that the possibility of a cross-contamination of non-target samples with target odour or of both target and non-target samples with foreign odours can be excluded (Furton and Myers, 2001). It should be guaranteed that there are no other systematic odour differences between target and non-target samples than the target odour, e.g. matching donors of target and non-target samples in respect to age, sex and other malignancies in the case of cancer detection (Willis et al., 2004). To prevent dogs from just recognizing individual samples known from the training procedure instead of searching for the target odour (Elliker et al., 2014), testing scent detection dogs should always be done with samples and odour carriers the dog had never had contact to before. Testing scent detection dogs should always be conducted in a double-blind fashion (Furton and Myers, 2001). In order to prevent the trainer as well as the dog from being able to guess the position of the target samples, sample position has to be randomized. Jeziersky et al. (2014) reported that in their study the dog handlers' intrinsic state (stress) may have influenced dogs' performance. To reduce this kind of bias the trainer should be positioned in a way that the dog is not able to interact with him while searching, i.e. by positioning the handler behind a screen with a one way mirror (Concha et al., 2014). For certain free searching tasks, it is impossible to remove the handler from the scenario. Here, constant training of testing scenarios can reduce potential bias as far as possible. In my first study, we speculated that the trainer's skills and previous training experience might influence the duration necessary for training a certain task and the training outcome and proposed further research (Johnen et al., 2013). The used training method should be reported in a way that the reader is able to reproduce the procedures. Until now, there is a lack of scientific evidence on suitability of particular breeds for detection tasks (Jezierski et al., 2014). Information about breed, age, sex, reproductive status and previous scent detection experience of the trained dogs should be provided in every study concerning canine scent detection.

6 SUMMARY

The overall objectives of this thesis were to develop a specific training protocol for training dogs to identify cows in estrus under practical conditions and to identify steps in the procedure of scent detection dog training and testing that can eventually bias the study outcome.

My first study revealed multiple differences in methodology and a high variability of the results of 14 studies on training and testing scent detection dogs recruited in a systematic literature research. For evaluation of the studies a check list detailing relevant information about the study design and the training and testing process was used. For the experimental part of this study seven dogs were trained to indicate black tea as target scent in two different testing systems, a testing platform and a scent detection board, and against two different types of negative samples, tap water and smelling distractions. We could achieve a high sensitivity (92.1%) and specificity (97.4%). There was no difference of performance in the two testing systems. Quality of negative samples had no influence on the results.

In the second study six dogs were trained for heat detection in dairy cows. A specific training protocol was developed for training dogs to identify cows in estrus from the feed alley. Four of those dogs participated in the final test after an average training time of 50 h per dog. Finally the dogs overall correctly identified positive cows as being positive in 23 out of 32 cases (i.e. sensitivity of 71.9%) and falsely classified positive cows as being negative in nine cases (28.1% type II errors). Out of 128 cases 119 were correctly classified as true negative cows (i.e. specificity of 93.0%) and in nine cases negative cows were falsely identified as positive cows (7.0% type I errors).

To process lessons we learned from the first and the second study, we wanted to identify factors in the procedure of dog training and testing that can eventually bias the study outcome. Therefore we performed a systematic literature research and systematic assessment. The objective was to evaluate publications on scent dog studies. Finally I wanted to summarize this information to recommend a best practice standard for scent dog studies. A total of 54 studies could be included for final evaluation. We were able to identify the target odour, the type of scent detection task and experimental set-up, the samples used for training and testing, the test design, the dog trainer and training procedure chosen and the used dog breeds as factors potentially influencing the outcome of scent detection studies with trained dogs. With these findings a best practice standard was recommended.

We were able to show that trained dogs can detect cows in estrus on farm with better sensitivity and specificity than is often obtained by conventional heat detection methods. Further research is warranted to develop an optimized training plan that allows training estrus detection dogs for practical use within an appropriate time and to assure a certain sensitivity and specificity in working estrus detection dogs over time. The use of dogs as a diagnostic

tool for medical and biological applications is becoming increasingly important. It is mandatory, however, to follow standardized test protocols for scent detection dogs to allow valid comparisons of results. We recommend the application of the best practice standards to reduce possible bias in training and testing scent detection dogs as far as possible.

