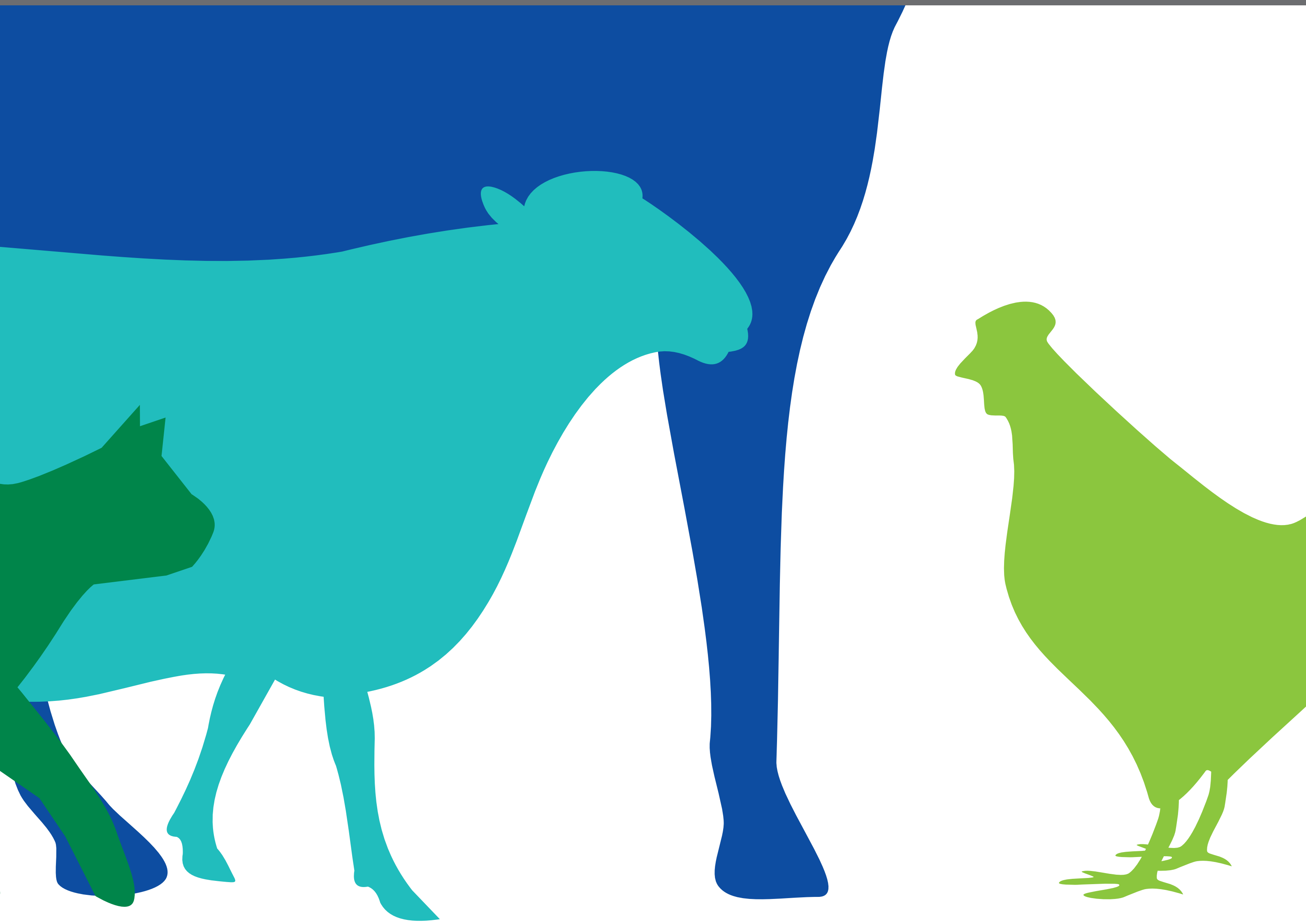
WORKING DOGS: FORM AND FUNCTION, VOLUME II

EDITED BY: Cynthia M. Otto, Nathaniel James Hall and Wendy Irene Baltzer
PUBLISHED IN: Frontiers in Veterinary Science



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-397-4
DOI 10.3389/978-2-88971-397-4

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

WORKING DOGS: FORM AND FUNCTION, VOLUME II

Topic Editors:
Cynthia M. Otto, University of Pennsylvania, United States
Nathaniel James Hall, Texas Tech University, United States
Wendy Irene Baltzer, The University of Sydney, Australia

Citation: Otto, C. M., Hall, N. J., Baltzer, W. I., eds. (2021). Working Dogs: Form and Function, Volume II. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88971-397-4

Table of Contents

frontiers in Veterinary Science

EDITORIAL
published: 29 July 2021
doi: 10.3389/fvets.2021.732304

Editorial: Working Dogs: Form and Function, Volume II

Nathaniel J. Hall[1], Cynthia M. Otto[2,3]* and Wendy I. Baltzer[4]

[1] Department of Animal and Food Science, Texas Tech University, Lubbock, TX, United States, [2] Penn Vet Working Dog Center, Philadelphia, PA, United States, [3] Department of Clinical Sciences and Advanced Medicine, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, United States, [4] Faculty of Science, The University of Sydney, Sydney, NSW, Australia

Keywords: olfaction, stress, behavior, service dog, detection dog, musculoskeletal, human animal bond, occupational hazards

Editorial on the Research Topic

Working Dogs: Form and Function, Volume II

OPEN ACCESS

Edited and reviewed by:
Marta Hernandez-Jover, Charles Sturt University, Australia

****Correspondence:***
Cynthia M. Otto
cmotto@vet.upenn.edu

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences, a section of the journal Frontiers in Veterinary Science

Received: *28 June 2021*
Accepted: *05 July 2021*
Published: *29 July 2021*

Citation:
Hall NJ, Otto CM and Baltzer WI (2021) Editorial: Working Dogs: Form and Function, Volume II. Front. Vet. Sci. 8:732304. doi: 10.3389/fvets.2021.732304

INTRODUCTION

Working dogs span the spectrum of careers from high powered physical athletes working in protection fields to highly cognitive dogs working in the service sector. Regardless of the career, all working dogs share a common requirement for physical capabilities and mental aptitude and partnership with their handler. In this Research Topic, Working Dogs: Form and Function, twenty manuscripts address the development and assessment of the physical Form, behavioral selection of dogs that contribute to Function, a focus on olfactory Function, occupational hazards that interfere with Form and Function and the relationship between the working dog and handler that addresses Form and Function of the working dog team.

OCCUPATIONAL HAZARDS/STRESSORS

Working dogs perform tasks in a variety of domains. Dogs must adapt both physically and mentally to work regardless of the environment. Physical stamina is necessary for many working dogs. During exercise, these dogs must maintain physiologic homeostasis and mental acuity for prolonged periods of time. The most common physical stresses that impact performance are environmental temperatures, particularly heat stress and the physical wear and tear of high impact activity. Hydration is a critical factor in maintaining function and reducing the risk of heat injury (1). In a study of tracking dogs in the desert, the median working body temperature was 41C (106F) (Niedermeyer et al.). Dogs receiving electrolyte solutions or flavored water had increased fluid consumption compared to dogs receiving plain water. No adverse effects of electrolyte solutions were observed, however, chicken-flavored water without electrolytes was associated with increased markers of muscle injury.

Heat from muscle activity or the environment can lead to physiological stress and cellular injury. The gastrointestinal tract is affected by both physical stress (exercise, heat) (2) and mental stress (3). Many working dogs have diarrhea during or after intense exercise, which could be influenced by diet or physical conditioning. A group of nine hunting dogs in Italy were monitored over time to determine the effect of training, hunting and off-season rest on fecal stress markers and the microbiota (Zannoni et al.). Training and hunting did not detectably alter most of the stress markers, and only transiently impacted the microbiota, suggesting that this cycle of activity was only mildly to moderately stressful.

In Nicaragua, hunting dogs are an important part of the community, but often are left to roam, share the food of the family that owns them and are at risk for malnutrition and dehydration. In a study evaluating hair cortisol as a marker of chronic stress in 454 Nicaraguan hunting dogs (Bowland et al.), cortisol concentrations were higher in dogs with light-colored fur, and those with low body condition score (emaciated or thin). While excessive body fat can lead to progression of osteoarthritis (4), emaciation is associated with increased stress.

In addition to physiological stress, working dogs undergo physical stress from repetitive motion. This type of stress typically manifests as musculoskeletal injury, altered mechanics (joint range of motion), swelling, pain or lameness. In a longitudinal study of 323 New Zealand working farm dogs (Isaksen et al.), 57% of dogs developed at least one musculoskeletal abnormality as determined by a veterinary examination. The carpus and stifle were most commonly associated with reduced range of motion, whereas pain was most commonly found in the hip joint. Over half of the dogs that developed one abnormality, also developed a second abnormality over time. This study is one of the first to document increased risk of further injury following development of a primary musculoskeletal problem.

The anatomical structure of a working dog, in combination with the dog's kinematics (5), may impact the amount of strength and power it may generate, as well as affect its risk of development of musculoskeletal injuries. In a comprehensive review, Zink and Schlehr synthesized published data with expert observations to systematically describe the critical relationship between structure and function of the canine musculoskeletal system. Further research and understanding of the link between form and both skill and injury will be vital to improving performance and longevity of working dogs. Some of the common structural abnormalities in working dogs involve the lower back and pelvis. Even with normal structure, the repeated motion of jumping up onto raised surfaces or standing on the hind limbs to search elevated locations can put strain on the low back. Computed tomography was used to retrospectively evaluate the pelvis and lumbosacral spine in two different working breeds. The sacral iliac joint, which is the connection between the pelvis and spine, was evaluated for lesions in a retrospective study of 22 working Labrador retrievers by Carnevale et al., The methodology may be useful for future minimally invasive evaluations of working dogs with low back pain. Computed tomographic images of working military German Shepherd and Belgian Malinois lumbar vertebra were compared (Dragicevich et al.), the German Shepherd dogs had a higher incidence of both funnel-shaped lumbar vertebral foramina and articular process dysplasia malformations which may be associated with low back pain.

In addition to selecting potential working dogs based on a physical structure that will support the expected workload, preventing injury requires strategies of physical fitness to build muscle strength and flexibility. A novel approach to standardized training and canine fitness testing was described by Farr et al. Longitudinal studies will be necessary to determine the impact of fitness and conditioning on injury prevention, but standardized testing will be invaluable in such studies.

ODOR DETECTION

This special issue covers an array of topics relevant to the performance, capabilities, and assessment of detection dogs. The use of conservation detection dogs is growing in popularity and in success [e.g., (6–8)]. Fukuzawa and Shibata investigated dogs' detection limits for the Carolina anole and found that dogs could successfully identify samples from enclosures housing a wide variety of anole population densities. This fundamental research is critical to the continued development of detection dogs as a new tool in conservation work battling invasive species as well as acting as an important contribution to a promising and burgeoning field for the scientific use of detection canines in conservation.

In a remarkable synergy across four laboratories and international borders, the importance of standardization and attention to subtle procedural differences in detection dog training and evaluation emerged as the zeitgeist for this field. Guest et al., humbly share an important lesson learned with their medical detection canines. When a precipitous drop in performance was noted, instead of simply moving on, they investigated the cause which revealed that an easily missed difference in the processing of urine samples (whether a "dip stick" for urinalysis was placed in the sample) was aiding the dogs in the initial training. This highlights the importance of controlling every step along sample collection and processing as well as a constant evaluation of ongoing performance. The future of medical detection dog research will likely require much more attention to every detail of sample collection and processing steps to help move this field forward.

In another investigation into the effects of subtle differences in procedure on medical detection dogs' performance, Essler et al. demonstrate that the topography of the alert behavior (stop and stare vs. a sit) can have significant impacts on sampling behavior. Dogs that made a stop and stare alert showed more differentiation in sampling times between sample types (e.g., targets and non-targets) than did dogs that made a sit response. Continuing in the series of investigations on important methodological variables DeChant et al. evaluated the effect of each handler's knowledge level for the search task on team performance metrics. When handlers had more information about the number of targets to find, and whether certain areas may be "blank," search behavior changed compared to teams without such knowledge. Teams spent more time searching and the dog more time looking back toward the handler when they had less knowledge of the search compared to teams that had more knowledge. In addition, there was no overall difference in performance when a trained researcher was monitoring searches in a single-blind (handler blind but researcher not) compared to a double-blind (handler and researcher blind) test.

The assessment of canine performance and measurement of a dog's capacity for high performance is critical for the successful procurement and training of detection dogs. In a series of studies, Rooney and Clark developed a monitoring instrument for detection dog performance in which observers rate dogs' performance during searches. Through systematic investigation, the authors investigated how subtle differences in the way the scoring system is presented and conducted (i.e., adding benchmarks to the rating scale or having handlers rate their own dogs) can lead to important differences in ratings (Clark and Rooney; Clark et al.). Together, these papers highlight the non-trivial nature of developing ratings for detection dog performance, and how minor changes in how questions are posed, whether benchmarks are given, or familiarity of the rater to the dog being rated, can all impact the results.

In addition to developing performance metrics, it is also important to identify and describe the behavioral characteristics associated with explosives detection performance. Lazarowski et al. review the prior research and identify behavioral characteristics consistently associated with optimal explosives detection dog performance. The authors focus on three broad categories, detection characteristics, overall trainability, and environmental soundness. Together, this review highlights how individual behavioral characteristics can have important functional consequences on a dog's suitability as an explosives detection dog and will lead to new and exciting research directions.

Lastly, two additional extensive narrative reviews provide thorough summaries of even more methodological considerations in canine detection work. In the first review, Lazarowski et al. discuss many of the same methodological considerations raised elsewhere in this Research Topic issue as well as many other considerations. This review will likely become a critical reference material for those interested in starting detection canine research. Similarly, Simon et al. provide a thorough review of the types of canine training aids. Training aids are the odor sources used for detection dogs. In many circumstances, the target material dogs are trained to detect (e.g., explosives) maybe too dangerous for frequent training. Therefore, the relevant odor needs to be presented in a safe and reliable manner and this extensive review highlights the varying approaches and limitations of each. As the use of detection dogs extends into new areas (biohazard detection such as COVID-19 or detection of critically endangered animals) methods to collect and store relevant target odorants will become of greater importance, and this review is a great starting point to learn about the benefits and limitations of each training aid technology.

HUMAN-ANIMAL BOND

Last, but not least, it is important to recognize that nearly all working dogs work within the context of a human partnership. The human-animal bond is therefore a critical aspect to evaluate and perhaps most important within the context of service dogs. Lloyd et al. examined feelings and experiences when a service dog partnership has come to an end. Their results highlight the grief and negative feelings associated with the loss of a guide dog drawing attention to the similarity of grief following the loss of a pet dog. These results underscore the importance of considering the human animal bond in the working dog industry, especially following the loss of a service dog.

The human-animal bond also comes into play when considering the training methodology selected. Different working dog (and pet dog) domains can have different training methodologies to achieve the goals needed for the dog's assigned tasks. Many methods are employed including the use of aversive stimuli such as electronic collars. China et al. show that when comparing the efficiency and performance of dogs trained on recall and sit using an electronic collar or positive reinforcement, there was no overall difference in the number of commands disobeyed between groups. This highlights that positive reinforcement procedures can be just as efficient and effective as electronic collars. Positive reinforcement procedures may build a stronger human-animal bond and avoid the potential welfare risks from aversive techniques, therefore, identifying training methods for working dogs that are not only highly effective but also promote animal welfare and the human-animal bond is an important future direction for this field. Together, these two important papers highlight the need for additional research investigating the human-animal bond in the context of working dogs, which still receives little research attention.

OVERALL SUMMARY

In order to maintain form and function, awareness of the physical and behavioral stressors of working dogs is necessary. These studies have shown that although working dogs are resilient, they are often at risk of stressors that can impact their welfare and performance. The topics of working dog welfare, nutrition, hydration, physical fitness and exercise are all timely topics that warrant continued investigation.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

1. Horowitz M, Nadel ER. Effect of plasma volume on thermoregulation in the dog. *Pflugers Arch Eur J Physiol.* (1984) 400:211–3. doi: 10.1007/BF00585045
2. Alves JC, Santos A, Jorge P, Pitães A. The use of soluble fibre for the management of chronic idiopathic large-bowel diarrhoea in police working dogs. *BMC Vet Res.* (2021) 17:100. doi: 10.1186/s12917-021-02809-w
3. Lambert GP. Stress-induced gastrointestinal barrier dysfunction and its inflammatory effects. *J Anim Sci.* (2009) 87:E101–8. doi: 10.2527/jas.2008-1339

4. Kealy RD, Lawler DF, Ballam JM, Mantz SL, Biery DN, Greeley EH, et al. Effects of diet restriction on life span and age-related changes in dogs. *J Am Vet Med Assoc*. (2002) 220:1315–20. doi: 10.2460/javma.2002.220.1315
5. Fischer MS, Lilje KE. *Dogs in Motion*. Dortmund: VDH Service GmbH (2011).
6. Cablk ME, Heaton JS. Accuracy and reliability of dogs in surveying for desert tortoise (*Gopherus agassizii*). *Ecol Appl*. (2006) 16:1926–35. doi: 10.1890/1051-0761(2006)0161926:AARODI2.0.CO;2
7. Jean-Marie B, Raphael G, Fabien R, Aurélien B, Sébastien C, Nicolas B, et al. Excellent performances of dogs to detect cryptic tortoises in Mediterranean scrublands. *Biodiver Conserv*. (2019) 28:4027–45. doi: 10.1007/s10531-019-01863-z
8. Smith DA, Ralls K, Hurt A, Adams B, Parker M, Davenport B, et al. Detection and accuracy rates of dogs trained to find scats of San Joaquin kit foxes (Vulpes macrotis mutica). *Anim Conserv*. (2003) 6:339–46. doi: 10.1017/S136794300300341X

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 13 March 2020
doi: 10.3389/fvets.2020.00118

Check for updates

Differences in the Search Behavior of Cancer Detection Dogs Trained to Have Either a Sit or Stand-Stare Final Response

Jennifer L. Essler[1*†], *Clara Wilson*[1†], *Alexander C. Verta*[1], *Rebecca Feuer*[1] and *Cynthia M. Otto*[1,2]

[1] *Penn Vet Working Dog Center, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, United States,* [2] *Department of Clinical Sciences and Advanced Medicine, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, United States*

OPEN ACCESS

Edited by:
Erik Hofmeister,
Auburn University, United States

Reviewed by:
John J. Ensminger,
Delta Hedge Consulting, New York, NY, United States
Lowell Paul Waggoner,
Auburn University, United States

***Correspondence:**
Jennifer L. Essler
esslerj@upenn.edu

† *These authors have contributed equally to this work and share first authorship*

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences, a section of the journal Frontiers in Veterinary Science

Received: *17 December 2019*
Accepted: *18 February 2020*
Published: *13 March 2020*

Citation:
Essler JL, Wilson C, Verta AC, Feuer R and Otto CM (2020) Differences in the Search Behavior of Cancer Detection Dogs Trained to Have Either a Sit or Stand-Stare Final Response. Front. Vet. Sci. 7:118. doi: 10.3389/fvets.2020.00118

Recent literature has demonstrated that dogs have the potential to detect, and communicate the presence of, various human diseases. However, there is a lack of investigation into whether commonplace training differences within the field could influence a dog's behavior during a biomedical detection task. Here we report on the behavior of four dogs trained to alert to blood plasma samples taken from individuals with ovarian cancer. One hundred trials per dog were selected from routine video recordings collected over a period of 13 months. Videos were coded frame by frame to quantify sample checking, alerting behavior, and durations of alert. Dogs had previously been trained to elicit a final response behavior once they had located the target odor. Two dogs had a "sit" response while the other two had a "stand-stare" response. Alert behavior was categorized as true positive (a correct alert to a cancer sample) or false positive (an incorrect alert to biological and non-biological controls and distractors). Hesitations were also recorded, where the dog either checks the sample twice or, spends a longer duration of time sniffing the sample than a true pass without carrying out their final response. Results show individual variation in the total frequency of false alerts elicited. However, the rate of hesitations appears to be influenced by alert style, with stand-stare dogs carrying out 40 and 32, respectively (total = 72) and sit dogs carrying out 7 and 8, respectively (total = 15). The stand-stare dogs had a non-significant difference in the duration of their true and false positive alerts. In contrast, the sit dogs showed a significant difference ($p < 0.001$), maintaining their false alerts for, on average, two times the duration of their true alerts. Stand-stare dogs increased the duration of time spent in contact with the port when plasma samples were present, whereas sit dogs spent on average 0.3 s in contact with the port regardless of what sample type it contained. These findings suggest that the type of operant response a biomedical detection dog has been trained may influence their sample checking and response behavior.

Keywords: canine, olfaction, biomedical detection, behavior, detection dogs, cancer

INTRODUCTION

Over the past decade, the use of dogs to detect and alert to human health conditions has expanded. There is growing evidence that dogs can be trained to alert to human disease samples, including, but not limited to: bladder cancer (1), breast cancer (2), cervical cancer (3), colorectal cancer (4), lung cancer (2, 5–7), ovarian cancer (8, 9), prostate cancer (10), melanoma (11), Clostridium difficile (12), and cystic fibrosis bacterial pathogens (13) [see Edwards et al. (14) for the most recent systematic review]. These studies employ a variety of human sample types, including breath, urine, blood plasma, excrement and sebum. Proficient training is fundamental to ensure a dog recognizes their target odor and is motivated to repeat the task over numerous trials. In most cancer detection studies, the dog is further exposed to samples taken from healthy controls, and samples taken from people who have benign tumors. During training, handlers attempt to specify the odor of cancer as the target, as opposed to general human odor or the presence of benign masses, by shaping the dog's response to the cancer positive samples. While individual training methods vary, most dogs are trained using positive reinforcement, with many using the aid of a marker cue (e.g., a clicker) to specify at the precise moment that the dog makes a correct choice (14). If correct, the dog will receive their reward, usually a toy [e.g., (15)] or food [e.g., (16)].

Biomedical detection dogs must be taught two components to be successful. Firstly, dogs must learn their target odor, and be able to discriminate between control and disease positive samples. Secondly, they must be taught a method of communicating that they have located the target odor, known as their "alert." To communicate with the experimenter, the dogs are conditioned to exhibit a specific behavior, most commonly sitting in front of the target odor. Employment of the sit alert in the biomedical field was likely influenced by passive alerts trained in other working dog fields, such as explosives detection [e.g., (17)]. Of the recent biomedical canine studies published, most reported that dogs had been trained to elicit a sit alert [e.g., (4, 11, 12, 15)]. Jezierski et al. (18) notes that their sampled dogs had a final response dependent on the dog's previous training and the "dog's preference," however usually consisted of the dog "sitting or lying down in front of the target sample." While this convention reduces ambiguity for the purposes of the experimenter, it is possible that the arbitrary nature of the behavior may impact their behavior and influence their decisions. It is imperative to minimize factors that may skew a dog's response on such a sensitive odor discrimination task to ensure that response behaviors are driven by the odor source rather than environmental variations. This highlights a potential issue in biomedical detection dog training, where the required alert behavior may actually impact a dog's performance at the task. Mancini et al. (19) highlight this issue, and argue that binary options (e.g., perform the trained alert behavior or do not perform the trained alert behavior) may limit the reliability of a canine's response to a sample. Mancini et al. (19) suggest an "honest signaling" method whereby trained alerts are not implemented, and instead the duration of non-trained behaviors, such as duration of sniffing the port, is used to distinguish between samples. This method, however, relies on the use of technology to accurately track behaviors to the millisecond and would be impossible for a trainer to reliably carry out by eye. Currently, most laboratories still rely on a behavioral cue from the dog to signal detection of the target odor.

The stand-stare alert, whereby the dog remains standing with their nose over the port and freezes, has been less widely used in the current biomedical detection literature. It is possible that dogs who carry out a stand-stare alert may receive more feedback from a sample as they are required to keep their nose on the sample as a function of their alert. Unlike sit alert dogs, to receive their reward, stand-stare dogs must maintain their nose in close proximity to the odor source. Sit alert dogs move back, away from the port, to carry out an alert, which may have an effect on the duration of their false alerts. This study asks whether the type of trained alert impacts a dog's sample checking and alert behaviors while detecting ovarian cancer from human blood plasma samples.

MATERIALS AND METHODS

The protocol was approved by the Institutional Animal Care and Use Committee at the University of Pennsylvania for dogs owned by the university (Protocol #804900).

Videos

Videos were pseudorandomly sampled from Penn Vet Working Dog Center's ovarian cancer detection program archives. Dogs in this program are routinely tested and video recorded in one to four sessions per week using the training protocol described in section Training Protocol. A Canon VICIA HF R700 camera, positioned on a wall mount, recorded all sessions. Videos were included under the restriction that the session had to have taken place once that dog had task acquisition (e.g., not during odor imprinting or alert development stages). Ten recorded sessions were selected per dog, representing 100 trials each. The videos sampled dated from between 08/12/2017 and 11/26/2018.

Subjects

Dogs included in the study were three females and one male, all neutered or spayed. Breeds were two German Shepherds, one Labrador Retriever and one English Springer Spaniel (min age: 2 years, max age: 7 years, mean age: 4.5 years). Dogs had been taught their alert behavior starting when they arrived at the center, at ∼8 weeks of age, and had been imprinted on ovarian cancer blood plasma a minimum of 3 months prior to when the study videos were recorded.

Training Protocol

As part of an ongoing project, dogs are trained one to four times per week to identify human blood plasma samples taken from an individual with ovarian cancer. Each session is video recorded and data is recorded at the time of the session, tracking which sample is in each port and the medical identification of the human biological samples. Trainers and experimenters are out-of-sight behind a wall for all trials, with the dogs observed on a

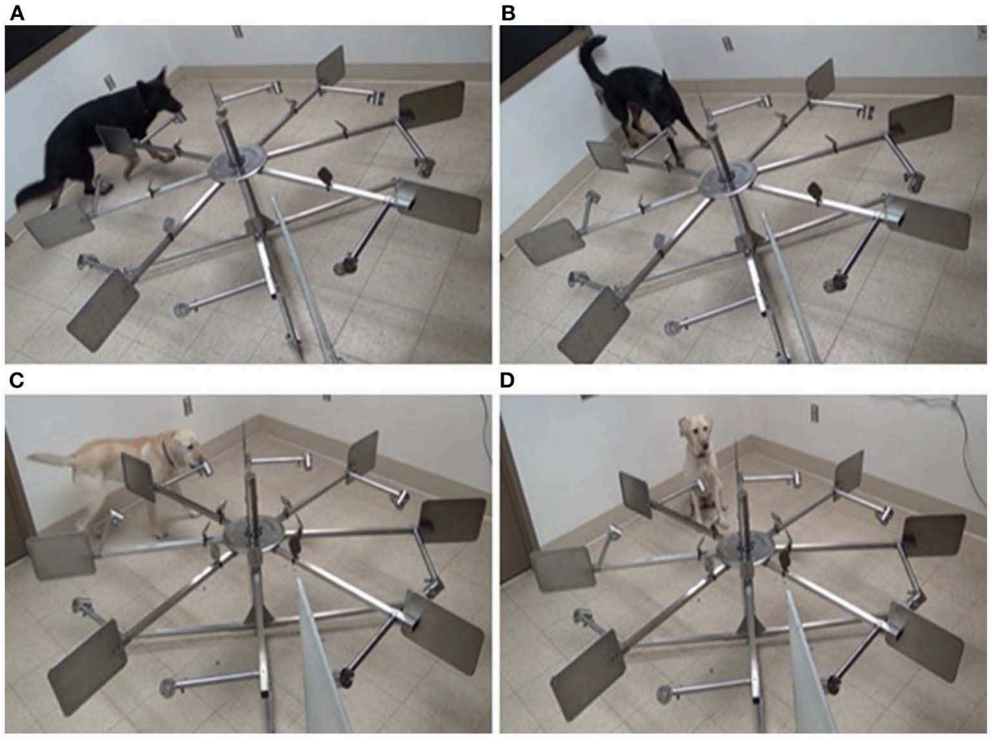


FIGURE 1 | (A) Bobbie checking port four (top left). **(B)** Bobbie carrying out a stand-stare alert at port four (top right). **(C)** Ffoster checking port four (bottom left). **(D)** Ffoster carrying out a sit-alert at port four (bottom right).

computer monitor screen via video. Dogs are trained on an eight-armed wheel with a "port" on each arm (Medical Detection Dogs, DEMAND—Design and Manufacture for Disability). Each port denotes a receptacle for one sample (see **Figure 1**). Within each port there is either (1) blood plasma taken from an individual with confirmed ovarian cancer, (2) blood plasma taken from an individual with a benign ovarian tumor (herein denoted as "benign", (3) blood plasma taken from a healthy individual (herein described as "normal"), (4) a control (a non-biological substance that is involved in the study process and may interfere with the identification of the target odor, e.g., latex gloves, as these are worn when handling samples), or (5) a distractor (a non-biological, unrelated, object e.g., paper clips). Dogs were presented with 75 µl of blood plasma during imprinting, and 50 µl in all subsequent training. For each "hot" trial, there is one cancer sample present (the target odor), and up to two benign or normal samples, the remaining ports contain distractors or controls. For one dog (McBaine, sit alert), an older version of the wheel that has twelve ports was being used at the time of recording (Anne B Kingsley Wheel). The distribution of and quantity of biological samples was identical, with the additional ports being used for additional distractor objects. For all dogs, each session contained ten trials, with 30-50% of these trials being "blanks" (no cancer sample is present). In blank trials, dogs are expected to check all ports of the wheel, and then walk to a raised platform to signal that there is no target odor present. During blank trials, normal, benign, control and distractor samples are present in the wheel.

Every dog was imprinted on the target odor using positive reinforcement and a clicker to mark their correct response. McBaine and Ffoster were shaped to elicit sit response during initial training at the center, and this was taken forward in the rest of their training, including ovarian cancer detection. During imprinting, the cancer odor was presented and the dog would sniff the sample, then told "sit." This was repeated until the verbal sit command could be phased out and the dog offered it automatically on smelling the cancer sample. Bobbie and Osa were shaped to have a stand-stare alert. This was trained by initially clicking as soon as they sniffed the cancer sample, then building up the duration of the nose-on-port behavior until a full stand-stare was established. During training, dogs were rewarded using either food or a toy, dependent on their preference. Once the target odor could be correctly identified on the wheel among non-biological odors (distractors and controls), other human biological samples were added; first normal samples and then benign samples. Dogs proceeded to each stage of training once they had reached a criterion of 80 percent of trials per session correct over three consecutive days. Videos were only included in the present study once the dogs had reached the final stage. This was carried out to safeguard from potential influences on behavior during the dog acquiring the task.

During all trials utilized for the present study, the dogs were sent to the scent wheel out-of-sight of the trainer and experimenters. The dog searched the wheel while the trainer watched on the computer monitor, and once the dog gave a correct alert on the cancer odor or correctly indicated that the

TABLE 1 | Behavioral ethogram used to code the videos.

Variable name	Description	Modifier	Measure
Pass	Dog checks port by making contact with their nose. Dog does not carry out alert behavior or hesitation and instead moves onto the next port or raised platform.	Sample type	Frequency
Hesitate	The dog maintains contact with the port for a greater duration of time than a true pass, but does not carry out their final response. Or, the dog passes the sample then flicks their head back to check the sample a second time.	Sample type	Frequency
Contact with port	Dog puts nose in contact with port.	Sample type	Duration
Stand-stare alert	Dog stands still with nose in contact with, or within one centimeter of, the port. Start behavior when the dog freezes. End behavior when the dog moves their head or body.	Sample type	Duration
Sit alert	Dog checks port and then sits behind port. Start behavior when dog's haunches touch the ground and all movement stops. End behavior when dog moves their head or body.	Sample type	Duration

wheel was free of cancer by moving to the raised platform, the trainer marked with a "click" and the dog came out for its reward, either food or a toy. Prior to the investigation into this study, there was no requirement specifically for the length of alert duration required from each dog for stand-stare dogs, and the duration decisions were left to the dog's specific trainer. Similarly, sit dogs were not required to hold a sit beyond it being a clear change of behavior on their target odor.

Coding

Videos were coded using The Observer XT 14. Behaviors included in the ethogram were based on a dog's response to each port and their alert behaviors (see **Table 1**). To ensure consistency of coding between alert types, alert behaviors were coded only once the dog had stopped motion. It was important to initiate coding of a sit alert once the dog's haunches touched the ground and the dog became motionless, to exclude the time taken for the dog to go from standing to sitting that would, by default, make the alert time longer. By the same measure, stand-stare alerts were initiated only once the dog had "frozen," and ended as soon as the dog moved out of their static position (see **Figure 1**).

Statistical Analysis

Twenty percent of trials were double coded. Inter-rater reliability was assessed using The Observer XT 14 Reliability Analysis function. Data was extracted from The Observer XT to Microsoft Excel version 16.25 for formatting. For each session, the dogs' duration data were averaged such that there was one number accounting for their duration of each behavior (**Table 1**). Statistical analyses were carried out on R version 3.5.1 (20). Using R package lme4 (21) a linear mixed effect model was run for mean duration of true and false positive alert with alert behavior (sit vs. stand-stare) and alert type (true positive vs. false positive) as fixed effects with an interaction, and dog name as a random effect. MASS package for R (22) was used to carry out generalized linear mixed effects model to compare duration of contact with port. A p-value of < 0.05 was considered statistically significant across all tests.

RESULTS

Inter-rater reliability was above 84% for each session, with an average of 87.73% agreement between observers (Kappa = 0.85, $p < 0.001$). All data can be found in **Supplementary Table 1**.

Rates of False Alerts and Hesitations

Over 200 trials (100 trials per dog), the sit dogs elicited a total of 78 false alerts (Ffoster: 41, McBaine: 37) and the stand-stare dogs a total of 48 false alerts (Bobbie: 34, Osa: 14). The sit dogs hesitated on samples only 15 times (Ffoster: 8, McBaine: 7) whereas the stand-stare dogs hesitated a total of 72 times (Bobbie: 42, Osa: 30).

Duration of True Positive and False Positive Alerts

A significant interaction was found between alert behavior (sit vs. stand-stare) and alert type (false positive vs. true positive) ($t = 4.07$, $p < 0.001$). The model was further split to compare true and false positive durations between alert behavior group (sit or stand-stare). A non-significant difference was found between the stand-stare dog's mean duration of true and false positive alerts ($t = -1.24$, $p = 0.223$). Bobbie had a mean duration of 2 s for true alerts (min = 0.5, max = 3.7 s), and 2 s for false alerts (min = 0.5, max = 3.3 s). Osa's true positive alerts were on average 1.1 s (min = 0.3, max = 2.1 s) and false positive alerts were 1.4 s (min = 0.8, max = 2.3 s). Conversely, sit dogs showed a significant difference in the duration of their true positive alerts as compared to their false positive alerts ($t = -7.179$, $p = <0.001$) (**Figure 2**). Foster had a mean duration of 1 s for true positive alerts (min = 0.4, max = 1.5), and 2.3 s for false positive alerts (min = 1.3 s, max = 6.2 s). McBaine's true positive alerts were on average 1.1 s (min = 0.5, max = 3.1 s) and false alerts on average 2 s (min = 1.4, max = 4.1 s).

Duration of Contact With Port

Dogs that show a sit alert spent on average 0.3 s in contact with the port, regardless of whether it contained a distractor, control, normal, benign or cancer sample (**Figure 3**). In contrast, dogs in the stand-stare group showed an increase in the duration spent in contact with the port, with a mean duration of 0.3 on non-human odor samples (distractors and controls), 0.5 s on normal samples, 0.6 s on benign samples and 1.5 s on cancer samples. Differences in the mean duration of contact with the port between the sit and stand-stare dogs were approaching significance (**Figure 3**).

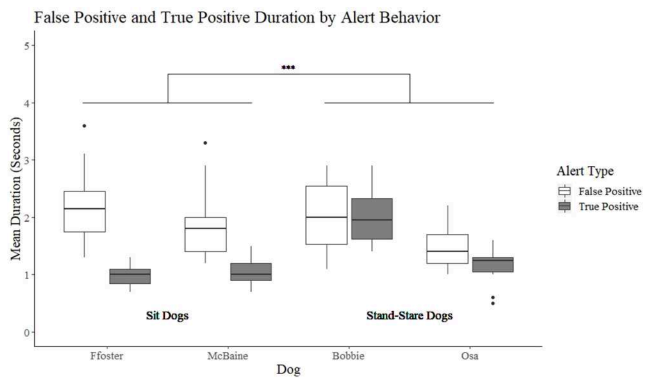


FIGURE 2 | The duration of the sit and stand-stare dogs' true positive and false positive alerts. ***Indicates a significant difference at $p < 0.001$.

DISCUSSION

Across 400 trials we see individual variation in the number of false alerts each dog performed. Bobbie (stand-stare) showed a similar number of false alerts to McBaine and Ffoster (sit alert). In contrast, Osa (stand-stare) carried out only 14 false alerts over her 100 trials. Given the small number of dogs sampled, a direct association between alert behavior and a dog's overall ability at this task cannot be made. It should also be noted that Osa had a more extensive training history where benign and normal were present but not rewarded [see (9) for details]. This may have contributed to her increased proficiency at the task overall. Of interest is the differences between sit and stand-stare dogs in their number of hesitations on the samples. Sit dogs showed a total of 15 hesitations over 200 trials, whereas the stand-stare dogs hesitated 72 times. These results suggest that the type of operant behavior required to signal an alert may impact on a dog's behavior while checking samples. While Mancini et al. (19) label hesitations as a "breakdown in communication" it could conversely be interpreted as gathering further information on a sample. For example, Mancini et al. (19) highlight the need for dogs to classify samples as "positive, negative or in-between." As long as strict criterion for reward marking is maintained (e.g., the dog must "freeze" to signal a final response) then it could be argued that a stand-stare alert allows hesitations on a sample to signify this "in-between" response. While hesitations are inherently ambiguous in terms of classifying the sample, it is also important in such a sensitive discrimination task that the dog can check the sample for as long as necessary to make an informed decision. As Mancini et al. (19) highlight, it is possible that dogs become more focused on performing their learnt behavior than on the stimulus coming from the sample. Perhaps within this argument however there are degrees of effect dependent on what the learnt behavior is (e.g., sit or stand-stare).

It could be assumed that a dog's false positive alert would be approximately the same length as their true positive alert. The duration of a true positive alert will be determined by the trainer, as it ends once the marker cue is given. For example, Bobbie showed the longest true positive alert mean duration (2 s) as her trainer used a criterion that Bobbie must be frozen in a true alert for between 1.5 and 3 s before using the clicker. If there was no effect of alert behavior on alert duration, it would be hypothesized that the dog would merely wait for a period of time approximate to when they usually hear their marker cue (the clicker), then move on if they do not hear the cue. By comparing each dog's false positive alert length to their true positive alert length, we were able to assess whether all dogs showed an approximately equal length of true and false positive, or if there was potential impact of alert type on false alert duration. We find that the stand-stare dogs conformed to this hypothesis, with Osa showing a difference of 0.3 s, and Bobbie a difference of 0 s, between their true and false positive alerts. In contrast, the sit dogs carried out false positive alerts for approximately double the duration of their true positive alerts, even though there was never an effort made by their trainers to increase their sit duration for their true positive alert. This may be influenced by the fact that they are no longer getting feedback from the odor source. It is possible that, because a stand-stare behavior requires a dog to keep their nose on the sample, a stand-stare trained dog can continue to receive information from the sample and may move on more quickly from an incorrect response than a sit alert trained dog who has, in carrying out their alert, created more distance from the sample.

For all four dogs, the mean duration of contact with the port was 0.3 s for non-human odor samples (distractors and controls). The sit dogs maintained this mean duration of 0.3 s across all samples, including human odor, whereas the stand-stare dogs elicited a mean duration of 0.5 s for normal samples, 0.6 s for benign samples and 1.5 s for cancer samples. It is not surprising that, for the stand-stare dogs, the longest duration was seen on cancer samples, as their alert behavior includes them making contact with the port. Of particular interest, however, is the increased duration on benign and normal (healthy control) samples. The stand-stare dogs show an increase of duration of contact when a plasma sample of any type is in the port, which may contribute to the increased number of hesitations seen in this group. Though we did not investigate sniffing rates here, Concha et al. (23) found that sniffing behavior in working detection dogs varied between true negatives and other odors. They found that true negatives saw the least number of sniffs by the dogs, compared to true positives, false positives, and false negatives, which elicited twice the number of sniffs. This initially seems to contradict our findings, where the stand-stare dogs spent more time in contact with the port when the odor was a plasma odor regardless of its cancer status (normal, benign, or cancer-positive), even when the dog left the port, marking a true negative. However, the Concha et al. (23) study investigated detection dogs working on the presence or absence of an odor, without controls of similar odor profiles, as seen in cancer detection dogs comparing blood plasmas of different cancer statuses. Nonetheless, future studies should investigate actual sniffing behavior to determine whether time spent with nose on port, prior to and during an alert, are true indicators of more sniffing.

Stand-stare and sit dogs differ in two main ways. Firstly, the sit dogs have the addition of a chained, arbitrary behavior to add on once they have located the target odor (the sit). Secondly, the sit dogs take their nose off the sample to carry out the alert

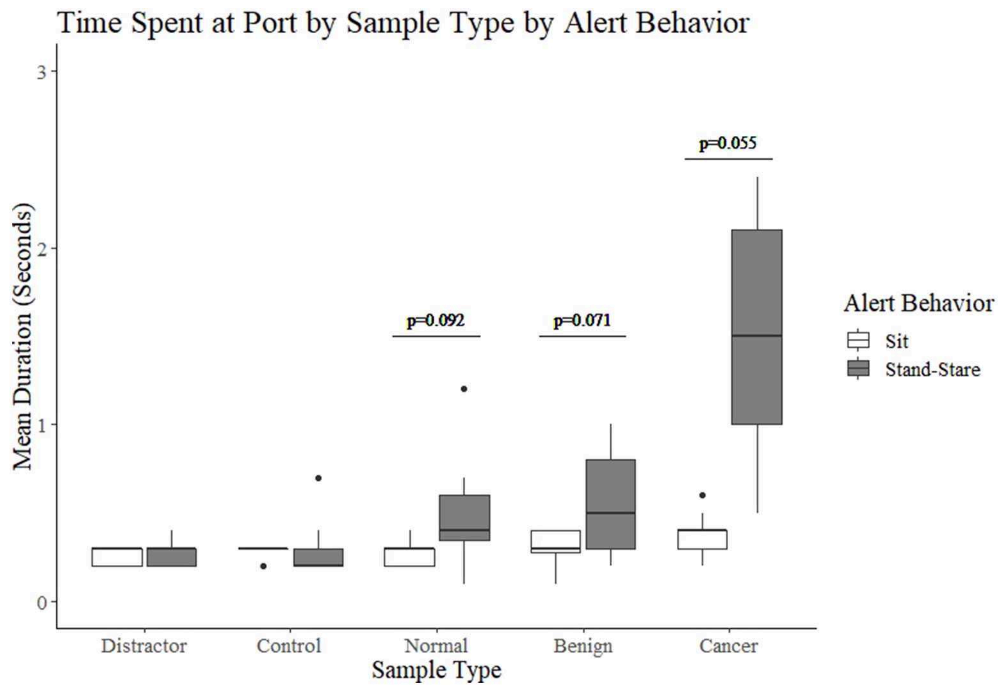


FIGURE 3 | The duration of contact with the port for sit and stand-stare dogs per sample type.

response. Finally, it is important to consider that the stand-stare is similar to a natural "pointing" behavior seen in many dogs and specifically selected for in some breeds (24). Thus, there may in fact be an advantage to using a more naturalistic behavior, that is often seen in response to odor already, rather than adding an arbitrary sit behavior. To disentangle whether one of these aspects may be influencing a dog's behavior more than the other, future studies may wish to include dogs who carry out sit alerts while keeping their nose on the sample. It is possible that training dogs with a sit alert to either keep their nose on the target odor, resulting in more of a "sit-stare" alert, or to engage in more sniffing behavior, may convey to these dogs the same potential advantages seen by the stand-stare dogs in this study.

While this study cannot disentangle whether the results are most influenced by the addition of the unnatural sit behavior or a by-product of their alert including them taking their nose off the sample, the reduced time spent checking each sample regardless of type indicates that perhaps the mere anticipation of carrying out a behavior which involves taking their nose off the port reduces the duration of time spent checking. Given the sensitive nature of the task and the low odor thresholds involved (up to parts per trillion), it may be most prudent to employ a system which does not limit a dog's interaction with the sample, such as training an alert which involves them moving away from the sample itself. While arguably ambiguous, hesitations may, in fact, further provide more information on a sample that a binary pass/alert response would fail to communicate. In training a stand-stare alert, it is important to establish a "freeze" to mark out the final response behavior. In doing so, the dog is able to check the sample for a greater amount of time prior to making their final response. The results of this study indicate that a stand-stare alert may facilitate this process to a greater extent than an operant response that involves the dogs moving off the sample.

It must be considered that these results were carried out on a limited sample of dogs. This is unfortunately a field-wide issue, as multiple laboratories test different human diseases, often with limited access or resources to train a sizeable sample of dogs. For example, several articles in this field offer important proof-of-principle data, but involve only a single canine [e.g., (12, 15)]. A lack of access to a large sample of trained dogs limits the scope to assess aspects such as alert behavior on task performance. Previous research has shown that individual characteristics of dogs' impact on their accuracy on human disease detection tasks [e.g., (25)], therefore a larger sample size would be needed to corroborate that these findings are related to the alert type rather than individual differences. However, results within the two "sit-stare" dogs were consistent to each other, and similarly results within the two "stand-stare" dogs, suggesting that there were effects of alert style as opposed to random variation between individual dogs. To compensate for the limited access to a wider pool of trained dogs, a larger number of trials per dog was chosen to establish robust findings within-dog. If these preliminary results can be established on a larger sample of trained dogs, there could be important applications to the field.