7 ZUSAMMENFASSUNG

Wie Hunde trainiert werden, brünstige Kühe am Geruch zu erkennen – Erfahrungen aus dem Training von Spürhunden

Ziel dieser Arbeit war es, wissenschafts-basiert ein spezifisches Protokoll für das Training von Spürhunden für brünstige Kühe unter Praxisbedingungen zu entwickeln. Zudem sollten diejenigen Faktoren innerhalb des Trainings- und Testprozesses von Spürhunden identifiziert werden, die letztendlich zu systematischen Fehlern in Studienergebnissen führen können.

In meiner ersten Studie wurden Unterschiede hinsichtlich der verwendeten Methodik und eine hohe Variabilität der Ergebnisse von 14 Studien zum Thema Training und Testen von Spürhunden offenbart, die im Rahmen einer systematischen Literaturrecherche ausgewählt wurden. Um die Studien auszuwerten, wurde eine Checkliste verwendet, mit deren Hilfe relevante Informationen hinsichtlich des Studiendesigns sowie des Trainings- und Testprozesses abgefragt wurden. Im experimentellen Teil der Studie wurden sieben Hunde trainiert, schwarzen Tee als Zielgeruch anzuzeigen. Mit einer Testplattform und einer Geruchsstrecke wurden zwei unterschiedliche Testsysteme verwenden sowie gegen zwei verschiede Arten von Negativproben diskriminiert, namentlich klares Wasser sowie Verleitgerüche. Es konnten eine hohe Sensitivität (92,1%) sowie Spezifität (97.4%) erreicht werden. Die Leistung der Hunde unterschied sich statistisch nicht zwischen den beiden Testsystemen. Die Art der verwendeten Negativproben hatte ebenfalls keinen Einfluss auf das Ergebnis.

Im Rahmen meiner zweiten Studie wurden sechs Hunde trainiert, brünstige Milchkühe am Geruch zu erkennen und anzuzeigen. Es wurde ein spezifisches Trainingsprotokoll entwickelt, das ein Training der Hunde für eine Suche vom Futtertisch aus ermöglichte. Vier der Hunde nahmen nach einer durchschnittlichen Trainingszeit von 50 h pro Hund an einem abschließenden Test teil. In diesem identifizierten die Hunde brünstige Kühe in 23 von 32 Fällen korrekt als brünstig (Sensitivität von 71,9%) und klassifizierten brünstige Kühe fälschlicherweise in neun Fällen als nicht-brünstig (28,1% Fehler II. Art). Von 128 nicht-brünstigen Kühen wurden 119 korrekt als nicht-brünstig klassifiziert (Spezifität von 93,0%), währen in neun Fällen nicht-brünstige Kühe fälschlicherweise als brünstig identifiziert wurden (7,0% Fehler I. Art).

Um während der Durchführung der ersten und zweiten Studie gemachte Erfahrungen zu verarbeiten, sollten diejenigen Faktoren innerhalb des Trainings- und Testprozesses von Spürhunden identifiziert werden, die letztendlich zu systematischen Fehlern in Studienergebnissen führen können. Daher wurde eine Literaturrecherche und systematische Auswertung mit dem Ziel durchgeführt, Veröffentlichungen von Studien mit Spürhunden auszuwerten, um einen Best-Practice-Standard für das Durchführen solcher Studien zu empfehlen. Insgesamt konnten 54 Studien in die Auswertung einbezogen werden. Als Faktoren, die möglicherweise Auswirkungen auf das Ergebnis von Studien mit Spürhunden haben können, wurden der Zielgeruch, die Art der gestellten Geruchsaufgabe sowie der Versuchsaufbau, die für Training und Test verwendeten Geruchsproben, das Testdesign, der Hundetrainer, die verwendete Trainingsmethode sowie die verwendeten Hunderassen identifiziert. Mit diesen Ergebnissen wurde ein Best-Practice -Standard empfohlen.

Es konnte gezeigt werden, dass trainierte Hunde brünstige Kühe auf einem landwirtschaftlichen Betrieb mit besserer Sensitivität und Spezifität anzeigen können, als in der Regel mit konventionellen Brunstbeobachtungsmethoden erreicht werden. Die Erforschung weiterer Lösungsansätze ist notwendig, um einen optimierten Trainingsplan zu entwickeln, der das Training von Brunstsuchhunden für den praktischen Einsatz innerhalb einer angemessenen Zeit und mit konstanter Zuverlässigkeit erlaubt. Die Bedeutung von Hunden als Diagnostikum in Gebieten der Medizin und Biologie nimmt mehr und mehr zu. In jedem Fall müssen bei der Arbeit mit Spürhunden standardisierte Testcharakteristika eingehalten werden, um gültige und vergleichbare Ergebnisse zu erhalten. Wir empfehlen die Anwendung der Best-Practice-Standards, um systematische Fehler beim Training und Testen von Spürhunden so weit wie möglich zu reduzieren.