It is currently commonplace to allow the dog's preference to guide their final alert behavior, as it was previously thought that, within operant trained responses, alert type does not impact task behaviors. The results of this study indicate otherwise. While passive alerts may be ideal in other detection dog roles, for example in providing a non-ambiguous response at great distances, in a laboratory setting a sit response may be sub-optimal. Biomedical detection dogs are tasked with comparing multiple odor sources, many with a similar odor profile, in close proximity to one another. In a line-up of eight human samples, where, for example, four are different healthy controls, three are from people with benign tumors and one is a cancer positive sample, the level of specificity needs to be extremely high. When considering further that the dog may be given as little as 50 μl of sample, it may be beneficial to intentionally train an operant response that, by definition, includes the dog keeping their nose on the sample longer. This may allow dogs to make more informed decisions as a product of them having an additional motivation to keep their nose on the sample. It may also reduce the likelihood of the dog making incorrect decisions without the ability to change response because they have moved away from the sample and are no longer able to get feedback from it. It should be considered that both sit and stand-stare alerts are still operant behaviors that need to be shaped and trained in a similar way. However, without the means to use "honest signaling" (e.g., using technology to measure non-trained responses to a sample), a stand-stare alert may offer trainers a more truthful method of communication than a sit response.

CONCLUSION

Currently in biomedical detection research a sit alert final response is most commonly used. Until now, it was widely considered that operant alert type would not impact on task-related behaviors. This study suggests that alert type may influence the duration of a dog's false positive alert, and the amount of time spent checking a sample. Individual differences in the total number of false alerts recorded prohibits judgment on whether alert type directly affects task accuracy. Given the potential lack of feedback available once a dog has sat back away from the sample, it is possible that training a stand-stare alert instead may provide more information to the canine and assist in their categorization of the sample.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Pennsylvania Institutional Review Board, Approval # 702679. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by University of Pennsylvania Institutional Animal Care and Use Committee.

AUTHOR CONTRIBUTIONS

JE, CW, AV, RF, and CO planned the study, edited and finalized the manuscript. JE, CW, AV, and RF collected the data. CW and AV analyzed the videos. JE analyzed the data. JE and CW drafted the original manuscript. CO acquired funding for the project.

FUNDING

The following agencies funded this project: the Kleberg Foundation, USA; Ovarian Cancer Symptom Awareness (OCSA); Kaleidoscope of Hope; the Kahn Foundation; and Monell Chemical Senses Center. RF was funded by a summer research grant provided by OCSA. CW was funded by a summer research grant funded by Boehringer Ingelheim. Bobbie was sponsored by Merck Animal Health. Ffoster was sponsored by the Kaleidoscope of Hope Foundation. McBaine was sponsored by Saint Germain Catering. Osa was sponsored by Air Animal Pet Movers.

ACKNOWLEDGMENTS

The authors would like to thank the funders of the project, as well as all volunteers, interns, and trainers who have helped on this project along the way.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.00118/full#supplementary-material

Supplementary Table 1 | Data used for this study.

REFERENCES

1. Willis CM, Britton LE, Harris R, Wallace J, Guest CM. Volatile organic compounds as biomarkers of bladder cancer: sensitivity and specificity using trained sniffer dogs. *Cancer Biomarkers.* (2010) 8:145–53. doi: 10.3233/CBM-2011-0208
2. McCulloch M, Jezierski T, Broffman M, Hubbard A, Turner K, Janecki T. Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. *Integr Cancer Ther.* (2006) 5:30–9. doi: 10.1177/1534735405285096
3. Guerrero-Flores H, Apresa-García T, Garay-Villar Ó, Sánchez-Pérez A, Flores-Villegas D, Bandera-Calderón A, et al. A non-invasive tool for detecting cervical cancer odor by trained scent dogs. *BMC Cancer.* (2017) 17:1–8. doi: 10.1186/s12885-016-2996-4
4. Sonoda H, Kohnoe S, Yamazato T, Satoh Y, Morizono G, Shikata K, et al. Colorectal cancer screening with odour material by canine scent detection. *Gut.* (2011) 60:814–9. doi: 10.1136/gut.2010.218305
5. Walczak M, Jezierski T, Górecka-Bruzda A, Sobczynska M, Ensminger J. Impact of individual training parameters and manner of taking breath odor samples on the reliability of canines as cancer screeners. *J Vet Behav Clin Appl Res.* (2012) 7:283–94. doi: 10.1016/j.jveb.2012.01.001
6. Ehmann R, Boedeker E, Friedrich U, Sagert J, Dippon J, Friedel G, et al. Canine scent detection in the diagnosis of lung cancer:

Revisiting a puzzling phenomenon. *Eur Respir J*. (2012) 39:669–76. doi: 10.1183/09031936.00051711
7. Amundsen T, Sundstrom S, Buvik T, Gederaas OA, Haaverstad R. Can dogs smell lung cancer? First study using exhaled breath and urine screening in unselected patients with suspected lung cancer. *Acta Oncol*. (2014) 53:307–15. doi: 10.3109/0284186X.2013.819996
8. Horvath G, Andersson H, Nemes S. Cancer odor in the blood of ovarian cancer patients: a retrospective study of detection by dogs during treatment, 3 and 6 months afterward. *BMC Cancer*. (2013) 10:643–9. doi: 10.1186/1471-2407-13-396
9. Murarka M, Vesley-Gross ZI, Essler JL, Smith PG, Hooda J, Drapkin R, et al. Testing ovarian cancer cell lines to train dogs to detect ovarian cancer from blood plasma: A pilot study. *J Vet Behav*. (2019) 32:42–8. doi: 10.1016/j.jveb.2019.04.010
10. Taverna G, Tidu L, Grizzi F, Torri V, Mandressi A, Sardella P, et al. Olfactory system of highly trained dogs detects prostate cancer in urine samples. *J Urol*. (2015) 193:1382–7. doi: 10.1016/j.juro.2014.09.099
11. Pickel D, Manucy GP, Walker DB, Hall SB, Walker JC. Evidence for canine olfactory detection of melanoma. *Appl Anim Behav Sci*. (2004) 89:107–16. doi: 10.1016/j.applanim.2004.04.008
12. Bomers MK, Van Agtmael MA, Luik H, Van Veen MC, Vandenbroucke-Grauls CMJE, Smulders YM. Using a dog's superior olfactory sensitivity to identify Clostridium difficile in stools and patients: Proof of principle study. *BMJ*. (2012) 345:1–8. doi: 10.1136/bmj.e7396
13. Davies JC, Alton E, Simbo A, Murphy R, Seth I, Williams K, et al. Training dogs to differentiate Pseudomonas aeruginosa from other cystic fibrosis bacterial pathogens: not to be sniffed at? *Eur Respir J*. (2019) 54:1900970. doi: 10.1183/13993003.00970-2019
14. Edwards TL, Browne CM, Schoon A, Cox C, Poling A. Animal olfactory detection of human diseases: Guidelines and systematic review. *J Vet Behav Clin Appl Res*. (2017) 20:59–73. doi: 10.1016/j.jveb.2017.05.002
15. Cornu JN, Cancel-Tassin G, Ondet V, Girardet C, Cussenot O. Olfactory detection of prostate cancer by dogs sniffing urine: A step forward in early diagnosis. *Eur Urol*. (2011) 59:197–201. doi: 10.1016/j.eururo.2010.10.006
16. Hardin DS, Anderson W, Cattet J. Dogs can be successfully trained to alert to hypoglycemia samples from patients with type 1 diabetes. *Diabetes Ther*. (2015) 6:509–17. doi: 10.1007/s13300-015-0135-x
17. Nolan RV, Gravitte DL. Mine-detecting Canines. (1977). Available online at: https://apps.dtic.mil/dtic/tr/fulltext/u2/a048748.pdf
18. Jezierski T, Walczak M, Ligor T, Rudnicka J, Buszewski B. Study of the art: canine olfaction used for cancer detection on the basis of breath odour. Perspectives and limitations. *J Breath Res*. (2015) 9:027001. doi: 10.1088/1752-7155/9/2/027001
19. Mancini C, Harris R, Aengenheister B, Guest C. Re-centering multispecies practices: a canine interface for cancer detection dogs. In: *Conference* on *Human Factors* in *Computing Systems-Proceedings*. (2015). p. 2673–82. Available online at: https://dl.acm.org/doi/proceedings/10.1145/2702123
20. R Core Team. *R: A Language and Environment for Statistical Computing*. (2018). Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/ (accessed August 15, 2019).
21. Bates D, Maechler M, Bolker BM, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. (2015) 67:1–48. doi: 10.18637/jss.v067.i01
22. Venables WN, Ripley BD. *Modern Applied Statistics with S, 4th Edn*. New York, NY: Springer (2002). Available online at: http://www.stats.ox.ac.uk/pub/MASS4
23. Concha A, Mills DS, Feugier A, Zulch H, Guest C, Harris R, et al. Using sniffing behavior to differentiate true negative from false negative responses in trained scent-detection dogs. *Chem Senses*. (2014) 39:749–54. doi: 10.1093/chemse/bju045
24. Parra D, Méndez S, Cañón J, Dunner S. Genetic differentiation in pointing dog breeds inferred from microsatellites and mitochondrial DNA sequence. *Anim Genet*. (2008) 31:1–7. doi: 10.1111/j.1365-2052.2007.01658.x
25. Gadbois S, Reeve C. The semiotic canine: scent processing dogs as research assistants in biomedical and environmental research. *Dog Behav*. (2016) 2:26–32. doi: 10.4454/db.v2i3.43

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 21 April 2020
doi: 10.3389/fvets.2020.00126

Non-invasive Assessment of Fecal Stress Biomarkers in Hunting Dogs During Exercise and at Rest

Augusta Zannoni[1,2], Marco Pietra[1], Alba Gaspardo[1], Pier Attilio Accorsi[1], Monica Barone[3], Silvia Turroni[3,4], Luca Laghi[4,5], Chenglin Zhu[5], Patrizia Brigidi[3,4] and Monica Forni[1,2]*

[1] Department of Veterinary Medical Sciences, University of Bologna, Bologna, Italy, [2] Health Sciences and Technologies—Interdepartmental Center for Industrial Research (CIRI-SDV), Alma Mater Studiorum—University of Bologna, Bologna, Italy, [3] Unit of Microbial Ecology of Health, Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy, [4] Interdepartmental Centre for Agri-Food Industrial Research, University of Bologna, Bologna, Italy, [5] Department of Agro-Food Science and Technology, Centre of Foodomics, University of Bologna, Cesena, Italy

OPEN ACCESS

Edited by:
Cynthia M. Otto,
University of Pennsylvania,
United States

Reviewed by:
Erin Beth Perry,
Southern Illinois University
Carbondale, United States
Federica Pirrone,
University of Milan, Italy

***Correspondence:**
Augusta Zannoni
augusta.zannoni@unibo.it

Specialty section:
This article was submitted to
Veterinary Humanities and Social
Sciences,
a section of the journal
Frontiers in Veterinary Science

Received: 29 October 2019
Accepted: 19 February 2020
Published: 21 April 2020

Citation:
Zannoni A, Pietra M, Gaspardo A,
Accorsi PA, Barone M, Turroni S,
Laghi L, Zhu C, Brigidi P and Forni M
(2020) Non-invasive Assessment of
Fecal Stress Biomarkers in Hunting
Dogs During Exercise and at Rest.
Front. Vet. Sci. 7:126.
doi: 10.3389/fvets.2020.00126

Intense exercise causes to organisms to have oxidative stress and inflammation at the gastrointestinal (GI) level. The reduction in intestinal blood flow and the exercise-linked thermal damage to the intestinal mucosa can cause intestinal barrier disruption, followed by an inflammatory response. Furthermore, the adaptation to exercise may affect the gut microbiota and the metabolome of the biofluids. The aim of the present research was to evaluate the presence of a GI derangement in hunting dogs through a non-invasive sampling as a consequence of a period of intense exercise in comparison with samples collected at rest. The study included nine dogs that underwent the same training regime for hunting wild boar. In order to counterbalance physiological variations, multiple-day replicates were collected and pooled at each experimental point for each dog. The samples were collected immediately at rest before the training (T0), after 60 days of training (T1), after 60 days of hunting wild boar (T2), and finally, at 60 days of rest after hunting (T3). A number of potential stress markers were evaluated: fecal cortisol metabolites (FCMs) as a major indicator of altered physiological states, immunoglobulin A (IgA) as an indicator of intestinal immune protection, and total antioxidant activity [total antioxidant capacity (TAC)]. Since stool samples contain exfoliated cells, we investigated also the presence of some transcripts involved in GI permeability [occludin (OCLN), protease-activated receptor-2 (PAR-2)] and in the inflammatory mechanism [interleukin (IL)-8, IL-6, IL-1b, tumor necrosis factor alpha (TNFα), calprotectin (CALP), heme oxygenase-1 (HO-1)]. Finally, the metabolome and the microbiota profiles were analyzed. No variation in FCM and IgA content and no differences in OCLN and CALP gene expression between rest and training were observed. On the contrary, an increase in PAR-2 and HO-1 transcripts, a reduction in total antioxidant activity, and a different profile of microbiota and metabolomics data were observed. Collectively, the data in the present study indicated that physical exercise in our model could be considered a mild stressor stimulus.

Keywords: dog, exercise, stress markers, stool, welfare

INTRODUCTION

Intense exercise is known to exacerbate body stressors, such as oxidative stress and inflammation, the latter at both the muscular (1–3) and the gastrointestinal (GI) (4–6) levels. As a consequence, in performance sports, there is a high prevalence of GI problems both in humans, such as endurance runners (6–8), and in animals, such as horses (9, 10) or dogs (11). In a review paper by ter Steege et al. (12), several studies were cited that suggested that the key culprit behind GI symptoms during exercise was splanchnic hypoperfusion, which could lead to intestinal ischemia, thus subsequently damaging the intestinal epithelial cells and compromising the intestinal barrier function. Multiple studies involving humans have reported an exercise-induced increase in intestinal permeability (13). The tight junction (TJ) plays an important role in regulating the epithelial permeability by means of modifying the multiprotein complex [claudins and occludin (OCLN)] and/or promoting dysfunction to TJ regulatory proteins (i.e., zona-occludens) (14).

A downregulation of OCLN expression has been observed in different intestinal models, in which the permeability was strongly altered [i.e., inflammatory bowel disease (IBD), ulcerative colitis], and was downregulated (15, 16). Gut permeability is also influenced by protease-activated receptor-2 (PAR-2) expressed in the apical and basolateral membranes of intestinal epithelial cells (17). As described by a review (17), its activation induces an increase in permeability by means of impairment of the TJ functions, as shown in several epithelial and endothelial cell models (18–21). In different models including colitis and ischemia and reperfusion (I/R), PAR-2 transcription was upregulated in mouse, rodent, and horse models (21–23). Other markers of intestinal inflammation are calprotectin (CALP) and pro-inflammatory cytokines, which have been shown to be upregulated in IBD models (24–26).

Heme oxygenase-1 (HO-1) is an inducible cytoprotective stress-responsive protein induced by various stimuli, including oxidative stress I/R, heavy metals, and cytokines (27), the induction of which is usually associated with antioxidant, anti-apoptotic, and anti-inflammatory effects as reported by a review paper (28). In studies using murine experimental colitis models, HO-1 activity and expression were markedly increased, associated with the development of colitis, and the inhibition of HO activity potentiates colonic damage and inflammation (29, 30). Moreover, the relationship between physical exercise and increased HO-1 mRNA and protein expression/activity in different cells and tissues has already been demonstrated in rodents (31–34) as well as in humans (35, 36).

Cortisol is a well-known indicator of the stress response in the majority of mammals including dogs, with previous studies showing increased levels after exercise, such as agility work (37) and training in outdoor conditions (38, 39).

Many factors contribute to the maintenance of GI homeostasis. One of them is the secretion of immunoglobulin A (IgA), which coats the bacteria, favoring a tolerant, non-inflammatory relationship with the host (40) and the homeostatic control of the intestinal redox environment (41). Previous papers have reported that exercise may affect the levels of IgA in mice (42) and cause oxidative stress in dogs (43).

Emerging research has suggested that intense exercise could also affect the gut microbiota. In particular, cross-sectional studies have shown an overall increase in biodiversity with some compositional alterations, mainly in mucin degraders, lactate utilizers, and short-chain fatty acid (SCFA) producers, in the intestinal microbial ecosystem of professional athletes (44). Several factors are likely to be involved, including changes in diet, hydration levels and metabolic flux, altered gut motility, and also impaired gut barrier function, as a result of exercise-induced heat stress and ischemia (44). Given the fundamental role of the gut microbiota in maintaining host metabolic and immunological homeostasis (45), its monitoring during periods of intense physical activity could help to elucidate the mechanisms underlying the microbial response to exercise and understand if and how these are related to host performance.

The metabolome of fluids, which is made up of the ensemble of low-weight organic molecules, results from a complex interaction between endogenous and exogenous host factors, including the gut microbiota. As such, it has been shown to give important information regarding the overall effects of exercise in both humans and animals, with specific reference to the inflammatory status. The fecal metabolome seems to be no exception, at least in rats (46).

The exfoliated enterocytes contained in feces have recently been used as a tool to investigate the impact of therapies and nutritional regimens on GI functions (47, 48). In fact, stool is easy to obtain and has already been used in quantifying intestinal gene expression profiles from exfoliated epithelial cells in neonates (49, 50), as well as under pathological conditions to detect candidate molecular biomarkers (51–53). Exercise induces multiple biochemical changes, which may affect the gene expression of the transcripts involved in the mitochondrial metabolism in muscle (54) and oxidative stress, as assessed non-invasively (i.e., in saliva) in avalanche military dogs (55).

The aim of the present research was to evaluate the presence of a GI derangement in hunting dogs through a non-invasive sampling as a consequence of a period of intense exercise in comparison with samples collected at rest. To reach this goal, we selected a number of potential stress markers in fecal samples, including cortisol metabolites [fecal cortisol metabolite (FCM)], transcripts involved in epithelial integrity and inflammatory mechanisms [cytokines: interleukin (IL)-8, IL-6, IL-1β, and tumor necrosis factor alpha (TNFα); OCLN; CALP; PAR-2; and HO-1], IgA, and total antioxidant capacity (TAC) levels. Furthermore, we decided to profile the fecal metabolome, by means of high-resolution proton magnetic resonance spectroscopy (^{1}H-NMR), and the microbiota, by 16S rRNA gene-based next-generation sequencing.

MATERIALS AND METHODS

Experimental Design and Exercise

Four experimental time points were set: T0, after 180 days of complete rest (rest before the training session, September); T1,

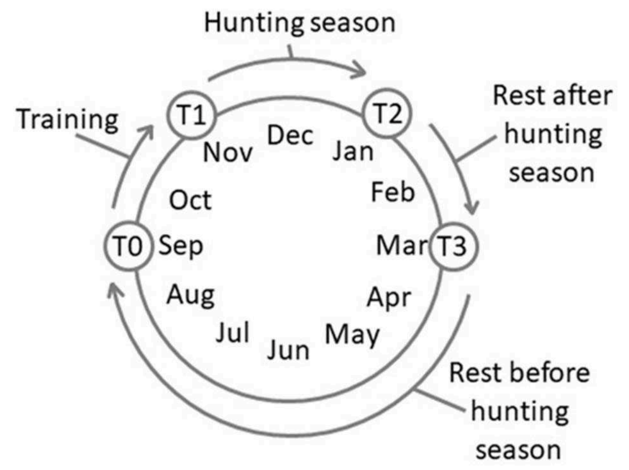


FIGURE 1 | Schedule of the experimental time points. T0, rest before the training; T1, 60 days of training; T2, 60 days of hunting season; T3, 60 days of rest after hunting.

after 60 days of training, 3 days a week, 3 h each day (November); T2, after 60 days of wild boar hunting three times a week, 5–6 h each day (January); and T3, after 60 days of complete rest (rest after hunting season, March) (**Figure 1**). The physical activity carried out during both the training (T1) and the hunting (T2) periods was similar and consisted of a first phase of identifying and locating prey and a subsequent chase phase. The duration of these phases, due to the nature of the hunting itself, varied and was therefore impossible to standardize. All the dogs equally and simultaneously participated in each training/hunting session The training activity occurred on alternative days and was always conducted by the same person, the owner (not a professional trainer but an expert hunter fully aware of the goal of the research project), without any type of reinforcement.

Animals

The exploratory study was carried out from September 2017 to March 2018 on nine hunting dogs. The dogs were of various ages (9.1 ± 5.0 years; mean ± SD), sexes (two unneutered males and seven spayed females), and breeds (three English Setter, three Segugio Italiano, two Dachsbacke, one Deutsch Kurzhaar) (**Table 1**). T0 body weight (BW) (19.3 ± 3.3 kg; mean ± SD) and the body condition score (BCS, calculated by using the 1–9 score proposed by Royal Canine SAS) are reported in **Table 1**. BW and BCS were also determined at each experimental point. The dogs, owned by a single owner, were housed in individual boxes and fed, once a day, with a commercial diet (Eko Adult, Russo Mangimi SpA, NA, Italy): crude protein 22%, crude fats and oils 9%, crude fiber 4.6%, and crude ash 11.2%. The food was administered in relation to the weight of the dog and to physical activity, increasing the dose by about 15% in T1 and T2 with respect to the rest periods. All the dogs underwent a physical examination by a veterinarian at the beginning of and during the trials. Only those who were clinically healthy were included in the study.

TABLE 1 | Dogs included in the study.

Dog	Breed	Gender	Age	BW (kg)	BCS 1–9
1	Segugio Italiano	SF	4	18	6
2	Dachsbracke	M	14	20	6
3	Deutsch Kurzhaar	SF	4	25	5
4	English Setter	M	13	22	6
5	Segugio Italiano	SF	7	18	4
6	Segugio Italiano	SF	16	20	5
7	Dachsbracke	SF	2	13	5
8	English Setter	SF	12	18	5
9	English Setter	SF	10	20	6

M, unneutered male; SF, spayed female; BW, body weight; BCS, body condition score at T0 (rest before training).

Collection of the Fecal Sample

The samples were collected during the last week of each experimental period. In order to counterbalance the physiological fluctuations that occur within individuals, three samples for each time point were collected on different days. Specifically, at T1 and T2, the three samples were collected during the last week of physical activity on the day after the exercise session, while at T0 and T3, the three samples were collected on 3 consecutive days. The sample collection time was the same at each experimental point (after feeding in the late afternoon).

In agreement with the Italian law transposition of European Directive 2010/63 (DL 26/2014), the collection of fecal samples is not classified as a procedure, and it did not require any kind of authorization. This non-invasive sampling method was performed without any discomfort for the animals.

In total, 108 samples were collected: three for each dog at each of the four experimental times. The aforementioned three samples were pooled for the assays, leading to an overall sample number of 36 (one for each dog at each experimental time point).

Fresh fecal samples were collected by the owner within 1 h of defecation (late afternoon) and immediately stored at −20°C until analysis.

RNA Extraction and Reverse Transcription

Lyophilized fecal samples (Modulyo EF4 1044, Edwards) were weighed and resuspended with Dulbecco's phosphate-buffered saline (DPBS) (w/v; 100 mg/ml) by vortex mixing (3 min). RNA extraction was performed using PureZol RNA isolation reagent (BioRad, Bio-RAD Laboratories Inc., California, USA) and a NucleoSpin RNA II kit (Macherey Nagel, Duren, Germany). Briefly, 1 ml of PureZol RNA isolation reagent was added to 100 μl of each sample and vortex mixed (3 min). Two hundred microliters of chloroform was then added to the suspension and mixed well. After incubation at room temperature (10 min), the samples were centrifuged (12,000 g for 10 min), and the aqueous phase was recovered. One volume of ethanol was added, and the resulting solution was loaded onto a NucleoSpin RNA Column (light blue ring) (NucleoSpin RNA II kit, Macherey Nagel). The RNA was then purified according to the manufacturer's instructions and spectrophotometrically quantified (A260 nm) (DeNovix Inc., Wilmington, DE, USA). RNA (1 µg) was then reverse-transcribed to cDNA using an iScript cDNA Synthesis Kit (Bio-RAD), arriving at a final volume of 20 µl. An additional sample of canine intestinal biopsy, collected from the duodenum of a dog with IBD (derived from a diagnostic procedure, performed at DIMEVET, with the express consent of the owner; endoscopy code 9290, March 19, 2018, sample code 14873), underwent RNA extraction, reverse transcription, and subsequent analysis (quantitative real-time PCR assay) as a positive control of inflammatory gene expression.

Quantitative Real-Time PCR

Real-time quantitative PCR was carried out using a CFX 96 Real Time System (Bio-RAD) and SsoAdvanced™ Universal SYBR® Green Supermix (Bio-RAD). All the samples were analyzed in duplicate (10 µl/well), and the qPCR assays were carried out for different references [glyceraldehyde-3-phosphate dehydrogenase (GAPDH), TATA-box binding protein (TBP), tight junction protein 1 (TJP1), ribosomal protein L32 (RPL32), succinate dehydrogenase (SDHA), and interest genes (Il-8, IL-1β, IL-6, TNFα, OCLN, CALP, PAR-2, HO-1]. Primer sequences are reported in **Table 2**.

Real-time efficiency was evaluated by amplification of a standardized amount of cDNA, starting from 150 ng with subsequent 5-fold dilutions (75, 15, 3, 0.6, and 0.12 ng), derived from both fecal sample–derived and intestinal cDNA (duodenal biopsy). The specificity of the amplified PCR products was verified by analysis of the melting curve and agarose gel electrophoresis. The relative gene expression was calculated as the fold increase using the $2^{-\Delta\Delta Ct}$ method (58) in relation to T0 ($\Delta\Delta Ct = \Delta Ct_{\text{T1 or T2 or T3 group}} - \Delta Ct_{\text{T0}}$).

Fecal Cortisol Metabolites Determination

Extraction from the feces was performed as previously described (59). Briefly, a methanol:water (v/v 4:1) solution was added to the feces in capped glass tube vials. The vials were then vortex mixed for 30 min using a multitube pulsing vortexer. Following centrifugation (1,500 g for 15 min), ethylic ether and $NaHCO_3$ (5%) were added to 1 ml of supernatant. This preparation was then vortex mixed for 1 min on a multitube pulsing vortexer and centrifuged for 5 min (1,500 g). The ether portion was then separated and evaporated to dryness under an air-stream suction hood at 37°C; finally, the dry residue was dissolved into phosphate-buffered saline (PBS) 0.05 M, pH 7.5.

TABLE 2 | List of primer pairs, amplicon size (bp), and accession number (AN) in the NCBI (National Center of Biotechnology Information) database.

Gene		Primer sequence (5′->3′)	PCR (bp)	AN	References
HO-1	F	GCCAGTGCCACGAAGTTC	164	NM_001194969	Present study
	R	TCCTCAGTGTCCTTGCTCAG			
CALP	F	ACCATGCTGACGGAACTGGAGAG	244	NM_001146144	Present study
	R	CCACGCCCACCTTTATCACCAATATG			
OCLN	F	CAGAGTCTTCCTATAAATCAAC	196	NM_001003195.1	Present study
	R	GTGTAGTCTGTCTCATAGTG			
PAR-2	F	TGAAGATCGCCTACCACATCCG	137	AB_458680	(56)
	R	CCAATACCGTTGCACACTGA			
IL-8	F	CTTCCAAGCTGGCTGTTGCTC	173	NM_001003200	(56)
	R	TGGGCCACTGTCAATCACTCTC			
IL-1β	F	GCTGCTGCCAAGACCTGAAC	112	XM_005630074	Present study
	R	GCTACAATGACTGACACGAAATGC			
TNFα	F	CCCAAGTGACAAGCCAGTAGCTC	146	NM_001003244	(56)
	R	ACAACCCATCTGACGGCACTATC			
IL-6	F	AAAGAGCAAGGTAAAGAATCAGGATG	126	NM_001003301	Present study
	R	CGCAGGATGAGGTGAATTGTTG			
GAPDH	F	TGTCCCCACCCCCAATGTATC	100	NM_001003142	(57)
	R	CTCCGATGCCTGCTTCACTACCTT			
TBP	F	CTATTTCTTGGTGTGCATGAG G	96	XM849432	(56)
	R	CCT CGG CATTCAGTCTTTTC			
TJP1	F	GCTGTGGAAGAAGATGAAGATG	175	NM_001003140	Present study
	R	CTCGGCAGACCTTGAAGTAG			
RPL32	F	GGCACCAGTCAGACCGATATG	209	NM_001252169	Present study
	R	GCACATCAGCAGCACTTCAAG			
SDHA	F	CGCATAAGAGCCAAGAAC	194	XM535807	Present study
	R	CCTTCCGTAATGAGACAAC			

HO-1, heme oxygenase-1; CALP, calprotectin; OCLN, occludin; PAR-2, protease-activated receptor-2; IL, interleukin; TNFα, tumor necrosis factor alpha; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; TBP, TATA-box binding protein; TJP1, tight junction protein 1; RPL32, ribosomal protein L32; SDHA, succinate dehydrogenase.

Radio immunological assay (RIA) was carried out according to Tamanini et al. (60). Analysis was carried out in duplicate. The parameters for analysis validation were: sensitivity 0.23 pg/mg; intra-assay variability 6.4%; inter-assay variability 9.7%; and specificity (%) of cortisol 100, corticosterone 9.5, 11α-hydroxy-progesterone 8.3, cortisone 5.3, 11α-deoxycortisol 5.0, progesterone 0.6, deoxycorticosterone 0.5, 20α-dihydrocortisone 0.4, testosterone 0.3, aldosterone 0.1, and dehydroepiandrosterone, 5α-pregnenolone, 17β-estradiol, and cholesterol $<$0.0001.

Determination of Total IgAs and TAC

The IgA extraction was performed essentially as reported by Peters et al. (61).

Briefly, the lyophilized fecal samples were placed in 1 ml (w/v; 100 mg/ml) of extraction buffer (PBS containing 0.5% Tween 20 (Sigma-Aldrich, St. Louis, MO, USA) and a protease inhibitor cocktail (Sigma, 1 tablet/25 ml), and after the addition of three 3 mm glass beads, the samples were homogenized for 1 min with TissueLyser (50 Hz) (QIAGEN, Hilden, Germany). The homogenates were then centrifuged (1,500 g for 15 min), and the recovered supernatants were additionally centrifuged (15,000 g for 20 min). The supernatants were frozen at −20°C until analysis.

The IgA level was measured by a specific enzyme-linked immunosorbent assay (ELISA) kit (Dog IgA ELISA Quantitation Set, Bethyl Laboratories Inc., Montgomery, TX, USA). The analyses were carried out in duplicate. The parameters for analysis validation were: intra-assay variability 2.1% and inter-assay variability 12.8%. After checking the parallelism ($R^2 = 0.9849$, unpublished data), we diluted the sample 1:75,000 and carried out the assay according to the manufacturer's instructions.

The TAC level was assayed by using an Antioxidant Assay Kit (item no. 709001; Cayman Chemical Company, Ann Arbor, MI, USA) according to the manufacturer's instructions and expressed as a Trolox equivalent.

Metabolomics

The fecal samples were prepared for ^{1}H-NMR analysis by vortex mixing for 5 min (80 mg of stool with 1 ml of deionized water). The mixtures were then centrifuged for 15 min at 18,630 g and 4°C. The supernatants (700 μl) were added to a D_2O solution of 3-(trimethylsilyl)-propionic-2,2,3,3-d4 acid sodium salt (TSP) 10 mM and NaN_3 2 mM, set at pH 7.00 ± 0.02 with 1 M potassium phosphate buffer. Before analysis, the samples were centrifuged again at the above conditions.

The ^{1}H-NMR spectra were recorded at 298 K using an AVANCE III spectrometer (Bruker, Milan, Italy) operating at a frequency of 600.13 MHz. In accord with Ventrella et al. (62), the signals from broad resonances originating from large molecules were suppressed by a Carr–Purcell–Meiboom–Gill (CPMG) filter composed by 400 echoes with a τ of 400 μs and a 180° pulse of 24 μs, for a total filter of 330 ms. The HOD residual signal was suppressed by means of pre-saturation. Each spectrum was acquired by summing up 256 transients using 32 K data points over a 7,184 Hz spectral window, with an acquisition time of 2.28 s. To apply NMR as a quantitative technique (63), the recycle delay was set to 5 s, taking into consideration the relaxation time of the protons under investigation. ^{1}H-NMR spectra were baseline-adjusted by means of the peak detection according to the "rolling ball" principle (64) implemented in the baseline R package (65). A linear correction was then applied to each spectrum, so as to make the points pertaining to the baseline randomly spread around zero. Spectra have been horizontally aligned by employing the signal of TSP as a reference. The differences in water and fiber content among the samples were taken into consideration using probabilistic quotient normalization (66), applied to the entire spectra array.

The signals were assigned by comparing their chemical shift and multiplicity with the Human Metabolome Database (67) and Chenomx software data bank (Chenomx Inc., Canada, version 8.1).

Microbial DNA Extraction and 16S rRNA Gene Sequencing

Microbial DNA was extracted from the fecal samples using the DNeasy Blood & Tissue kit (QIAGEN), with a modified protocol as previously described (68). Briefly, 250 mg of feces were resuspended in 1 ml of lysis buffer (500 mM NaCl, 50 mM Tris-HCl pH 8, 50 mM EDTA, 4% SDS). Four 3 mm glass beads and 0.5 g of 0.1 mm zirconia beads (BioSpec Products, Bartlesville, OK) were added to the fecal samples and homogenized with three bead-beating steps using the FastPrep instrument (MP Biomedicals, Irvine, CA) at 5.5 movements/s for 1 min, keeping the samples on ice for 5 min after each treatment. The samples were heated at 95°C for 15 min and centrifuged for 5 min at 13,000 g to pellet stool particles. The supernatants were collected, and 260 μl of 10 M ammonium acetate was added; the samples were then incubated on ice for 5 min and then centrifuged for 10 min at 13,000 g. One volume of isopropanol was added, and the supernatants were incubated on ice for 30 min. The nucleic acids were collected by centrifugation for 15 min at 13,000 g and washed with 70% ethanol. The pellets were then resuspended in 100 μl of Tris-EDTA (TE) buffer and treated with 2 μl of DNase-free RNase (10 mg/ml) for 15 min at 37°C. Protein removal and DNA purification using QIAamp Mini Spin columns (QIAGEN) were carried out according to the kit protocol. The DNA extracted was quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE).

For each sample, the V3–V4 region of the 16S rRNA gene was sequenced as previously reported (69). Briefly, the DNA was amplified using the S-D-Bact-0341-b-S-17/S-D-Bact-0785-a-A-21 primers (70) with Illumina overhang adapter sequences. PCR products of ∼460 bp were purified using a magnetic bead-based system (Agencourt AMPure XP; Beckman Coulter, Brea, CA), indexed by limited-cycle PCR using Nextera technology, and were additionally purified using Agencourt AMPure XP magnetic beads. Indexed libraries were pooled at an equimolar concentration, denatured, and diluted to 6 pmol/L before loading onto the MiSeq flow cell. Sequencing was carried out on an Illumina MiSeq platform using a 2 × 250 bp paired-end protocol, according to the manufacturer's instructions (Illumina, San Diego, CA). Sequencing reads were deposited in the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA; BioProject ID PRJNA 589580).

Bioinformatics and Statistical Analysis

Statistical analysis was carried out in R computational language (71). Differences among sampling points were assessed using the analysis of variance (ANOVA) test for repeated measures (P-value < 0.05 was considered statistically significant). Robust principal component analysis (rPCA) models were calculated as described by Hubert et al. (72), namely, by accepting an alpha value of 0.75. Differences in the mRNA data were evaluated using one-way ANOVA (P-value < 0.05 was considered statistically significant).

As for the gut microbiota analysis, raw sequences were processed using a pipeline combining PANDAseq (73) and QIIME 2 (74). High-quality reads were filtered and clustered into amplicon sequence variants (ASVs) at 99% similarity by means of an open-reference strategy carried out using dada2 (75). Taxonomy was assigned using the vsearch classifier (76) and the Greengenes database as a reference (release May 2013). Alpha rarefaction was carried out using Faith's phylogenetic index and the number of observed ASVs, while beta diversity was estimated by computing weighted and unweighted UniFrac distances. All the statistical analyses were carried out using R (version 3.1.3) and the packages vegan and made4. UniFrac distances were used for the principal coordinate analysis (PCoA), and the significance of data separation was tested using a permutation test with pseudo-F ratios (function adonis of vegan) and the ANOSIM test. The Wilcoxon test for paired data was used to assess significant differences in alpha diversity and taxon relative abundance between groups, while the Kruskal–Wallis test was used for multiple comparisons. A P-value < 0.05 was considered statistically significant.

RESULTS

Animals

In **Figure 2**, we report the variation in BW of the dogs during the trial.

The physical activity induced a statistically significant decrease of BW after 60 days of hunting season (T2, 16.9 ± 3.4) with respect to the rest periods (T0, 19.3 ± 3.3, and T3, 18.8 ± 3.2) ($P = 0.017$, repeated measures ANOVA, Tukey's multiple comparison test, $P < 0.05$). On the contrary, the training period did not significantly influence the BW (T1, 18.7 ± 3.6) (repeated measures ANOVA, Tukey's multiple comparison test, $P < 0.05$). The percentages of BW reduction at T1, T2, and T3 with respect to T0 were 3.2, 12.7, and 2.9%, respectively.

The BCSs of the dogs recorded during the trial were (median, min–max): T0 (4, 5, 5, 6); T1 (4, 5, 5, 6); T2 (3, 4, 4); and T3 (4, 5, 5, 6). Similarly to BW, only T2 (60 days after hunting season) was statistically different from rest periods (T0 and T3) and the period after 60 days of training (T1) (repeated measures ANOVA, Friedman test, Dunn's multiple comparison test, $P < 0.05$).

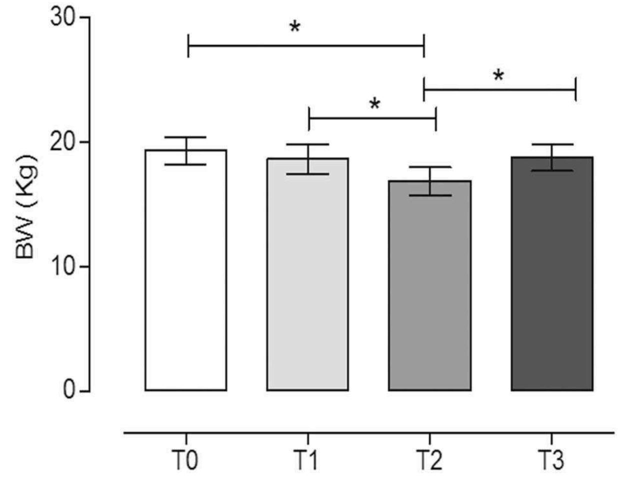


FIGURE 2 | Body weight (BW) of dogs at the different time points. The physical activity induced a statistically significant decrease in BW after 60 days of hunting season (T2) (mean ± SEM) ($P = 0.017$). *Indicates $P < 0.05$ (repeated measures ANOVA, Tukey's multiple comparison test, $P < 0.05$).

Real-Time Quantitative Reverse Transcription PCR for PAR-2, HO1, CALP, OCLN, IL-8, IL-1β, IL-6, and TNFα

RNA was extracted from all the samples with a yield of 336.35 ± 147.8 ng/10 mg dry feces. Of the reference genes analyzed, only GAPDH was always detectable; therefore, it was used as a reliable internal reference for qPCR normalization. To evaluate the matrix effect, we determined qPCR efficiency for GAPDH in the stool and tissue samples. The results showed that the efficiency was similar in both samples (97 and 91.7%, respectively) (**Figure 3**), indicating that RNA isolated from feces did not contain particular PCR inhibitors.

The presence and specificity of the PCR products were verified using melting curve analysis and agarose gel electrophoresis. The transcripts of GAPDH, HO-1, CALP, OCLN, and PAR-2 were detectable in the majority of the samples analyzed (GAPDH 33/36, HO-1 29/36, PAR-2 27/36, CALP 21/36, OCLN 26/36), although with a huge variability regarding the range of gene expression both between the dogs and regarding the time points.

The expression levels of OCLN and CALP did not show significant differences among groups ($P = 0.6338$ and $P = 0.1704$, respectively) (one-way ANOVA, Tukey's multiple comparison test, $P < 0.05$, **Figure 4**). On the contrary, a statistically significant increase was observed at T2 (after 60 days of hunting season) for PAR-2 and HO-1 as compared to T0 ($P = 0.042$ and $P = 0.028$, respectively) (one-way ANOVA, Tukey's multiple comparison test, $P < 0.05$, **Figure 4**). Very low or undetectable expression levels were observed for the genes encoding the cytokines (IL-8, IL-1β, IL-6, TNFα) (very low 7/36, undetectable 29/36) and for the other reference genes (TPB, TJP1, RPL32, SDHA) (very low 8/36, undetectable level 28/36).

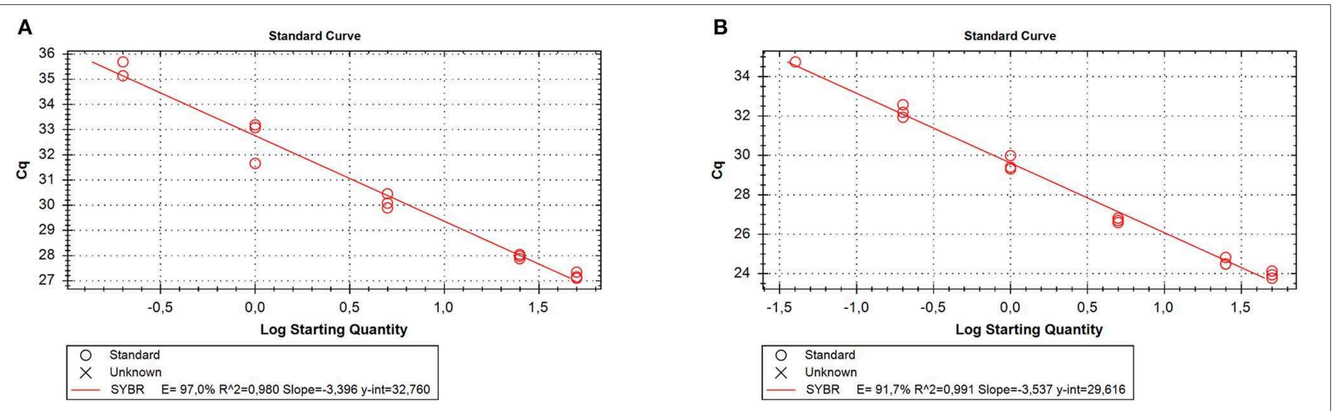


FIGURE 3 | Quantitative real time PCR (qRT-PCR) efficiency for the reference gene [glyceraldehyde-3-phosphate dehydrogenase (GAPDH)]. Five different 5-fold dilutions of the stool **(A)** or tissue **(B)** samples were assayed. Cq, cycle quantification; E, efficiency.

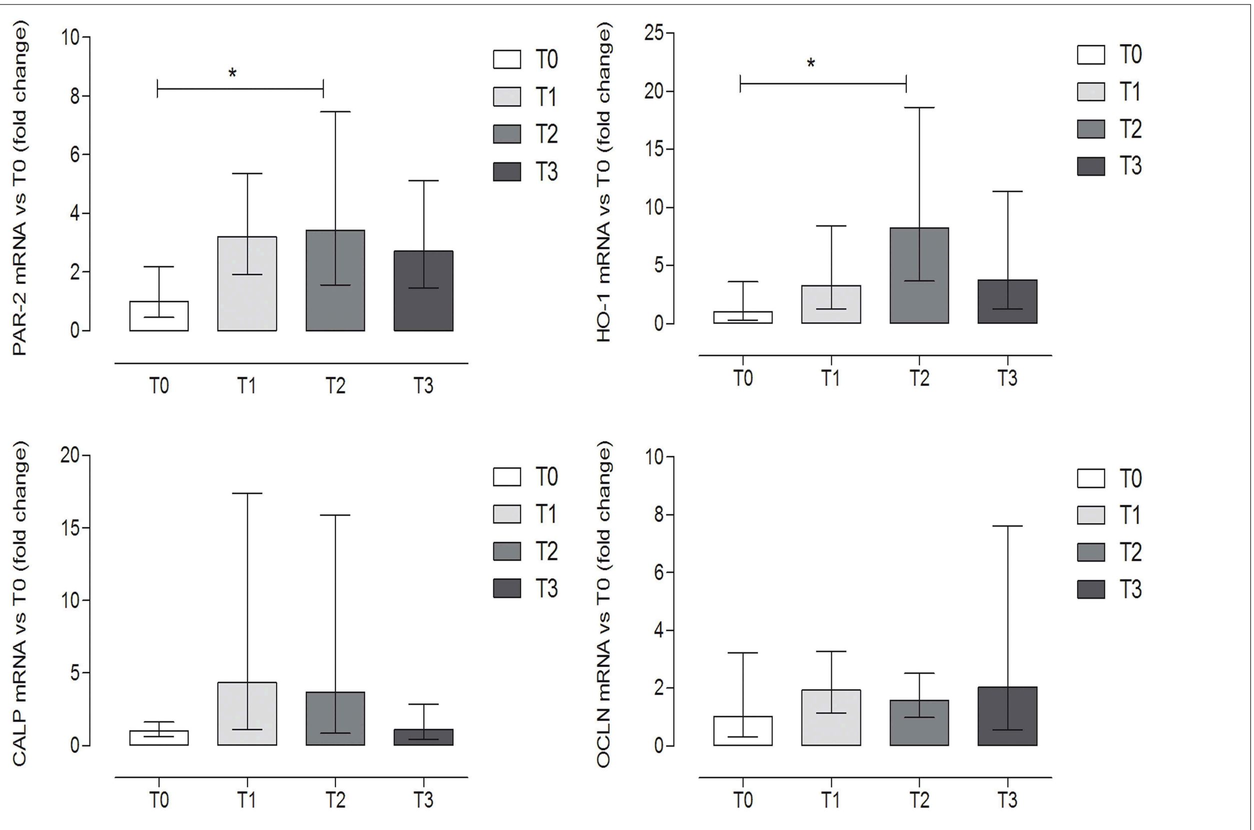


FIGURE 4 | Gene expression of protease-activated receptor-2 (PAR-2), heme oxygenase-1 (HO-1), calprotectin (CALP), and occludin (OCLN) evaluated by qRT-PCR, at the different time points. Relative gene expression of PAR-2, HO-1, CALP, and OCLN in the fecal samples of dogs at rest before training (T0), after 60 days of training (T1), after 60 days of hunting season (T2), and at 60 days of rest after hunting (T3). The mRNA data are expressed as fold change with respect to T0. *Indicates $P < 0.05$ (one-way ANOVA, $P < 0.05$, *post-hoc* Tukey's test). *P*-values: $P = 0.042$ for PAR-2; $P = 0.028$ for HO-1; $P = 0.1704$ for CALP; $P = 0.6338$ for OCLN. Error bars represent the range of gene expression.