8 REFERENCES FOR INTRODUCTION AND DISCUSSION

Arner, L.D., Johnson, G.R., Skovronek, H.S., 1986. Delineating toxic areas by canine olfaction. J. Hazard. Mater. 13, 375-381.

At-Taras, E.E., Spahr, S.L., 2001. Detection and Characterization of Estrus in Dairy Cattle with an Electronic Heatmount Detector and an Electronic Activity Tag. J. Dairy Sci. 84, 792-798.

Becker, F., Kanitz, W., Heuwieser, W., 2005. Advantages and disadvantages of different methods of heat detection in cattle. Zuchtungskunde 77, 140-150.

Bijland, L.R., Bomers, M.K., Smulders, Y.M., 2013. Smelling the diagnosis: a review on the use of scent in diagnosing disease. Neth. J. Med. 71, 300-307.

Browne, C., Stafford, K., Fordham, R., 2006. The use of scent-detection dogs. Ir. Vet. J. 59, 97-102.

Cablk, M.E., Sagebiel, J. C., Heaton, J. S., Valentin, C., 2008. Olfaction-based Detection Distance: A Quantitative Analysis of How Far Away Dogs Recognize Tortoise Odor and Follow It to Source. Sensors 8, 2208-2222.

Concha, A., Mills, D.S., Feugier, A., Zulch, H., Guest, C., Harris, R., Pike, T.W., 2014. Using sniffing behavior to differentiate true negative from false negative responses in trained scent-detection dogs. Chem. Senses 39, 749-754.

Cooper, R., Wang, C., Singh, N., 2014. Accuracy of Trained Canines for Detecting Bed Bugs (Hemiptera: Cimicidae). J. Econ. Entomol. 107, 2171-2181

Cornu, J.N., Cancel-Tassin, G., Ondet, V., Girardet, C., Cussenot, O., 2011. Olfactory detection of prostate cancer by dogs sniffing urine: a step forward in early diagnosis. Eur. Urol. 59, 197-201.

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

Fischer-Tenhagen, C., Tenhagen, B.A., Heuwieser, W., 2013. Short communication: Ability of dogs to detect cows in estrus from sniffing saliva samples. J. Dairy Sci. 96, 1081-1084.

Fischer-Tenhagen, C., Wetterholm, L., Tenhagen, B.-A., Heuwieser, W., 2011. Training dogs on a scent platform for oestrus detection in cows. Appl. Anim. Behav. Sci. 131, 63-70.

Fjellanger, R., Andersen, E.K., McLean, I.G., 2002. A Training Program for Filter-Search Mine Detection Dogs. Int. J. Comp. Psychol. 15, 278-287.

Furton, K.G., Myers, L.J., 2001. The scientific foundation and efficacy of the use of canines as chemical detectors for explosives. Talanta 54, 487-500.

Gsell, A., Innes, J., Monchy, P.d., Brunton, D., 2010. The success of using trained dogs to locate sparse rodents in pest-free sanctuaries. Wildl. Res. 37, 39-46.