FCM Determination

No statistically significant differences were observed in FCM content during the trial ($P = 0.270$) (repeated measures ANOVA, $P < 0.05$). The concentration of FCMs at T0 was 0.31 ± 0.03 pg/mg feces, while at T1, the level was 0.63 ± 0.29 pg/mg feces (**Figure 5**).

Determination of Total IgA in Stools

The IgA content in the canine fecal samples at the different time points is reported in **Figure 6**. No statistically significant differences among the groups were observed ($P = 0.065$) (repeated measures ANOVA, $P < 0.05$).

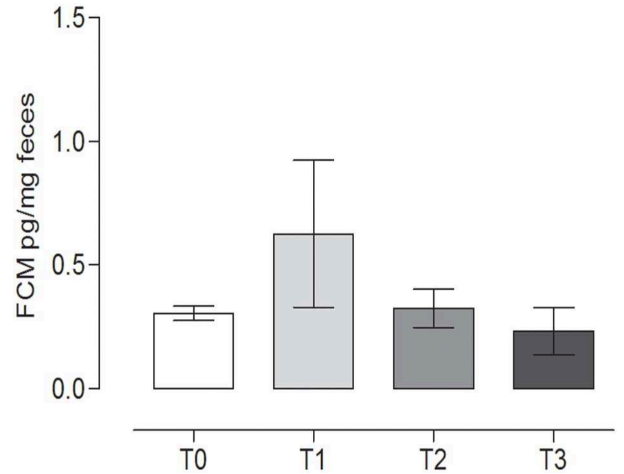


FIGURE 5 | Fecal cortisol metabolites (FCMs) at the different time points. The concentration of cortisol metabolites (mean ± SEM) in the fecal samples of the dogs at rest before training (T0), after 60 days of training (T1), after 60 days of hunting season (T2), and at 60 days of rest after hunting (T3). No statistically significant differences ($P = 0.2760$) were observed (repeated measures ANOVA, Tukey's multiple comparison test, $P < 0.05$).

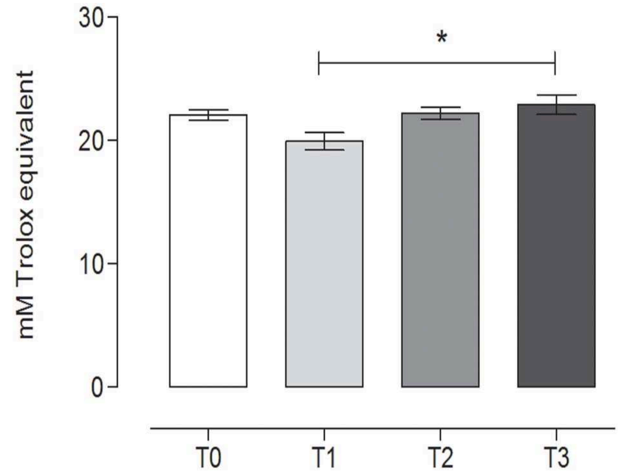


FIGURE 7 | Total antioxidant capacity (TAC) in the fecal samples at the different time points. The TAC value (mean ± SEM) in the fecal sample at rest before training (T0, 22.74 ± 0.46), after 60 days of training (T1, 19.82 ± 0.79), after 60 days of hunting season (T2, 22.16 ± 0.56), and at 60 days of rest after hunting (T3, 22.89 ± 0.89). The TAC was significantly lower at T1 than at T3 ($P = 0.0213$). *Indicates $P < 0.05$ (repeated measures ANOVA, Tukey's multiple comparison test, $P < 0.05$).

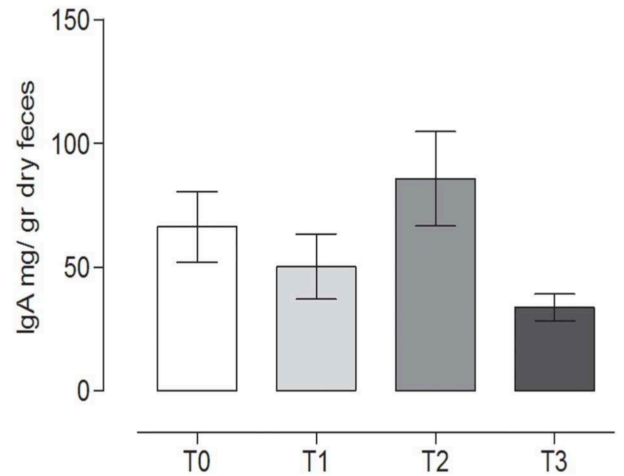


FIGURE 6 | Immunoglobulin A (IgA) concentrations in the stool at the different time points. The IgA concentrations (mean ± SEM) in the fecal samples of dogs at rest before training (T0), after 60 days of training (T1), after 60 days of hunting season (T2), and at 60 days of rest after hunting (T3). No statistically significant differences ($P = 0.065$) were observed (repeated measures ANOVA, Tukey's multiple comparison test, $P < 0.05$).

Determination of Total Antioxidant Activity

TAC showed a slight variation during the study, with a statistically significant difference between T1 (19.82 ± 0.79, mean ± SD) (after 60 days of training) and the rest after the hunting season, T3 (22.89 ± 0.89, mean ± SD) (60 days of rest after hunting) ($P = 0.0213$) (repeated measures ANOVA, $P < 0.05$, **Figure 7**).

Metabolomics of the Feces

In order to explore the changes in the fecal metabolome of the dogs involved in the study, the ^{1}H-NMR spectra were registered. Seventy-three molecules could be quantified. Seventeen molecules, reported in **Table 3**, showed a concentration that differed among the time points investigated.

To observe the overall trends driving the changes that these molecules underwent, their concentrations were used as a basis for an rPCA model, as depicted in **Figure 8**. Along PC1 of its score plot (**Figure 8A**), representing as much as 62.7% of the entire sample's variability explained by the PCA, the metabolomes of the dogs at T0 and T1 were characterized by the highest and the lowest scores, respectively, while the fecal metabolomes of the dogs at T2 and T3 appeared in intermediate positions. Specifically, the samples at T0, T1, and T2 appeared to be significantly separated from one other, while the metabolome at T3 was not distinguishable from that at T1 or T2. **Figure 8C** is a pictorial representation that highlights how all the molecules that have changed significantly over time tended to have the lowest concentrations at T0. The molecules mainly responsible for grouping the samples in this respect were proline, galacturonate and formate, 1,3-dihydroxyacetone, uridine, malate, 3-hydroxyphenylacetate, methylamine, and fucose.

The Structure and the Variations of the Gut Microbiota of Hunting Dogs as Related to Physical Activity

The 16S rRNA gene-based next-generation sequencing yielded a total of 1,390,231 high-quality reads, with an average of

TABLE 3 | Temporal dynamics of the fecal metabolome of hunting dogs following physical activity.

	T0	T1	T2	T3	*P*-value
Formate	$4.95 \times 10^{-5} \pm 1.47 \times 10^{-4\ b}$	$2.32 \times 10^{-4} \pm 5.96 \times 10^{-5\ a}$	$2.37 \times 10^{-4} \pm 4.25 \times 10^{-5\ a}$	$2.82 \times 10^{-4} \pm 1.76 \times 10^{-4\ a}$	6.72E-04
Uridine	$1.85 \times 10^{-4} \pm 8.83 \times 10^{-5\ b}$	$3.65 \times 10^{-4} \pm 1.14 \times 10^{-4\ a}$	$3.89 \times 10^{-4} \pm 1.30 \times 10^{-4\ a}$	$3.98 \times 10^{-4} \pm 1.71 \times 10^{-4\ a}$	3.89E-04
3-Hydroxyphenylacetate	$2.63 \times 10^{-4} \pm 4.25 \times 10^{-4\ b}$	$1.69 \times 10^{-3} \pm 9.38 \times 10^{-4\ a}$	$1.71 \times 10^{-3} \pm 8.45 \times 10^{-4\ a}$	$1.26 \times 10^{-3} \pm 7.45 \times 10^{-4\ ab}$	1.13E-05
Galactose	$3.02 \times 10^{-5} \pm 9.33 \times 10^{-5\ c}$	$5.16 \times 10^{-4} \pm 3.48 \times 10^{-4\ b}$	$2.78 \times 10^{-4} \pm 1.07 \times 10^{-4\ a}$	$3.39 \times 10^{-4} \pm 2.84 \times 10^{-4\ ab}$	4.16E-06
Arabinose	$8.57 \times 10^{-4} \pm 4.92 \times 10^{-4\ b}$	$2.99 \times 10^{-3} \pm 1.16 \times 10^{-3\ a}$	$2.50 \times 10^{-3} \pm 8.47 \times 10^{-4\ a}$	$2.05 \times 10^{-3} \pm 1.11 \times 10^{-3\ a}$	1.13E-05
Fucose	$4.52 \times 10^{-5} \pm 2.02 \times 10^{-4\ b}$	$5.88 \times 10^{-4} \pm 2.30 \times 10^{-4\ a}$	$4.72 \times 10^{-4} \pm 3.19 \times 10^{-4\ a}$	$4.36 \times 10^{-4} \pm 1.89 \times 10^{-4\ a}$	6.88E-05
1,3-Dihydroxyacetone	$1.23 \times 10^{-5} \pm 2.44 \times 10^{-5\ b}$	$1.39 \times 10^{-4} \pm 1.24 \times 10^{-4\ a}$	$9.51 \times 10^{-5} \pm 7.63 \times 10^{-5\ a}$	$9.42 \times 10^{-5} \pm 8.98 \times 10^{-5\ a}$	3.35E-04
Galacturonate	$6.51 \times 10^{-5} \pm 9.02 \times 10^{-5\ c}$	$1.94 \times 10^{-4} \pm 8.84 \times 10^{-5\ b}$	$1.08 \times 10^{-4} \pm 6.96 \times 10^{-5\ abc}$	$1.35 \times 10^{-4} \pm 5.67 \times 10^{-5\ a}$	1.30E-05
Malate	$7.92 \times 10^{-4} \pm 6.07 \times 10^{-4\ b}$	$1.83 \times 10^{-3} \pm 9.29 \times 10^{-4\ a}$	$1.46 \times 10^{-3} \pm 1.01 \times 10^{-3\ ab}$	$2.32 \times 10^{-3} \pm 1.93 \times 10^{-3\ a}$	3.48E-02
Threonine	$8.07 \times 10^{-4} \pm 5.81 \times 10^{-4\ b}$	$2.14 \times 10^{-3} \pm 6.66 \times 10^{-4\ a}$	$1.91 \times 10^{-3} \pm 3.42 \times 10^{-4\ a}$	$1.99 \times 10^{-3} \pm 5.16 \times 10^{-4\ a}$	1.17E-03
Glycine	$2.07 \times 10^{-3} \pm 4.68 \times 10^{-4\ b}$	$4.82 \times 10^{-3} \pm 4.85 \times 10^{-3\ ab}$	$2.80 \times 10^{-3} \pm 7.04 \times 10^{-4\ ab}$	$3.33 \times 10^{-3} \pm 9.62 \times 10^{-4\ a}$	1.70E-03
Methanol	$2.37 \times 10^{-4} \pm 2.06 \times 10^{-4\ b}$	$6.02 \times 10^{-4} \pm 2.77 \times 10^{-4\ a}$	$4.48 \times 10^{-4} \pm 1.29 \times 10^{-4\ ab}$	$4.98 \times 10^{-4} \pm 2.82 \times 10^{-4\ ab}$	2.02E-02
Proline	$2.39 \times 10^{-4} \pm 1.31 \times 10^{-4\ b}$	$6.64 \times 10^{-4} \pm 1.49 \times 10^{-4\ a}$	$6.58 \times 10^{-4} \pm 3.27 \times 10^{-4\ ab}$	$6.57 \times 10^{-4} \pm 1.36 \times 10^{-4\ a}$	2.10E-05
Trimethylamine (TMA)	$3.99 \times 10^{-4} \pm 2.36 \times 10^{-4\ ab}$	$2.65 \times 10^{-4} \pm 1.44 \times 10^{-4\ b}$	$4.04 \times 10^{-4} \pm 1.27 \times 10^{-4\ a}$	$4.44 \times 10^{-4} \pm 3.00 \times 10^{-4\ ab}$	4.53E-02
Homocystine	$3.17 \times 10^{-4} \pm 5.05 \times 10^{-4\ b}$	$2.42 \times 10^{-3} \pm 1.56 \times 10^{-3\ a}$	$2.17 \times 10^{-3} \pm 1.41 \times 10^{-3\ a}$	$1.52 \times 10^{-3} \pm 8.90 \times 10^{-4\ a}$	6.01E-05
Methylamine	$1.84 \times 10^{-4} \pm 8.64 \times 10^{-5\ b}$	$4.20 \times 10^{-4} \pm 2.34 \times 10^{-4\ ab}$	$3.81 \times 10^{-4} \pm 1.08 \times 10^{-4\ a}$	$3.02 \times 10^{-4} \pm 1.21 \times 10^{-4\ b}$	1.13E-03
Valine	$1.80 \times 10^{-3} \pm 7.53 \times 10^{-4\ b}$	$2.51 \times 10^{-3} \pm 8.29 \times 10^{-4\ ab}$	$2.66 \times 10^{-3} \pm 5.34 \times 10^{-4\ ab}$	$2.85 \times 10^{-3} \pm 9.57 \times 10^{-4\ a}$	4.42E-02

Concentration (mmol/g, mean ± SD) of the molecules significantly differed among groups (Repeated Measure ANOVA, $P < 0.05$).
**For each molecule, different superscript letters identify significant differences among the groups ($P < 0.05$). For each molecule P value was reported.*

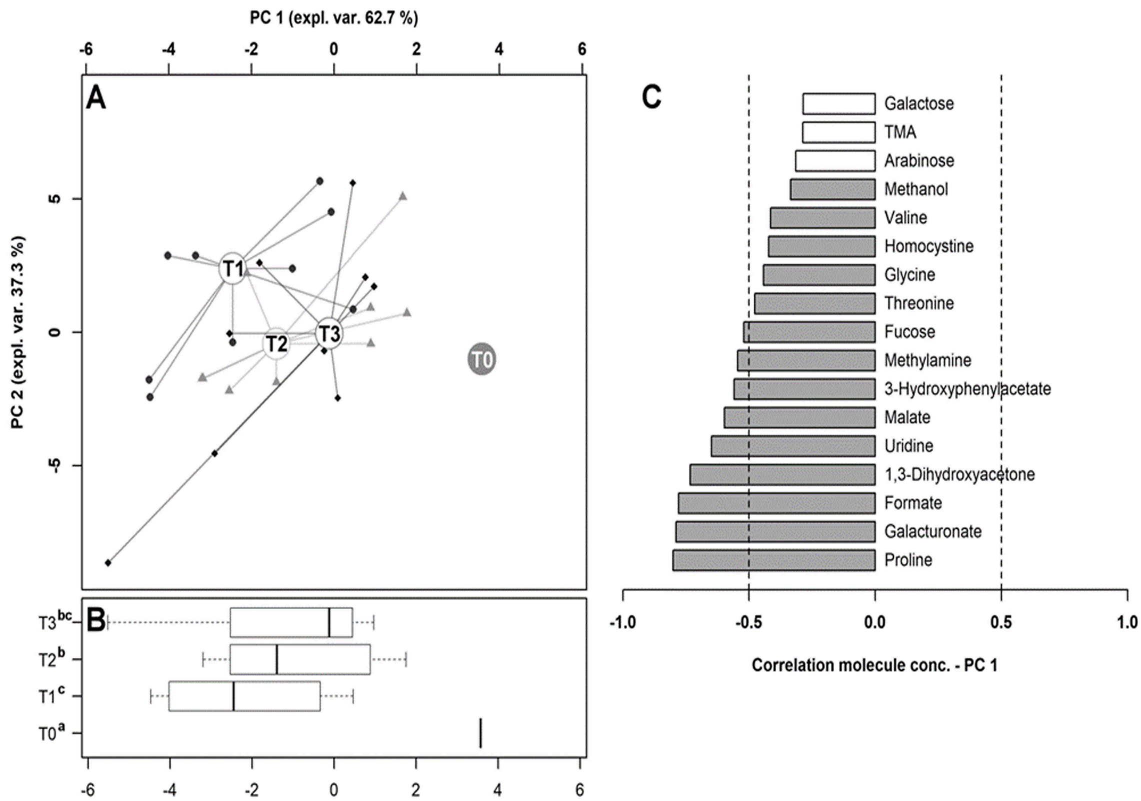


FIGURE 8 | Diversity of the fecal metabolome in hunting dogs. The robust principal component analysis (rPCA) model built on the space constituted by the molecules listed in **Table 3**, the concentration of which at T0 was subtracted from the other time points. In the score plot **(A)**, the empty circles highlight the medians of the samples collected at each time point. The position of the samples along PC1 is summarized in the box plots in **(B)**, where the different superscript letters identify the significant differences among the groups (repeated measures ANOVA, $P < 0.05$). The loading plot **(C)** reports the correlation between the concentration of each substance and its importance over PC1. The significant correlations (repeated measures ANOVA, $P < 0.05$) are highlighted with gray bars.

39,720 ± 12,005 sequences per sample, binned in 1,460 ASVs at 99% similarity.

The PCoA of inter-sample variation based on weighted and unweighted UniFrac distances showed significant separation among the study groups ($P < 0.03$, permutation test with pseudo-F ratios; $P \leq 0.02$, ANOSIM) (**Figure 9A**). In particular, according to both the adonis and the ANOSIM statistics applied to the unweighted UniFrac-based ordination, the samples at T1 and T2 segregated from those at T0 ($P < 0.005$), while the T3 samples occupied an intermediate position (**Table S1**). No significant differences were found in alpha diversity, even though Faith's phylogenetic index showed an increasing trend over time (**Figure 9B**).

In line with the literature available regarding the gut microbiota of healthy dogs (77, 78), the fecal microbial profiles at the baseline were dominated by the phylum Firmicutes (relative abundance, mean ± SEM, 69.6 ± 8.1%), with Bacteroidetes (12.0 ± 5.2%), Actinobacteria (6.7 ± 3.1%), Proteobacteria (6.0 ± 2.4%), and Fusobacteria (5.5 ± 4.1%) as minor components. Similar proportions were observed during training, hunting, and the subsequent rest period, except for a reduction in the relative abundance of Proteobacteria after training ($P < 0.01$, Wilcoxon test). *Clostridiaceae*, *Erysipelotrichaceae*, and *Lactobacillaceae* were the major families of the baseline microbiota (relative abundance > 10%). Following training, an increase in the relative abundance of *Streptococcaceae* and *Enterococcaceae* was observed ($P < 0.05$). Such an increase persisted for *Streptococcaceae* ($P = 0.008$) until the rest period after hunting, while for *Enterococcaceae,* relative abundance values comparable to the baseline were restored (**Figure 10**). In contrast, diminished proportions were observed for *Prevotellaceae* and *Ruminococcaceae* after training ($P < 0.04$).

Consistent with the above results, the main discriminant genera were *Streptococcus* and *Enterococcus*, the relative abundance of which was significantly greater at T1 than at T0 ($P < 0.04$), and *Prevotella*, the proportions of which decreased after training ($P \leq 0.03$) (**Figure 10**).

Although not significant, a decreasing trend was observed for *Faecalibacterium* and *Bacteroides* after physical activity (i.e., training and hunting) compared to both rest periods (i.e., before

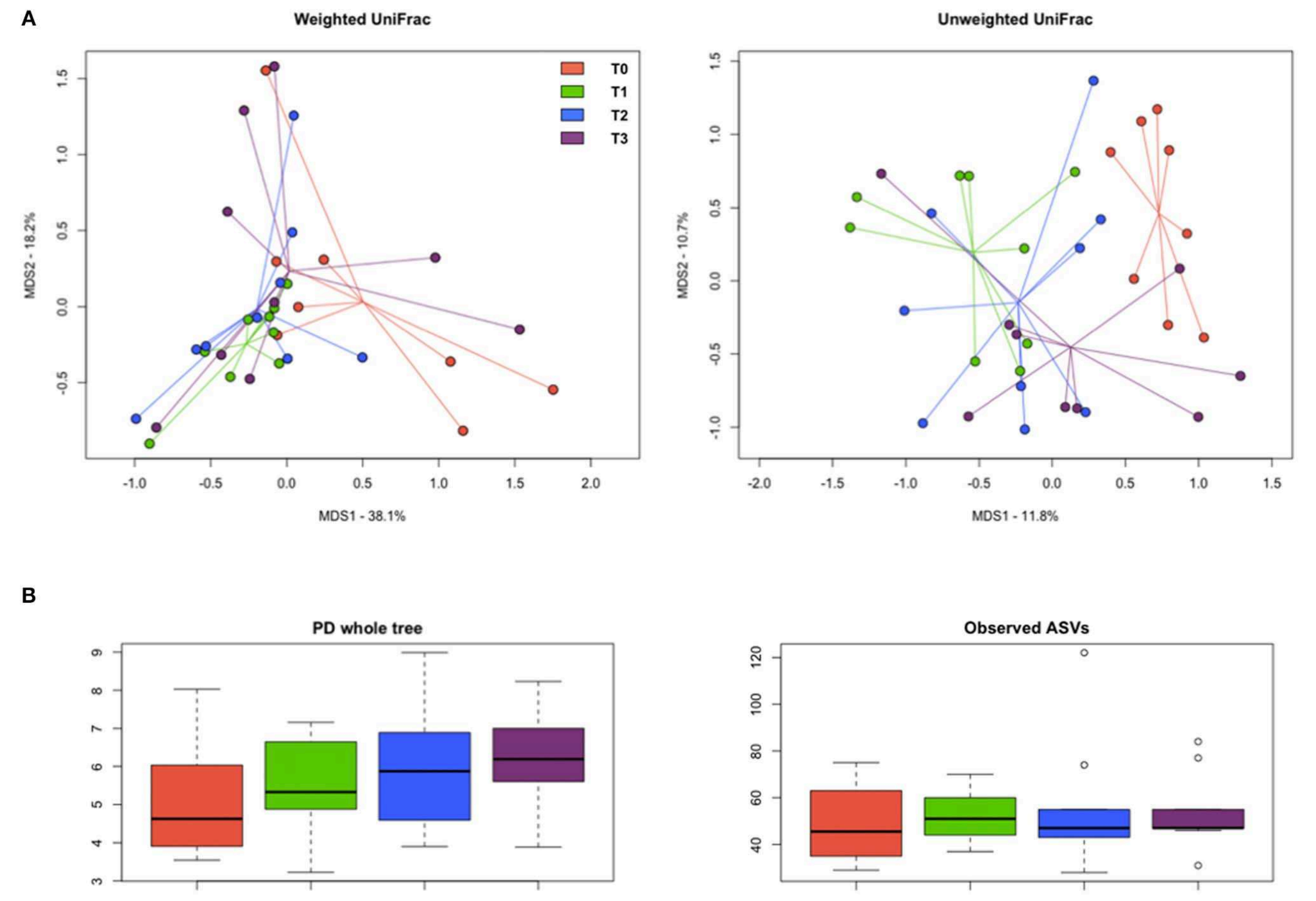


FIGURE 9 | Diversity of the gut microbiome in hunting dogs. **(A)** Principal coordinate analysis (PCoA) plots showing the beta diversity of the gut microbial communities of the study groups (rest before training, T0; after 60 days of training, T1; after 60 days of hunting season, T2; at 60 days of rest after hunting, T3), based on weighted and unweighted UniFrac distances. A significant separation among groups was found ($P < 0.03$, permutation test with pseudo-F ratios; $P \leq 0.02$, ANOSIM). **(B)** Box plots showing alpha diversity, computed with Faith's phylogenetic index (PD whole tree) and the number of observed amplicon sequence variants (ASVs).

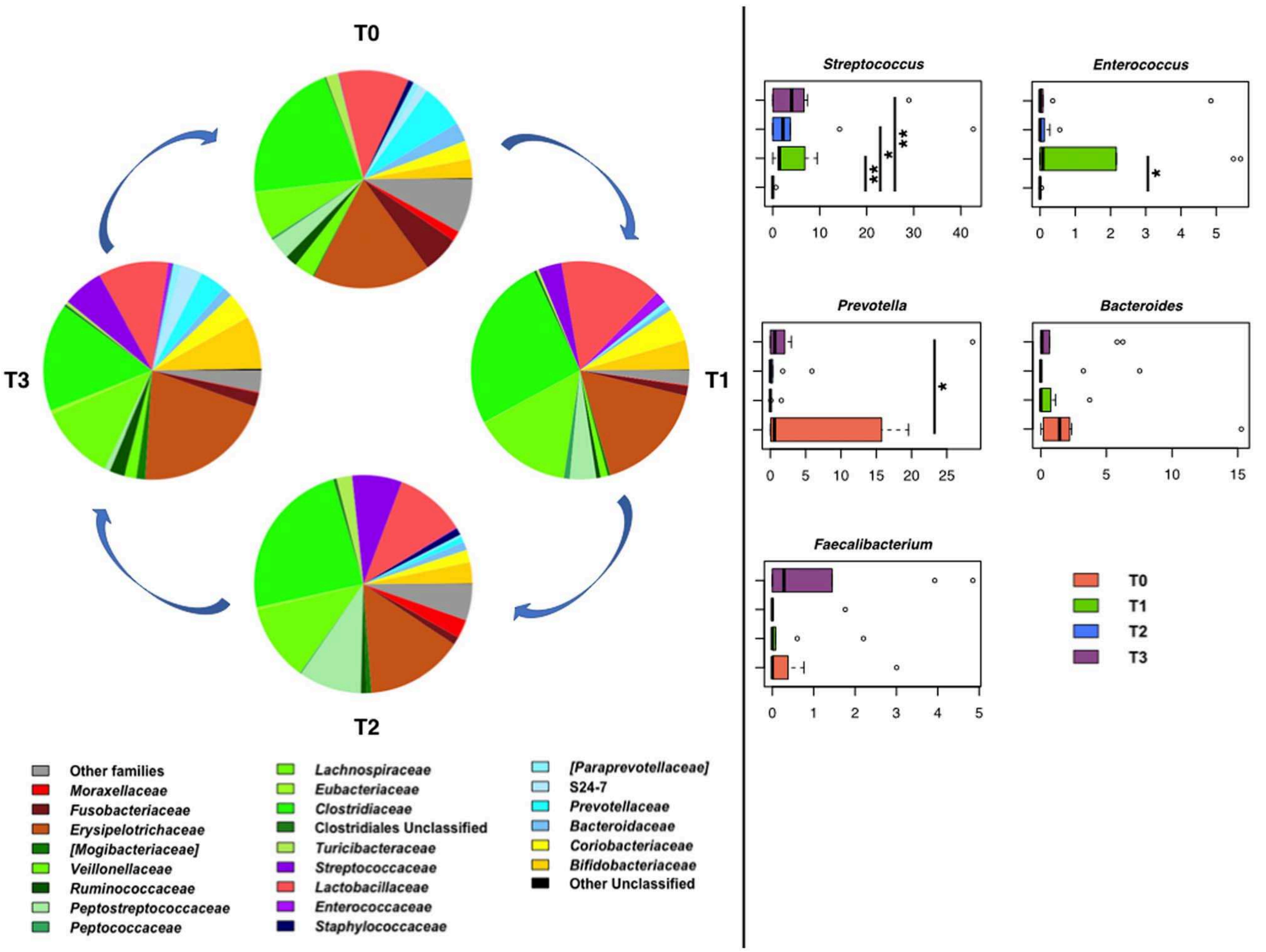


FIGURE 10 | Temporal dynamics of the gut microbiome in hunting dogs following physical activity. **Left**, pie charts representing the average values of family-level relative abundances at each time point (rest before training, T0; after 60 days of training, T1; after 60 days of hunting season, T2; at 60 days of rest after hunting, T3). **Right**, box plots showing the distribution of the relative abundances of significantly enriched or depleted bacterial genera over time. $^*P < 0.04$; $^{**}P < 0.001$ (Wilcoxon test). For *Bacteroides* and *Faecalibacterium*, only a decreasing trend was observed.

the training and after the hunting season). At rest, the baseline relative abundance of *Enterococcus* was restored, whereas the proportions of *Streptococcus* remained higher than the baseline ($P = 0.008$).

DISCUSSION

Inx this study, we evaluated the fluctuations of different stress markers in fecal samples of hunting dogs during physical activity and at rest. The main limitation for such studies lies within the difficulty in standardizing the training protocol (wild boar hunting) and the management of privately owned animals (diet, housing, treatments, etc.). In order to try and overcome this problem, we chose a group of dogs owned by the same person, in this case, one of the animal technicians of the Veterinary Department. He is indeed routinely involved in the husbandry and care of animals for both clinical and experimental purposes, and he was fully aware of the goals of the experiment and of the potential biases imputable to variations in the management of animals enrolled in such trials.

This choice has added a limiting factor to the study, which, being exploratory, included a low number of dogs of different breeds and ages, variables known to potentially influence the results (79); nonetheless, the study design allowed for a high level of standardization in terms of dogs' management, making for reliable results despite the relatively low sample size.

Typically, performance dogs are kept at 4–5/9 BCS due to the great chance of body condition loss during endurance activity, and the diet was calculated to support this condition. Weight loss is known to be related with some of the parameters measured in this research, for instance, microbiota (80), metabolomics profile (81), and cortisol (82). The expected weight loss observed during the trial has to be interpreted as a direct consequence of physical activity and not of a caloric restriction, so its potential effects on the measured parameters could be considered as a direct consequence of the physical activity.

In the model used in the present study, fecal cortisol, a well-known marker of stress in dogs, did not show any difference across the time points. In previous investigations, increases in cortisol concentration after sustained exercise had been observed in horses and humans, while the data regarding dogs were contradictory. In fact, some papers have reported increased levels of cortisol (83, 84), while others agreed with the present study in reporting no significant changes (85, 86). In particular, Pastore et al. (37) and Ando et al. (39) reported that cortisol increased right after exercise but returned to baseline levels shortly after, suggesting a mild transient stress. Similarly, in the present study, cortisol showed a transient non-significant increase during the first phase of activity (60 days of training, T1) only. Moreover, all samples were collected during a short-day period (autumn–winter), avoiding the reported interference of photoperiod on the cortisol concentrations (87).

Intestinal IgA secretion is considered to be an important indicator of mucosal immunity. Similar to cortisol, the literature regarding the effect of exercise on IgA secretion is contradictory, reporting either an increase or a decrease in intestinal IgA in mice (42, 88, 89). Based on the present data, training might not influence IgA concentration, confirming that exercise does not drastically alter canine intestinal immune homeostasis.

A previous paper indicated an increase in oxidative stress in hunting dogs after exercise (43): in accordance with this paper, the data in the present study also showed a significant and transient reduction in TAC during T1 (60 days of training), in relation to T3 (60 days of rest after hunting), suggesting an increase in oxidative stress following the resumption of physical activity.

In agreement with the human results, in the present study, we were able to detect different biomarkers' transcripts in dog stool samples. Among the studied genes, PAR-2 and HO-1 were significantly altered after the hunting period. To date, the relationship between exercise and increased HO-1 expression has been well-documented in different tissue and animal models (31–34). Such an increase is likely to restore HO-1 protein expression levels after 60 days of training (T1), when oxidative stress is high, as confirmed by the TAC data. As for PAR-2, it is well-described in intestinal models of I/R injury that the receptor is strongly activated by the tryptase released, for the most part by the mast cell infiltrate, with a consequent increase in paracellular permeability by means of the activation of myosin light chain kinase (MLCK) and myosin phosphatase (MP) (17, 90); once activated, the receptor is translocated to the lysosomes and degraded (23, 91). In different animal models regarding intestinal I/R injury, an increase in the PAR-2 transcript has been observed (21–23), consistent with the present data showing a slight but significant increase in PAR-2 mRNA levels at the end of the hunting period (T2). This similar trend in different models may be due to the fact that during exercise, the blood flow is diverted from the gut to the periphery, creating an I/R-like scenario (92) with the potential consequent activation of PAR-2. It has been reported that PAR-2 activation may directly affect cytoskeleton contraction by triggering the phosphorylation of MLCK with subsequent changes in TJ permeability, as demonstrated in *in vitro* epithelial models (19, 20). However, the unchanged expression level of OCLN suggests that the PAR-2 receptor activation in our model is insufficient to induce damage at the TJ level, and so we were unable to predict the impairment of barrier permeability.

The lack of the detection of cytokine transcripts and the absence of changes in CALP mRNA levels additionally reinforced the authors' assumptions, i.e., that physical exercise in the present model could be considered mild and did not result in a strong inflammatory GI response.

Nevertheless, metabolomics data indicate that some intestinal disorder occurred. A two-step approach regarding the metabolome of the feces, based on univariate/multivariate analyses, allowed hypothesizing the overall trends that the fecal molecule profiles underwent as a consequence of resting, training, and hunting. The samples collected at T2, T3, and T0 showed median scores along PC1 of −1.39, −0.12, and 3.69, respectively. From a metabolomic perspective, therefore, the recovery of baseline conditions seemed to be linearly related to time. The metabolomes of the dogs at rest before the training (T0) were markedly different from all the other time points. The greatest modifications from this long period of rest were associated with training, while the subsequent activities seemed to lead to a progressive return of the metabolome to the baseline characteristics. This confirmed a metabolic shift between rest and activity. Of the molecules leading to such a circular trend, some, as expected, pertained to the biochemical processes connected to energy (46). This was the case for malate, which is part of the TCA cycle. Interestingly, of the sugars, glucose showed no significant differences, while fucose and galacturonate did. Of the molecules that were, for the most part, modified in the present study, 1,3-dihydroxyacetone, formate, and uridine should be mentioned. In a previous experiment (93), these three molecules were found to be altered in mouse feces after the administration of probiotics, probably as a result of the modification of the intestinal microbiota. In particular, the increase in 1,3-dihydroxyacetone, an intermediate in fructose metabolism, was found to lead to an increase in intestinal permeability, which is a known consequence of prolonged strenuous exercise in both dogs (94) and humans (7).

Consistent with the abovementioned assumptions, the gut microbiota structure also underwent a rearrangement during training and tended to approach the initial configuration in the rest period following the hunt. In line with the literature available regarding exercise and gut microbiota, this rearrangement was characterized by: (1) a tendency toward increased biodiversity (95); (2) decreased relative abundance of widely prevalent commensals (i.e., *Prevotella* and *Ruminococcaceae* members) (96–98); and (3) increased proportions of subdominant taxa, including *Streptococcus*, *Enterococcus*, and *Slackia* (96, 99). The majority of these changes were transient, which additionally reinforced the hypothesis of a reversible non-drastic alteration of the intestinal ecosystem. However, this was not true for *Streptococcus*, which, similarly to *Enterococcus*, includes species known to act as pathobionts, i.e., capable of pathogenic expansion under unfavorable conditions, compromising and eventually translocating across the epithelial barrier, with potentially severe implications for the host health (100). It is also worth noting that *Streptococcus* spp. are capable of proteolytically interacting with PARs (101) and have previously been positively correlated with uridine levels, probably by means of the activity of cytidine deaminase (102), which suggests a major role for this bacterial genus in exercise response. On the other hand, negative correlations have so far been found between uridine as well as DHA and *Bacteroides* (103), the relative abundance of which tended to be gradually reduced over the course of activity and no longer restored. Although transient and non-significant, the depletion of *Faecalibacterium*, a well-known butyrate producer with multiple health-promoting activities (104), constitutes another red flag for possible GI (and systemic) complications and should be monitored in cases of intense and prolonged physical activity.

CONCLUSION

The aim of the present explorative study was to evaluate the presence of a GI derangement in hunting dogs through a non-invasive sampling as a consequence of a period of intense exercise in comparison with samples collected at rest.

We evaluated a number of potential stress markers in canine fecal samples. In particular, FCMs, IgA levels, and the TAC were measured. Moreover, the expression of selected genes was investigated, and microbiota and metabolomics analyses were carried out. Exercise induced a variation in gene expression, a reduction in TAC, and a modulation of the microbiome and metabolome profiles. Despite the intense physical activity required for hunting wild boar, the animals did not seem to show signs of particularly high stress under conditions of programmed training; all the data were consistent with a limited degree of alteration of intestinal homeostasis. Despite the limited statistical power of the study related to the relatively low number of subjects enrolled, the present findings are encouraging for the development of a non-invasive monitoring method for detecting the effect of exercise in dogs using a multidisciplinary integrated approach.

DATA AVAILABILITY STATEMENT

The microbiota dataset generated for this study can be found in the NCBI SRA (BioProject ID PRJNA 589580).

AUTHOR CONTRIBUTIONS

MF and PB conceived the project. MP and AG participated in the inclusion and clinical examination of the dogs, and performed the diagnostic endoscopy. AZ carried out the gene expression and IgA and TAC analyses. LL and CZ were responsible for the metabolomics investigation. MB and ST were responsible for the gut microbiota analysis. PA carried out the FCM analysis. AZ, LL, ST, and MB carried out the statistical analyses and wrote the original draft. MF and PB reviewed and edited the draft. All authors read and approved the final version of the manuscript.

FUNDING

This study was funded with grants provided by the University of Bologna (Programma di Ricerca Fondamentale Orientata, RFO 2018).

ACKNOWLEDGMENTS

The authors are very grateful to Daniele Vanti, the owner of the dogs included in the study. CZ gratefully acknowledges financial support from the Chinese Scholarship Council (grant no. 201606910076).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.00126/full#supplementary-material

REFERENCES

1. Powers SK, Jackson MJ. Exercise-induced oxidative stress: cellular mechanisms and impact on muscle force production. *Physiol Rev*. (2008) 88:1243–76. doi: 10.1152/physrev.00031.2007
2. Criswell D, Powers S, Dodd S, Lawler J, Edwards W, Renshler K, et al. High intensity training-induced changes in skeletal muscle antioxidant enzyme activity. *Med Sci Sports Exerc*. (1993) 25:1135–40. doi: 10.1249/00005768-199310000-00009
3. Finaud J, Lac G, Filaire E. Oxidative stress. *Sport Med*. (2006) 36:327–58. doi: 10.2165/00007256-200636040-00004
4. Lambert GP. Stress-induced gastrointestinal barrier dysfunction and its inflammatory effects. *J Anim Sci*. (2009) 87:E101–8. doi: 10.2527/jas.2008-1339
5. JanssenDuijghuijsen LM, Mensink M, Lenaerts K, Fiedorowicz E, Protégé study group, van Dartel DAM, et al. The effect of endurance exercise on intestinal integrity in well-trained healthy men. *Physiol Rep*. (2016) 4:e12994. doi: 10.14814/phy2.12994
6. de Oliveira EP, Burini RC, Jeukendrup A. Gastrointestinal complaints during exercise: prevalence, etiology, and nutritional recommendations. *Sports Med*. (2014) 44:S79–85. doi: 10.1007/s40279-014-0153-2
7. Lamprecht M, Frauwallner A. Exercise, intestinal barrier dysfunction and probiotic supplementation. *Med Sport Sci*. (2012) 59:47–56. doi: 10.1159/000342169
8. Choi SC, Choi SJ, Kim JA, Kim TH, Nah YH, Yazaki E, et al. The role of gastrointestinal endoscopy in long-distance runners with gastrointestinal symptoms. *Eur J Gastroenterol Hepatol*. (2001) 13:1089–94. doi: 10.1097/00042737-200109000-00016
9. Murray MJ, Schusser GF, Pipers FS, Gross SJ. Factors associated with gastric lesions in thoroughbred racehorses. *Equine Vet J*. (1996) 28:368–74. doi: 10.1111/j.2042-3306.1996.tb03107.x
10. Vatistas NJ, Snyder JR, Carlson G, Johnson B, Arthur RM, Thurmond M, et al. Cross-sectional study of gastric ulcers of the squamous mucosa in thoroughbred racehorses. *Equine Vet J Suppl*. (1999) 31:34–9. doi: 10.1111/j.2042-3306.1999.tb05166.x