Hawk, H.W., Conley, H.H., Kiddy, C.A., 1984. Estrus-Related Odors in Milk Detected by Trained Dogs. J. Dairy Sci. 67, 392-397.
Heersche Jr, G., Nebel, R.L., 1994. Measuring Efficiency and Accuracy of Detection of Estrus. J. Dairy Sci. 77, 2754-2761.
Horvath, G., Jarverud, G.A., Jarverud, S., Horvath, I., 2008. Human ovarian carcinomas detected by specific odor. Integr. Cancer Ther. 7, 76-80.
Jezierski, T., 1992. The Effectiveness of Estrus Detection in Cows by a Trained Dog. Anim. Sci. Papers Rep. 10, 57-65.
Jezierski, T., Adamkiewicz, E., Walczak, M., Sobczynska, M., Gorecka-Bruzda, A., Ensminger, J., Papet, E., 2014. Efficacy of drug detection by fully-trained police dogs varies by breed, training level, type of drug and search environment. Forensic Sci. Int. 237, 112-118.
Jezierski, T., Walczak, M., Górecka, A., 2008. Information-seeking behaviour of sniffer dogs during match-to-sample training in the scent lineup. Polish Psychol. Bull. 39, 71-80.
Johnen, D., Heuwieser, W., Fischer-Tenhagen, C., 2013. Canine scent detection—Fact or fiction? Appl. Anim. Behav. Sci. 148, 201-208.
Kerley, L.L., 2010. Using dogs for tiger conservation and research. Integr Zool 5, 390-396.
Kiddy, C.A., Mitchell, D.S., 1981. Estrus-Related Odors in Cows: Time of Occurrence. J. Dairy Sci. 64, 267-271.
Kiddy, C.A., Mitchell, D.S., Bolt, D.J., Hawk, H.W., 1978. Detection of estrus-related odors in cows by trained dogs. Biol. Reprod. 19, 389–395.
Kiddy, C.A., Mitchell, D.S., Hawk, H.W., 1984. Estrus-Related Odors in Body Fluids of Dairy Cows. J. Dairy Sci. 67, 388-391.
Lit, L., Schweitzer, J.B., Oberbauer, A.M., 2011. Handler beliefs affect scent detection dog outcomes. Anim. Cogn. 14, 387-394.
Moll, A., 1904. Mitteilungen über den "klugen Hans". Z. Pädagog. Psychol. 3, 369.
Partyka, M.L., Bond, R.F., Farrar, J., Falco, A., Cassens, B., Cruse, A., Atwill, E.R., 2014. Quantifying the sensitivity of scent detection dogs to identify fecal contamination on raw produce. J. Food Prot. 77, 6-14.
Paula, J., Leal, M.C., Silva, M.J., Mascarenhas, R., Costa, H., Mascarenhas, M., 2011. Dogs as a tool to improve bird-strike mortality estimates at wind farms. J. Nat. Conserv. 19, 202-208.
Pickel, D., Manucy, G.P., Walker, D.B., Hall, S.B., Walker, J.C., 2004. Evidence for canine olfactory detection of melanoma. Appl. Anim. Behav. Sci. 89, 107-116.
Reynolds, S.D., Christian, D.P., Reynolds, K.G., 2008. Illicit discharge detection is going to the dogs. American Society of Civil Engineers, Reston.
Richards, K.M., Cotton, S.J., Sandeman, Richard M., 2008. The use of detector dogs in the diagnosis of nematode infections in sheep feces. J. Vet. Behav.r: Clin. Appl. Res. 3, 25-31.
Roelofs, J., Lopez-Gatius, F., Hunter, R.H., van Eerdenburg, F.J., Hanzen, C., 2010. When is a cow in estrus? Clinical and practical aspects. Theriogenology 74, 327-344.

Savidge, J.A., Stanford, J.W., Reed, R.N., Haddock, G.R., Adams, A.A.Y., 2011. Canine detection of free-ranging brown treesnakes on Guam. N. Z. J. Ecol. 35, 174-181.
Senger, P.L., 1994. The estrus detection problem: new concepts, technologies, and possibilities. J. Dairy Sci. 77, 2745-2753.
Vaidyanathan, R., Feldlaufer, M.F., 2013. Bed bug detection: current technologies and future directions. Am. J. Trop. Med. Hyg. 88, 619-625.
Walczak, M., Jezierski, T., Górecka-Bruzda, A., Sobczyńska, M., Ensminger, J., 2012. Impact of individual training parameters and manner of taking breath odor samples on the reliability of canines as cancer screeners. J. Vet. Behav.: Clin. Appl. Res. 7, 283-294.
Walker, D.B., Walker, J.C., Cavnar, P.J., Taylor, J.L., Pickel, D.H., Hall, S.B., Suarez, J.C., 2006. Naturalistic quantification of canine olfactory sensitivity. Appl. Anim. Behav. Sci. 97, 241-254.
Welch, J.B., 1990. A detector dog for screwworms (Diptera: Calliphoridae). J. Econ. Entomol. 83, 1932-1934.
Willis, C.M., Church, S.M., Guest, C.M., Cook, W.A., McCarthy, N., Bransbury, A.J., Church, M.R., Church, J.C., 2004. Olfactory detection of human bladder cancer by dogs: proof of principle study. Br. Med. J. 329, 712.