11. Davis MS, Willard MD, Nelson SL, Mandsager RE, McKiernan BS, Mansell JK, et al. Prevalence of gastric lesions in racing alaskan sled dogs. *J Vet Intern Med*. (2003) 17:311–4. doi: 10.1111/j.1939-1676.2003.tb02453.x
12. ter Steege RWF, Kolkman JJ. Review article: the pathophysiology and management of gastrointestinal symptoms during physical exercise, and the role of splanchnic blood flow. *Aliment Pharmacol Ther*. (2012) 35:516–28. doi: 10.1111/j.1365-2036.2011.04980.x
13. van Wijck K, Lenaerts K, van Loon LJC, Peters WHM, Buurman WA, Dejong CHC. Exercise-induced splanchnic hypoperfusion results in gut dysfunction in healthy men. *PLoS ONE*. (2011) 6:e22366. doi: 10.1371/journal.pone.0022366
14. Zuhl M, Schneider S, Lanphere K, Conn C, Dokladny K, Moseley P. Exercise regulation of intestinal tight junction proteins. *Br J Sports Med*. (2014) 48:980–6. doi: 10.1136/bjsports-2012-091585
15. Poritz LS, Garver KI, Green C, Fitzpatrick L, Ruggiero F, Koltun WA. Loss of the tight junction protein ZO-1 in dextran sulfate sodium induced colitis. *J Surg Res*. (2007) 140:12–9. doi: 10.1016/j.jss.2006.07.050
16. Zhang Z, Li S, Cao H, Shen P, Liu J, Fu Y, et al. The protective role of phloretin against dextran sulfate sodium-induced ulcerative colitis in mice. *Food Funct*. (2019) 10:422–31. doi: 10.1039/C8FO01699B
17. Bueno L, Fioramonti J. Protease-activated receptor 2 and gut permeability: a review. *Neurogastroenterol Motil*. (2008) 20:580–7. doi: 10.1111/j.1365-2982.2008.01139.x
18. Jacob C, Yang P-C, Darmoul D, Amadesi S, Saito T, Cottrell GS, et al. Mast cell tryptase controls paracellular permeability of the intestine. Role of protease-activated receptor 2 and beta-arrestins. *J Biol Chem*. (2005) 280:31936–48. doi: 10.1074/jbc.M506338200
19. Enjoji S, Ohama T, Sato K. Regulation of epithelial cell tight junctions by protease-activated receptor 2. *J Vet Med Sci*. (2014) 76:1225–9. doi: 10.1292/jvms.14-0191
20. Nadeau P, Henehan M, de Benedetto A. Activation of protease-activated receptor 2 leads to impairment of keratinocyte tight junction integrity. *J Allergy Clin Immunol*. (2018) 142:281–4.e7. doi: 10.1016/j.jaci.2018.01.007
21. Cenac N, Coelho A-M, Nguyen C, Compton S, Gordon AP, MacNaughton WK, et al. Induction of intestinal inflammation in mouse by activation of proteinase-Activated receptor-2. *Am J Pathol*. (2002) 161:1903–15. doi: 10.1016/S0002-9440(10)64466-5
22. Liu D, Gan X, Huang P, Chen X, Ge M, Hei Z. Inhibiting tryptase after ischemia limits small intestinal ischemia-reperfusion injury through protease-activated receptor 2 in rats. *J Trauma Acute Care Surg*. (2012) 73:1138–44. doi: 10.1097/TA.0b013e318265d08d
23. Romagnoli N, Zannoni A, Bernardini C, Gobbetti T, Bombardi C, Rambaldi AM, et al. Proteinase-activated receptor 2 distribution and expression in equine small intestine tracts following herniation through the epiploic foramen. *Res Vet Sci*. (2017) 125:434–40. doi: 10.1016/j.rvsc.2017.10.006
24. Heilmann RM, Nestler J, Schwarz J, Grützner N, Ambrus A, Seeger J, et al. Mucosal expression of S100A12 (calgranulin C) and S100A8/A9 (calprotectin) and correlation with serum and fecal concentrations in dogs with chronic inflammatory enteropathy. *Vet Immunol Immunopathol*. (2019) 211:64–74. doi: 10.1016/j.vetimm.2019.04.003
25. Amara J, Saliba Y, Hajal J, Smayra V, Bakhos J-J, Sayegh R, et al. Circadian rhythm disruption aggravates DSS-Induced colitis in mice with fecal calprotectin as a marker of colitis severity. *Dig Dis Sci*. (2019) 64:3122–33. doi: 10.1007/s10620-019-05675-7
26. Perše M, Cerar A. Dextran sodium sulphate colitis mouse model: traps and tricks. *J Biomed Biotechnol*. (2012) 2012:718617. doi: 10.1155/2012/718617
27. Takagi T, Naito Y, Mizushima K, Nukigi Y, Okada H, Suzuki T, et al. Increased intestinal expression of heme oxygenase-1 and its localization in patients with ulcerative colitis. *J Gastroenterol Hepatol*. (2008) 23:S229–33. doi: 10.1111/j.1440-1746.2008.05443.x
28. Waza AA, Hamid Z, Ali S, Bhat SA, Bhat MA. A review on heme oxygenase-1 induction: is it a necessary evil. *Inflamm Res*. (2018) 67:579–88. doi: 10.1007/s00011-018-1151-x
29. Berberat PO, A-Rahim YI, Yamashita K, Warny MM, Csizmadia E, Robson SC, et al. Heme oxygenase-1-generated biliverdin ameliorates experimental murine colitis. *Inflamm Bowel Dis*. (2005) 11:350–9. doi: 10.1097/01.MIB.0000164017.06538.8a
30. Wang WP, Guo X, Koo MWL, Wong BCY, Lam SK, Ye YN, et al. Protective role of heme oxygenase-1 on trinitrobenzene sulfonic acid-induced colitis in rats. *Am J Physiol Liver Physiol*. (2001) 281:G586–94. doi: 10.1152/ajpgi.2001.281.2.G586
31. George L, Lokhandwala MF, Asghar M. Exercise activates redox-sensitive transcription factors and restores renal D_1 receptor function in old rats. *Am J Physiol Physiol*. (2009) 297:F1174–80. doi: 10.1152/ajprenal.00397.2009
32. Miyagi MYS, Seelaender M, Castoldi A, de Almeida DC, Bacurau AVN, Andrade-Oliveira V, et al. Long-term aerobic exercise protects against cisplatin-Induced nephrotoxicity by modulating the expression of IL-6 and HO-1. *PLoS ONE*. (2014) 9:e108543. doi: 10.1371/journal.pone.0108543
33. Ren C, Qi J, Li W, Zhang J. The effect of moderate-intensity exercise on the expression of HO-1 mRNA and activity of HO in cardiac and vascular smooth muscle of spontaneously hypertensive rats. *Can J Physiol Pharmacol*. (2016) 94:448–54. doi: 10.1139/cjpp-2015-0122
34. Bostick B, Aroor AR, Habibi J, Durante W, Ma L, deMarco VG, et al. Daily exercise prevents diastolic dysfunction and oxidative stress in a female mouse model of western diet induced obesity by maintaining cardiac heme oxygenase-1 levels. *Metabolism*. (2017) 66:14–22. doi: 10.1016/j.metabol.2016.09.005
35. Ghio AJ, Case MW, Soukup JM. Heme oxygenase activity increases after exercise in healthy volunteers. *Free Radic Res*. (2018) 52:267–72. doi: 10.1080/10715762.2018.1428965
36. Atamaniuk J, Stuhlmeier KM, Vidotto C, Tschan H, Dossenbach-Glaninger A, Mueller MM. Effects of ultra-marathon on circulating DNA and mRNA expression of pro- and anti-apoptotic genes in mononuclear cells. *Eur J Appl Physiol*. (2008) 104:711–7. doi: 10.1007/s00421-008-0827-2
37. Pastore C, Pirrone F, Balzarotti F, Faustini M, Pierantoni L, Albertini M. Evaluation of physiological and behavioral stress-dependent parameters in agility dogs. *J Vet Behav Clin Appl Res*. (2011) 6:188–94. doi: 10.1016/j.jveb.2011.01.001
38. Angle CT, Wakshlag JJ, Gillette RL, Stokol T, Geske S, Adkins TO, et al. Hematologic, serum biochemical, and cortisol changes associated with anticipation of exercise and short duration high-intensity exercise in sled dogs. *Vet Clin Pathol*. (2009) 38:370–4. doi: 10.1111/j.1939-165X.2009.00122.x
39. Ando I, Karasawa K, Matsuda H, Tanaka A. Changes in serum NGF levels after the exercise load in dogs: a pilot study. *J Vet Med Sci*. (2016) 78:1709–12. doi: 10.1292/jvms.16-0258
40. Palm NW, de Zoete MR, Cullen TW, Barry NA, Stefanowski J, Hao L, et al. Immunoglobulin a coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell*. (2014) 158:1000–10. doi: 10.1016/j.cell.2014.08.006
41. Circu ML, Aw TY. Intestinal redox biology and oxidative stress. *Semin Cell Dev Biol*. (2012) 23:729–37. doi: 10.1016/j.semcdb.2012.03.014
42. Serrano DME, Victoria GM, Padilla LE, Albor RA, Garfias RH, Martínez AI, et al. Moderate exercise enhances expression of SIgA in mouse ileum. *Int J Sports Med*. (2012) 33:1020–5. doi: 10.1055/s-0032-1312607
43. Pasquini A, Luchetti E, Cardini G. Evaluation of oxidative stress in hunting dogs during exercise. *Res Vet Sci*. (2010) 89:120–3. doi: 10.1016/j.rvsc.2010.01.004
44. Mailing LJ, Allen JM, Buford TW, Fields CJ, Woods JA. Exercise and the gut microbiome: a review of the evidence, potential mechanisms, and implications for human health. *Exerc Sport Sci Rev*. (2019) 47:75–85. doi: 10.1249/JES.0000000000000183
45. Turroni S, Brigidi P, Cavalli A, Candela M. Microbiota-Host transgenomic metabolism, bioactive molecules from the inside. *J Med Chem*. (2018) 61:47–61. doi: 10.1021/acs.jmedchem.7b00244
46. Deda O, Gika H, Panagoulis T, Taitzoglou I, Raikos N, Theodoridis G. Impact of exercise on fecal and cecal metabolome over aging: a longitudinal study in rats. *Bioanalysis*. (2017) 9:21–36. doi: 10.4155/bio-2016-0222
47. Cho Y, Kim H, Turner ND, Mann JC, Wei J, Taddeo SS, et al. A chemoprotective fish oil- and pectin-containing diet temporally alters gene expression profiles in exfoliated rat colonocytes throughout oncogenesis. *J Nutr*. (2011) 141:1029–35. doi: 10.3945/jn.110.134973
48. Kaeffer B, des Robert C, Gouabau AMC, Pagniez A, Legrand A, Amarger V, et al. Recovery of exfoliated cells from the gastrointestinal tract of premature infants: a new tool to perform "Noninvasive Biopsies?" *Pediatr Res*. (2007) 62:564–9. doi: 10.1203/PDR.0b013e318155a402

49. Chapkin RS, Zhao C, Ivanov I, Davidson LA, Goldsby JS, Lupton JR, et al. Noninvasive stool-based detection of infant gastrointestinal development using gene expression profiles from exfoliated epithelial cells. *Am J Physiol Gastrointest Liver Physiol*. (2010) 298:G582–9. doi: 10.1152/ajpgi.00004.2010
50. Donovan SM, Wang M, Monaco MH, Martin CR, Davidson LA, Ivanov I, et al. Noninvasive molecular fingerprinting of host-microbiome interactions in neonates. *FEBS Lett*. (2014) 588:4112–9. doi: 10.1016/j.febslet.2014.07.008
51. Davidson LA, Aymond CM, Jiang YH, Turner ND, Lupton JR, Chapkin RS. Non-invasive detection of fecal protein kinase C betaII and zeta messenger RNA: putative biomarkers for colon cancer. *Carcinogenesis*. (1998) 19:253–7. doi: 10.1093/carcin/19.2.253
52. Davidson LA, Lupton JR, Miskovsky E, Fields AP, Chapkin RS. Quantification of human intestinal gene expression profiles using exfoliated colonocytes: a pilot study. *Biomarkers*. (2003) 8:51–61. doi: 10.1080/1354750021000042268
53. Zhao C, Ivanov I, Dougherty ER, Hartman TJ, Lanza E, Bobe G, et al. Noninvasive detection of candidate molecular biomarkers in subjects with a history of insulin resistance and colorectal adenomas. *Cancer Prev Res (Phila)*. (2009) 2:590–7. doi: 10.1158/1940-6207.CAPR-08-0233
54. Ren JM, Semenkovich CF, Gulve EA, Gao J, Holloszy JO. Exercise induces rapid increases in $GLUT_4$ expression, glucose transport capacity, and insulin-stimulated glycogen storage in muscle. *J Biol Chem*. (1994) 269:14396–401.
55. Diverio S, Guelfi G, Barbato O, di Mari W, Egidi MG, Santoro MM. Non-invasive assessment of animal exercise stress: real-time PCR of GLUT4, COX2, SOD1 and HSP70 in avalanche military dog saliva. *Animal*. (2015) 9:104–9. doi: 10.1017/S1751731114002304
56. Maeda S, Ohno K, Uchida K, Igarashi H, Goto-Koshino Y, Fujino Y, et al. Intestinal protease-activated receptor-2 and fecal serine protease activity are increased in canine inflammatory bowel disease and may contribute to intestinal cytokine expression. *J Vet Med Sci*. (2014) 76:1119–27. doi: 10.1292/jvms.14-0060
57. Okanishi H, Kagawa Y, Watari T. Expression of selectins and P-selectin glycoprotein ligand-1 in dogs with lymphocytic–plasmacytic enteritis. *Vet Immunol Immunopathol*. (2014) 161:42–8. doi: 10.1016/j.vetimm.2014.06.005
58. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *Methods*. (2001) 25:402–8. doi: 10.1006/meth.2001.1262
59. Accorsi PA, Carloni E, Valsecchi P, Viggiani R, Gamberoni M, Tamanini C, et al. Cortisol determination in hair and faeces from domestic cats and dogs. *Gen Comp Endocrinol*. (2008) 155:398–402. doi: 10.1016/j.ygcen.2007.07.002
60. Tamanini C, Giordano N, Chiesa F, Seren E. Plasma cortisol variations induced in the stallion by mating. *Acta Endocrinol (Copenh)*. (1983) 102:447–50. doi: 10.1530/acta.0.1020447
61. Peters IR, Calvert EL, Hall EJ, Day MJ. Measurement of immunoglobulin concentrations in the feces of healthy dogs. *Clin Diagn Lab Immunol*. (2004) 11:841–8. doi: 10.1128/CDLI.11.5.841-848.2004
62. Ventrella D, Laghi L, Barone F, Elmi A, Romagnoli N, Bacci ML. Age-related 1H NMR characterization of cerebrospinal fluid in newborn and young healthy piglets. *PLoS ONE*. (2016) 11:e0157623. doi: 10.1371/journal.pone.0157623
63. Simmler C, Napolitano JG, McAlpine JB, Chen SN, Pauli GF. Universal quantitative nMR analysis of complex natural samples. *Curr Opin Biotechnol*. (2014) 25:51–9. doi: 10.1016/j.copbio.2013.08.004
64. Kneen MA, Annegarn HJ. Algorithm for fitting XRF, SEM and PIXE X-ray spectra backgrounds. *Nucl Instruments Methods Phys Res Sect B Beam Interact with Mater Atoms*. (1996) 109–110:209–13. doi: 10.1016/0168-583X(95)00908-6
65. Liland KH, Almøy T, Mevik BH. Optimal choice of baseline correction for multivariate calibration of spectra. *Appl Spectrosc*. (2010) 64:1007–16. doi: 10.1366/000370210792434350
66. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in1H NMR metabonomics. *Anal Chem*. (2006) 78:4281–90. doi: 10.1021/ac051632c
67. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the human metabolome database. *Nucleic Acids Res*. (2007) 35: D521–6. doi: 10.1093/nar/gkl923
68. Barone M, Turroni S, Rampelli S, Soverini M, D'Amico F, Biagi E, et al. Gut microbiome response to a modern paleolithic diet in a western lifestyle context. *PLoS ONE*. (2019) 14:e0220619. doi: 10.1371/journal.pone.0220619
69. Turroni S, Fiori J, Rampelli S, Schnorr SL, Consolandi C, Barone M, et al. Fecal metabolome of the hadza hunter-gatherers: a host-microbiome integrative view. *Sci Rep*. (2016) 6:32826. doi: 10.1038/srep32826
70. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res*. (2013) 41:e1. doi: 10.1093/nar/gks808
71. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing (2018).
72. Hubert M, Rousseeuw PJ, Vanden Branden K. ROBPCA: a new approach to robust principal component analysis. *Technometrics*. (2005) 47:64–79. doi: 10.1198/004017004000000563
73. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics*. (2012) 13:31. doi: 10.1186/1471-2105-13-31
74. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. (2019) 37:852–857. doi: 10.1038/s41587-019-0209-9
75. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from illumina amplicon data. *Nat Methods*. (2016) 13:581–3. doi: 10.1038/nmeth.3869
76. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*. (2016) 4:e2584. doi: 10.7717/peerj.2584
77. Sandri M, Dal Monego S, Conte G, Sgorlon S, Stefanon B. Raw meat based diet influences faecal microbiome and end products of fermentation in healthy dogs. *BMC Vet Res*. (2017) 13:65. doi: 10.1186/s12917-017-0981-z
78. Omatsu T, Omura M, Katayama Y, Kimura T, Okumura M, Okumura A, et al. Molecular diversity of the faecal microbiota of toy poodles in Japan. *J Vet Med Sci*. (2018) 80:749–54. doi: 10.1292/jvms.17-0582
79. Masuoka H, Shimada K, Yasuda KT, Kiyosue M, Oishi Y, Kimura S, et al. Transition of the intestinal microbiota of dogs with age. *Biosci Microbiota Food Heal*. (2017) 36:27–31. doi: 10.12938/bmfh.BMFH-2016-021
80. Santacruz A, Marcos A, Wärnberg J, Martí A, Martin-Matillas M, Campoy C, et al. Interplay between weight loss and gut microbiota composition in overweight adolescents. *Obesity (Silver Spring)*. (2009) 17:1906–15. doi: 10.1038/oby.2009.112
81. Morris C, O'Grada C, Ryan M, Roche HM, Gibney MJ, Gibney ER, et al. The relationship between BMI and metabolomic profiles: a focus on amino acids. *Proc Nutr Soc*. (2012) 71:634–8. doi: 10.1017/S0029665112000699
82. Johnstone AM, Faber P, Andrew R, Gibney ER, Elia M, Lobley G, et al. Influence of short-term dietary weight loss on cortisol secretion and metabolism in obese men. *Eur J Endocrinol*. (2004) 150:185–94. doi: 10.1530/eje.0.1500185
83. Wakshlag JJ, Snedden K, Reynolds AJ. Biochemical and metabolic changes due to exercise in sprint-racing sled dogs: implications for postexercise carbohydrate supplements and hydration management. *Vet Ther*. (2004) 5:52–9.
84. Durocher LL, Hinchcliff KW, Williamson KK, McKenzie EC, Holbrook TC, Willard M, et al. Effect of strenuous exercise on urine concentrations of homovanillic acid, cortisol, and vanillylmandelic acid in sled dogs. *Am J Vet Res*. (2007) 68:107–11. doi: 10.2460/ajvr.68.1.107
85. Rovira S, Muñoz A, Benito M. Hematologic and biochemical changes during canine agility competitions. *Vet Clin Pathol*. (2007) 36:30–5. doi: 10.1111/j.1939-165X.2007.tb00178.x
86. Arokoski J, Miettinen P V, Säämänen AM, Haapanen K, Parviainen M, Tammi M, et al. Effects of aerobic long distance running training (up to 40 $km \cdot day^{-1}$) of 1-year duration on blood and endocrine parameters of female beagle dogs. *Eur J Appl Physiol Occup Physiol*. (1993) 67:321–9. doi: 10.1007/BF00357630
87. Gallelli MF, Monachesi N, Miceli DD, Cabrera Blatter MF, Gomez NV, Meikle A, et al. Plasma ACTH, α-MSH and cortisol variations in the dog during the

oestrous cycle in different photoperiods. *Vet Med (Praha)*. (2016) 60:567–. doi: 10.17221/8495-VETMED
88. Shibuya T, Kaburagi T, Nagai R, Oshiro S. The effects of moderate exercise on secretory igA production in mice depends on dietary carbohydrate intake. *J Clin Biochem Nutr*. (2015) 57:44–9. doi: 10.3164/jcbn.15-21
89. Viloria M, Padilla LE, Rodríguez CR, Luna JA, Garfias RH, Sánchez LP, et al. Effect of moderate exercise on IgA levels and lymphocyte count in mouse intestine. *Immunol Invest*. (2011) 40:640–56. doi: 10.3109/08820139.2011.575425
90. Grosche A, Morton AJ, Graham AS, Sanchez LC, Blikslager AT, Polyak MMR, et al. Ultrastructural changes in the equine colonic mucosa after ischaemia and reperfusion. *Equine Vet J Supl*. (2011) 43:8–15. doi: 10.1111/j.2042-3306.2011.00402.x
91. Ossovskaya V, Bunnett NW. Protease-activated receptors: contribution to physiology and disease. *Physiol Rev*. (2004) 84:579–621. doi: 10.1152/physrev.00028.2003
92. Dokladny K, Zuhl MN, Moseley PL. Intestinal epithelial barrier function and tight junction proteins with heat and exercise. *J Appl Physiol*. (2016) 120:692–701. doi: 10.1152/japplphysiol.00536.2015
93. Biagioli M, Laghi L, Carino A, Cipriani S, Distrutti E, Marchianò S, et al. Metabolic variability of a multispecies probiotic preparation impacts on the anti-inflammatory activity. *Front Pharmacol*. (2017) 8:505. doi: 10.3389/fphar.2017.00505
94. Davis MS, Willard MD, Williamson KK, Steiner JM, Williams DA. Sustained strenuous exercise increases intestinal permeability in racing alaskan sled dogs. *J Vet Intern Med*. (2005) 19:34–9. doi: 10.1111/j.1939-1676.2005.tb02655.x
95. Clarke SF, Murphy EF, O'Sullivan O, Lucey AJ, Humphreys M, Hogan A, et al. Exercise and associated dietary extremes impact on gut microbial diversity. *Gut*. (2014) 63:1913–20. doi: 10.1136/gutjnl-2013-306541
96. Karl JP, Margolis LM, Madslien EH, Murphy NE, Castellani JW, Gundersen Y, et al. Changes in intestinal microbiota composition and metabolism coincide with increased intestinal permeability in young adults under prolonged physiological stress. *Am J Physiol Gastrointest Liver Physiol*. (2017) 312:G559–71. doi: 10.1152/ajpgi.00066.2017
97. Lambert JE, Myslicki JP, Bomhof MR, Belke DD, Shearer J, Reimer RA. Exercise training modifies gut microbiota in normal and diabetic mice. *Appl Physiol Nutr Metab*. (2015) 40:749–52. doi: 10.1139/apnm-2014-0452
98. Huang W-C, Chen Y-H, Chuang H-L, Chiu C-C, Huang C-C. Investigation of the effects of microbiota on exercise physiological adaption, performance, and energy utilization using a gnotobiotic animal model. *Front Microbiol*. (2019) 10:1906. doi: 10.3389/fmicb.2019.01906
99. Cerdá B, Pérez M, Santiago PJD, Aguilera TJF, Soltero GR, Larrosa M. Gut microbiota modification: another piece in the puzzle of the benefits of physical exercise in health? *Front Physiol*. (2016) 7:51. doi: 10.3389/fphys.2016.00051
100. Kim S, Covington A, Pamer EG. The intestinal microbiota: antibiotics, colonization resistance, and enteric pathogens. *Immunol Rev*. (2017) 279:90–105. doi: 10.1111/imr.12563
101. van den Boogaard FE, Brands X, Duitman J, de Stoppelaar SF, Borensztajn KS, Roelofs JJTH, et al. Protease-activated receptor 2 facilitates bacterial dissemination in pneumococcal pneumonia. *J Infect Dis*. (2018) 217:1462–71. doi: 10.1093/infdis/jiy010
102. Edlund A, Yang Y, Yooseph S, He X, Shi W, McLean JS. Uncovering complex microbiome activities via metatranscriptomics during 24 hours of oral biofilm assembly and maturation. *Microbiome*. (2018) 6:217. doi: 10.1186/s40168-018-0591-4
103. Carbajo-Pescador S, Porras D, García-Mediavilla MV, Martínez-Flórez S, Juarez-Fernández M, Cuevas MJ, et al. Beneficial effects of exercise on gut microbiota functionality and barrier integrity, and gut-liver crosstalk in an *in vivo* model of early obesity and non-alcoholic fatty liver disease. *Dis Model Mech*. (2019) 12:dmm.039206. doi: 10.1242/dmm.039206
104. Martín R, Miquel S, Benevides L, Bridonneau C, Robert V, Hudault S, et al. Functional characterization of novel *Faecalibacterium prausnitzii* strains isolated from healthy volunteers: a step forward in the use of *F. prausnitzii* as a next-generation probiotic. *Front Microbiol*. (2017) 8:1226. doi: 10.3389/fmicb.2017.01226

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 27 May 2020
doi: 10.3389/fvets.2020.00275

Computed Tomographic Measures of Funnel-Shaped Lumbar Vertebral Canal and Articular Process Dysplasia Malformations Differ Between German Shepherd and Belgian Malinois Military Working Dogs

Constance J. Dragicevich[1], *Jeryl C. Jones*[1,2*†], *William Bridges*[3] and *Heather Dunn*[1,4]

[1] Department of Animal and Veterinary Sciences, Clemson University, Clemson, SC, United States, [2] South Carolina Translational Research Improving Musculoskeletal Heath Center, Clemson, SC, United States, [3] School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC, United States, [4] Department of Bioengineering, Clemson University, Clemson, SC, United States

OPEN ACCESS

Edited by:
Cynthia M. Otto,
University of Pennsylvania,
United States

Reviewed by:
Kelly Alan Mann,
Colorado State University,
United States
Elizabeth Hare,
Dog Genetics LLC, United States

***Correspondence:**
Jeryl C. Jones
jerylj@clemson.edu

†ORCID:
Jeryl C. Jones
orcid.org/0000-0002-6851-9947

Specialty section:
This article was submitted to
Veterinary Humanities and Social
Sciences,
a section of the journal
Frontiers in Veterinary Science

Received: 17 February 2020
Accepted: 24 April 2020
Published: 27 May 2020

Citation:
Dragicevich CJ, Jones JC, Bridges W and Dunn H (2020) Computed Tomographic Measures of Funnel-Shaped Lumbar Vertebral Canal and Articular Process Dysplasia Malformations Differ Between German Shepherd and Belgian Malinois Military Working Dogs. Front. Vet. Sci. 7:275. doi: 10.3389/fvets.2020.00275

Researchers who study the selection and breeding program criteria for military working dogs aim to help maximize the years of active duty service. Computed tomographic (CT) quantitative phenotyping has been previously described as a method for supporting these research studies. Funnel-shaped lumbar vertebral foramen malformations have been previously described in Labrador retriever military working dogs and proposed to be risk factors for impaired arterial perfusion of nerve tissues during exercise. Articular process dysplasia malformations have been previously described in varying dog breeds and proposed to be risk factors for articular process degenerative joint disease and vertebral foramen stenosis. Aims of this retrospective, cross-sectional study were to describe quantitative CT phenotyping methods for characterizing funnel-shaped lumbar vertebral foramina and articular process dysplasia malformations and to apply these methods in a comparison between groups of German shepherd and Belgian Malinois military working dogs. A military working dog hospital's database was searched for German shepherd and Belgian Malinois dogs aged <6 years that had CT scans of the lumbosacral region during the period of 2008–2016. Observers unaware of CT findings recorded available clinical data for each of the dogs. An observer unaware of clinical data recorded CT measures of funnel-shaped lumbar vertebral foramina and articular process dysplasia malformations for each of dogs and each of the lumbar vertebrae that were available in the scans. A total of 59 dogs were sampled: 41 German shepherd and 18 Belgian Malinois. Articular process dysplasia and funnel-shaped vertebral foramen phenotypic traits were present in both breeds in this sample, with the frequency and quantitative measure of these traits being greater in German shepherd dogs and heavier dogs. Lower weight dogs had a lesser degree of a funnel-shaped foramen at all sampled vertebral locations. A consistent relationship between articular process dysplasia measures and body weight was not seen. Computed tomography measures of funnel shaped vertebral foramina were greater in German shepherd vs. Belgian Malinois dogs at the L7 vertebra ($P < 0.01$). The CT measures of cranial articular process dysplasia were greater in German shepherd vs. Belgian Malinois dogs at the L4 ($P < 0.01$) and L5 ($P < 0.05$) vertebrae.