9 PUBLICATIONS

Research articles

Johnen, D., Heuwieser, W., Fischer-Tenhagen, C., (2013):
Canine scent detection - Fact or fiction? Applied Animal Behaviour Science. 148: 201-208.

Johnen, D.; Heuwieser, W.; Fischer-Tenhagen, C. (2014):
Suchhunde testen - zu Risiken und Nebenwirkungen: Canine Scent Detection - Fact or Fiction. Forschungsergebnisse aus dem Institut für Rechtsmedizin der Universität Hamburg. 28: 39–46.

Johnen, D., Heuwieser, W., Fischer-Tenhagen, C. (2015):
How to train a dog to detect cows in heat - Training and success. Applied Animal Behaviour Science. 171: 39-46.

Fischer-Tenhagen, C., Johnen, D., Le Danvic, C., Gatien, J., Salvetti, P., Tenhagen, B.A., Heuwieser, W. (2015):
Validation of Bovine Oestrous-Specific Synthetic Molecules with Trained Scent Dogs; Similarities Between Natural and Synthetic Oestrous Smell. Reproduction in Domestic Animals. 50, 7-12.

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

10 ACKNOWLEDGEMENTS

Ich bedanke mich bei Herrn Prof. Dr. Heuwieser für die Möglichkeit, diese Arbeit durchführen zu können, die gute Betreuung, das kritische Korrekturlesen und die zahlreichen Anregungen. Dem Verein Tiergyn e.V. danke ich für die finanzielle Unterstützung.

Mein besonderer Dank gilt meiner Betreuerin Dr. Carola Fischer-Tenhagen für die großartige Idee, das in mich gesetzte Vertrauen und ihre tolle Unterstützung. Es hat viel Spaß gemacht, mit Dir zu arbeiten.

Vielen Dank an die Kollegen, die mir vertrauensvoll ihre Hunde zur „Mitarbeit" überlassen haben. Weiterhin bedanke ich mich bei allen Mitgliedern der Tierklinik für Fortpflanzung für ihre Hilfe, Anregungen und das Interesse an diesem doch etwas außergewöhnlichen Thema. Es war eine tolle Zeit mit Euch!

Ein großes Dankeschön an Herrn Rollandt und seine Familie und Mitarbeiter dafür, dass ich mit den Hunden über einen so langen Zeitraum auf seinem Betrieb willkommen war und so großartige Unterstützung erfahren habe.

Ich bedanke mich bei Viviane Theby für alles, was ich von ihr lernen durfte und noch darf. Ohne Dich wäre ich niemals so tief in dieses spannende Thema eingetaucht. Ein weiteres Dankeschön geht an Michaela Hares für die Unterstützung beim Hundetraining und die tatkräftige Hilfe beim Abschlusstest.

Claudia Neuber danke ich dafür, dass sie seit vielen Jahren meine beste Freundin ist, mich seit Anfang des Studiums bis zum Ende dieser Promotion begleitet hat und mir immer zur Seite stand und steht.

Der größte Dank gilt meinen Eltern, Großeltern und meinem Freund Daniel, die mich immer und in allen Belangen vorbehaltlos unterstützt haben.

Zu guter Letzt bedanke ich mich bei den Hunden, die meine wichtigsten und motiviertesten Mitarbeiter waren und mit Feuereifer jedes „silly little game" mitgespielt haben, dass ich mir ausgedacht habe. Ihr wart die besten Lehrmeister, die man sich wünschen kann.

11 DECLARATION OF INDEPENDENCE

Hiermit erkläre ich, dass ich, Dorothea Johnen, alle Studien selbständig durchgeführt und die vorliegende Arbeit selbständig angefertigt habe. Ich versichere, dass ich ausschließlich die angegebenen Quellen und Hilfen in Anspruch genommen habe.

Tabelle 1: Eigener Anteil[1] an den Forschungsprojekten der vorliegenden Dissertation

	Studie 1[a]	Studie 2[b]	Studie 3[c]
Studienplanung	+++	+++	+++
Datenerhebung	+++	+++	+++
Datenanalyse	+++	+++	+++
Verfassen des Manuskripts	+++	+++	+++
Editieren des Manuskripts	++	++	++

[1]Legende: +++: > 70 %
++: 50-70 %
+: < 50 %

[a] Canine Scent Detection – Fact or Fiction?

[b] How to train a dog to detect cows in heat – Training and success

[c] An approach to identify bias in scent detection dog testing

Berlin, den 08.01.2016

Dorothea Johnen