Keywords: spine, multi-level stenosis, sports medicine, deep phenotyping, athlete

INTRODUCTION

Breeding and procurement programs for working dogs aim to select dogs with the most desirable phenotypic traits for mission-specific working tasks and the highest likelihood for maximizing years of active duty service (1–6). Development of quantitative, deep phenotyping methods is important for supporting researchers who study the criteria used by these programs because these methods allow the use of stronger statistical comparison tests (7). Clinical phenotyping can also be insensitive in stoic, high-drive working dogs due to their tendency to mask clinical signs until disease is advanced (2). Degenerative joint disease and cauda equina syndrome have been reported to be important causes of death or euthanasia in military working dogs (1). Lumbosacral disease has been reported to be a predominant cause of euthanasia or retirement in police working dogs (5). German shepherd dogs and Belgian Malinois are widely used as police and military working dogs around the world (3, 5, 8).

"Funnel-shaped" vertebral foramen malformations have been described in Doberman Pinschers and Labrador retrievers (7–10). This type of malformation has been defined as a trapezoidal or cone-shaped vertebral foramen, with the cranial portion of the foramen being smaller than the caudal portion. A previous study in Labrador retriever military working dogs proposed that this malformation is an undesirable phenotypic trait in the lumbosacral region of high-performance canine athletes due to the potential for impaired arterial blood flow in the cauda equina during strenuous exercise (7). Articular process (facet) dysplasia has been defined as absence (aplasia), incomplete formation (hypoplasia), or increased size (hyperplasia) of the cranial or caudal articular processes (11, 12) and has been described in German shepherds and other breeds (9, 12–15). Dysplasia and/or aplasia of the articular processes in the cervical, thoracic, and vertebral spine has been associated with spinal stenosis at corresponding locations in dogs (9, 12, 13, 15). Abnormal shapes or variable joint angles for articular processes have been described as risk factors for biomechanical instability of the vertebral column and degenerative disease (12, 14–16). Computed tomography (CT) has been previously established as a non-invasive, method for quantifying canine lumbosacral vertebral morphology (7, 14, 17).

The objective of the current, preliminary study was to provide background for future research studies by describing quantitative CT phenotyping methods for characterizing funnel shaped lumbar vertebral foramina and articular process dysplasia malformations and to apply these methods in a comparison between groups of German shepherd and Belgian Malinois military working dogs. The research hypothesis was that CT measures for these three malformations would be greater in German shepherd vs. Belgian Malinois breed groups.

MATERIALS AND METHODS

This study was a retrospective, cross-sectional design. The sample size was based on convenience sampling, i.e., the number of animals that met inclusion criteria during the records search period. With hospital director approval, an ACVR-certified veterinary radiologist (J.J.) searched the database of a tertiary referral military working dog hospital (LTC Daniel E. Holland Military Working Dog Hospital at Lackland Joint Base, San Antonio; IACUC Exempt Protocol No. 2019-04) and retrieved medical record and CT data for dogs that had CT scans of the lumbosacral region from 2008 to 2016. This range began with the year that the CT machine was installed at the hospital, up through the most recent full calendar year of data at the time of initiation of this study. For inclusion in the study, dogs had to be German Shepherd or Belgian Malinois breed, and aged $\leq$6 years at the date of their first presentation for lumbosacral region CT scanning. Dogs of this age group were chosen based on a previous study reporting evidence that dogs younger than 6 years of age were more likely to successfully recover from treatment for lumbosacral stenosis (2). This age group choice was also intended to help minimize possible effects of degenerative disease on vertebral measurements.

An undergraduate research student assigned a research number to each dog and created a randomized list of dogs based on these numbers (https://www.random.org/). In consultation with the veterinary radiologist, a graduate student (C.D.) recorded CT measurements in this random order without knowledge of medical record findings (Horos for Mac, version 3.0.1, www.horosproject.org; Mac Pro, Apple Inc., Cupertino, CA). Hip-extension CT studies were used for all measurements. Measurements were acquired for each available lumbar vertebra in the lumbosacral region. For purposes of this study, the lumbosacral region was defined as L4-S1 (7). If both bone and standard algorithm studies were available, the bone algorithm studies were selected. If scans were acquired using variable slice thicknesses, scans acquired with the smallest available slice thickness were selected. If more than one scan was acquired for a dog, the first scan was selected. Measurements were made using the software program's three-dimensional (3D) multiplanar reformatting (MPR) tool and standard bone window settings (WW: 1,500, WL: 300). For each slice location, the 3D

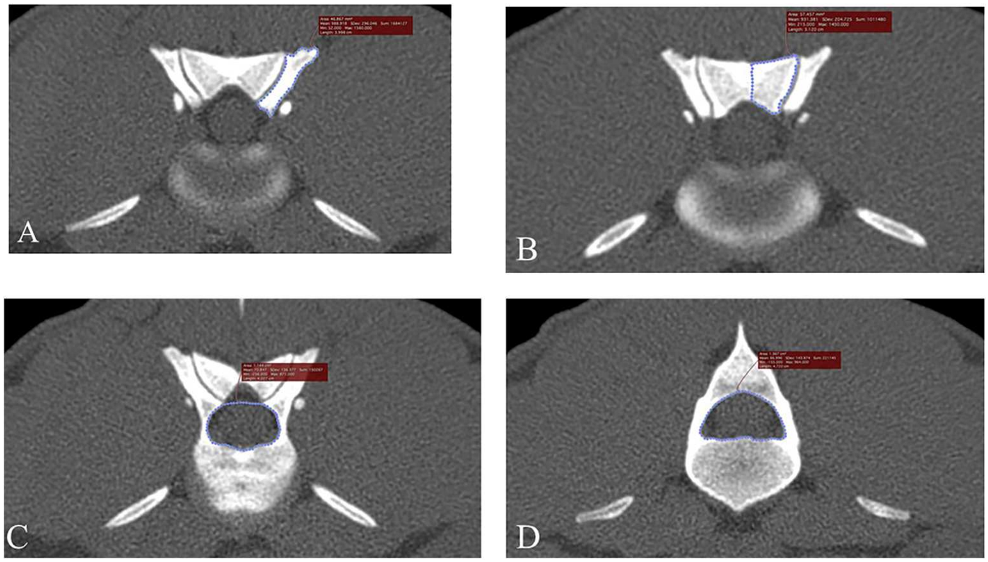


FIGURE 1 | Transverse, bone window CT images illustrating measurement methods for quantifying articular process dysplasia and funnel shaped vertebral foramen phenotypes (Slice thickness 0.625 mm, WW 1500, WL 300, Bone Algorithm). **(A)** Left cranial articular process transverse area; **(B)** left caudal articular process transverse area; **(C)** cranial vertebral foramen transverse area; **(D)** caudal vertebral foramen transverse area. Images are oriented so that dorsal is at the top, ventral is at the bottom, and the patient's left is to the viewer's right.

MPR tool was used to correct for any positioning obliquity before making transverse area measurements.

In addition to randomizing the order of dogs by research number and blinding the observer to medical record findings, bias for CT measurements was further minimized by determining the order for the side of measurements (right, left) using a coin flip. To minimize outside effects of intra-observer variation, each CT measurement was made in triplicate and averaged values were used in further analyses. **Figure 1** illustrates how measurement decisions were made. Articular process transverse areas were recorded in mm^2 and foramen transverse areas were recorded in cm^2. Articular process transverse area was measured using a modified version of previously described methods (9). The articular process measurement area decisions were made in the same way as with foramen areas described above; a transverse plane was placed bisecting the intervertebral disc space, and the first slice that showed both cranial and caudal articular processes was used for measurements (**Figures 1A,B**). The medial border of the caudal articular process was defined as a vertical line traced in the median plane, from the dorsal margin of the articular process to the dorsal margin of the vertebral foramen. When osseous proliferations were present, they were included in articular process ROI tracings only if they were of the same opacity as the adjacent bone. Cranial and caudal vertebral foramen transverse area measurements were performed using freehand tracing of regions of interest (ROIs), and were based on previously published methods (7, 17) (**Figures 1C,D**). The decision for choosing slice locations to perform measurements of the cranial foramen area was made by beginning at a slice bisecting the intervertebral space of the vertebra of interest and the vertebra cranial to it, and then moving caudally one slice at a time until the pedicles around the foramen opening were fully enclosed on transverse view. Likewise, the caudal foramen measurement decision was made by beginning at the intervertebral space between the vertebra of interest and the vertebra caudal to it, and then moving cranially slice-by-slice until the caudal foramen opening was fully enclosed. Data were recorded on a hard-copy data recording sheet first and then transferred to a spreadsheet (Microsoft Excel for Office 365, version 1812).

All statistical analyses were performed by a graduate student (C.D.) in consultation with a statistician (W.B). The following variables were evaluated (see **Table 1** for formulas): cranial articular process transverse area difference (CrAPd), cranial articular process transverse area ratio (CrAPp), caudal articular process transverse area difference (CdAPd), caudal articular process transverse area ratio (CdAPp), vertebral foramen transverse area difference (VFD), and vertebral foramen transverse area ratio (VFP). The primary statistical model used in this study included terms for breed, individual dogs within each breed, and vertebral location. Analysis of variance was used to calculate F-tests for the effects of breed and vertebral location. This model was used to make the comparisons among breeds and among locations. Some additional (or secondary) statistical models were used to determine if the covariates of age, sex, and weight class [dogs were assigned to lower (<32.884 kg) and heavier (>32.885 kg) weight classes] had any influence on breed and locations comparisons. These models included the terms from the primary model, one of the covariates, and the interactions of the covariate with breed and location. Analysis of covariance was used to calculate F-tests for the effects of the covariates. All statistical calculations were performed using statistical analysis software (JMP Pro for Windows 10, Version

TABLE 1 | CT measures of vertebral malformations used for breed comparisons.

Vertebral malformation	CT measure	Abbreviation	Formula[a]
Cranial articular process dysplasia	Cranial articular process transverse area ratio	CrAPp	[(R CrAP-L CrAP)/R CrAP]*100
	Cranial articular process transverse area difference	CrAPd	R CrAP – L CrAP
Caudal articular process dysplasia	Caudal articular process transverse area ratio	CdAPp	[(R CdAP – L CdAP)/R CdAP]*100
	Caudal articular process transverse area difference	CdAPd	R CdAP – L CdAP
Funnel-shaped vertebral foramen	Vertebral foramen transverse area ratio	VFP	[(Caudal foramen transverse area – Cranial foramen transverse area)/Caudal foramen transverse area]*100
	Vertebral foramen transverse area difference	VFD	Caudal foramen transverse area – Cranial foramen transverse area

[a]*R, right; L, left; Cr, cranial; Cd, caudal, AP, articular process; Ca, foramen.*

13.2.0, SAS Institute Inc., Cary, NC) and statistical significance was defined as $P < 0.05$.

RESULTS

Descriptions of the sampled dogs are provided in **Table 2**. A total of 59 dogs met initial inclusion criteria, comprising 18 Belgian Malinois and 41 German shepherds. All dogs were scanned based on a consensus opinion between the primary care clinician and a veterinary radiologist. Reasons listed for scanning in the "other" category of **Table 2** included the following: tenesmus, hypertension, possible torn ureter, pain on palpation, previous lumbosacral surgery, left metacarpal five fracture, possible pain, and electromyography abnormalities. One dog underwent lumbosacral surgery prior to scanning and a small defect was noted in the margins of the cranial L5 foramen, however this defect did not alter the ability to perform region of interest tracing and the dog was therefore included in analyses. Entire L4 through L7 vertebrae were available in CT scans of 52 dogs. Portions of L4 were not available for five dogs, portions of L4 and L5 were not available for one dog, and portions of L7 were not available for one dog. Missing vertebral locations were listed as "not available" for analyses. The German shepherd group included 32 males and nine females. The Belgian Malinois group included 10 males and eight females. Weights ranged from 20.4 to 43.8 kg. The average weight of the German shepherds was 32.84 ± 4.42 kg and the average weight of the Belgian Malinois was 30.58 ± 5.48 kg.

All animals were scanned in dorsal recumbency, using the same multi-slice CT scanner (Lightspeed VCT, GE Medical Systems, Pewaukee, WI). Technical parameters were as follows: slice thickness 0.625–1.000 mm, kVp 120, mAs 16–700, and matrix 512 × 512 mm. While hip extension scans were available for all included dogs, the degree of hip extension was not standardized. Some dogs were positioned with hips maximally extended and stifle joints adducted (OFA position) and some dogs were positioned with hips in a relaxed extension position and stifle joints abducted (frogleg position).

Figure 2 illustrates an example of articular process dysplasia (**Figure 2A**) and funnel-shaped vertebral foramen malformations (**Figures 2B–D**). Detailed descriptions of vertebral measurements that were used for calculating the six variables used for breed comparison analyses are provided in **Table 3**. **Figures 3–5** provide graphical breed group comparisons for the six measured variables.

TABLE 2 | Clinical characteristics of sampled dogs.

Variable	Category	German shepherd *N* = 41	Belgian Malinois *N* = 18
Average age (years, SD)		3.68 ± 1.78	4 ± 1.94
Age range (years)		1–6	1–6
Average weight (kg, SD)		32.84 ± 4.42	30.58 ± 5.48
Weight range (kg)		24.95–44.0	20.41–41.28
Number of dogs in each weight class[a]	Lower	21	12
	Upper	20	6
Sex	Male	32	10
	Female	9	8
Reasons for CT scanning	Neurologic deficits	15	2
	Lumbosacral pain	15	5
	Vertebral pain (other than lumbosacral)	1	0
	Problems with hindlimbs	22	6
	Inability/reluctance to perform certain actions	3	2
	Possible lumbosacral disease	4	1
	Diagnosed with vertebral disease	6	1
	Unspecified lameness	3	2
	Research study	4	0
	Other	4	3
	Reason not specified	1	1

[a]*Weight class defined as lower if ≤32.884 kg and upper if >32.884 kg. "Other reasons" included dogs that were scanned for reasons unrelated to musculoskeletal or neurological issues of the lumbosacral region or hindlimbs. Twenty dogs had more than one reason for scan listed.*

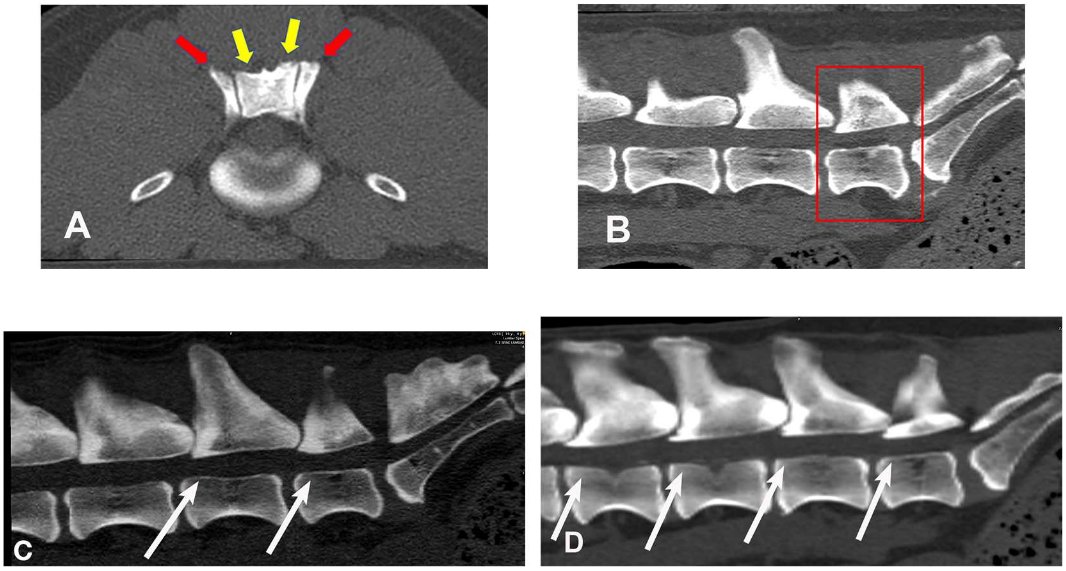


FIGURE 2 | Bone window, CT images illustrating examples of articular process dysplasia and funnel-shaped vertebral foramen malformations (slice thickness 0.625 mm, WW 1500, WL 300, Bone algorithm). **(A)** transverse planar image illustrating articular process dysplasia involving cranial (red arrows) and caudal (yellow arrows) articular processes; **(B)** sagittal planar image illustrating funnel-shaped vertebral foramen at L7 (boxed region); **(C)** sagittal planar image illustrating funnel-shaped vertebral foramina at L6 and L7 (white arrows); **(D)** sagittal, planar image illustrating funnel-shaped vertebral foramina at L4, L5, L6, and L7 (white arrows). Sagittal planar images are oriented so that dorsal is at the top, ventral is at the bottom, cranial is to the viewer's left, and caudal is to the viewer's right.

TABLE 3 | Vertebral foramen and articular process CT measurements, by breed and vertebra[a].

CT measurement	L4		L5		L6		L7	
	GSD	BM	GSD	BM	GSD	BM	GSD	BM
R CrAP	$N = 37$	$N = 16$	$N = 40$	$N = 18$	$N = 41$	$N = 18$	$N = 41$	$N = 18$
Mean (mm^2)	53.7 (16.9–98.3)	61.0 (30.3–94.3)	47.7 (18.2–86.00)	53.9 (36.6–79.8)	49.6 (17.0–83.3)	51.6 (33.3–72.7)	72.2 (26.2–110.7)	69.6 (44.9–86.0)
Range (mm^2)								
L CrAP	$N = 37$	$N = 16$	$N = 40$	$N = 18$	$N = 41$	$N = 18$	$N = 41$	$N = 18$
Mean (mm^2)	60.0 (18.9–99.4)	58.4 (33.0–84.0)	52.6 (14.1–90.8)	51.8 (33.3–67.8)	52.8 (14.1–86.5)	56.1 (36.7–73.4)	73.6 (26.4–113.7)	68.3 (50.8–88.3)
Range (mm^2)								
R CdAP	$N = 39$	$N = 18$	$N = 41$	$N = 18$	$N = 41$	$N = 18$	$N = 40$	$N = 18$
Mean (mm^2)	48.5 (27.9–75.0)	50.3 (33.8–69.7)	44.1 (20.6–70.0)	49.2 (29.0–71.2)	54.8 (27.4–96.1)	56.8 (31.4–76.9)	76.3 (42.3–113.8)	67.2 (47.2–84.4)
Range (mm^2)								
L CdAP	$N = 39$	$N = 18$	$N = 41$	$N = 18$	$N = 41$	$N = 18$	$N = 40$	$N = 18$
Mean (mm^2)	51.6 (27.8–72.2)	55.3 (32.7–72.1)	46.8 (19.2–74.9)	50.1 (30.4–68.6)	56.9 (26.4–84.4)	57.1 (33.5–72.6)	78.7 (40.1–111.3)	69.8 (35.3–86.56)
Range (mm^2)								
CrVF	$N = 37$	$N = 16$	$N = 40$	$N = 18$	$N = 41$	$N = 18$	$N = 41$	$N = 18$
Mean (cm^2)	1.21 (0.88–1.56)	1.23 (1.02–1.58)	1.27 (0.96–1.61)	1.33 (1.15–1.67)	1.12 (0.78–1.49)	1.19 (0.98–1.52)	0.85 (0.62–1.16)	0.97 (0.76–1.15)
Range (cm^2)								
CdVF	$N = 39$	$N = 18$	$N = 41$	$N = 18$	$N = 41$	$N = 18$	$N = 40$	$N = 18$
Mean (cm^2)	1.42 (0.93–1.84)	1.50 (1.23–1.85)	1.52 (1.11–1.95)	1.62 (1.26–2.03)	1.45 (1.01–1.91)	1.53 (1.16–1.93)	1.43 (1.19–1.81)	1.43 (1.09–1.88)
Range (cm^2)								

CT, computed tomographic; L4, 4th lumbar vertebra; L5, 5th lumbar vertebra; L6, 6th lumbar vertebra; L7, 7th lumbar vertebra; GSD, German Shepherd dog; BM, Belgian Malinois dog; R, right; L, left; Cr, cranial; Cd, caudal; AP, articular process; VF, vertebral foramen; mm^2, millimeters squared; cm^2, centimeters squared.

German shepherds displayed a wider variance than Belgian Malinois among all six variables at all vertebral locations, except for the cranial vertebral foramen at L7. Analysis of covariance revealed no significant interaction between breed and sex or age among any of the variables at any of the vertebral locations ($P > 0.05$). Breed had a significant effect on CrAP and VFD

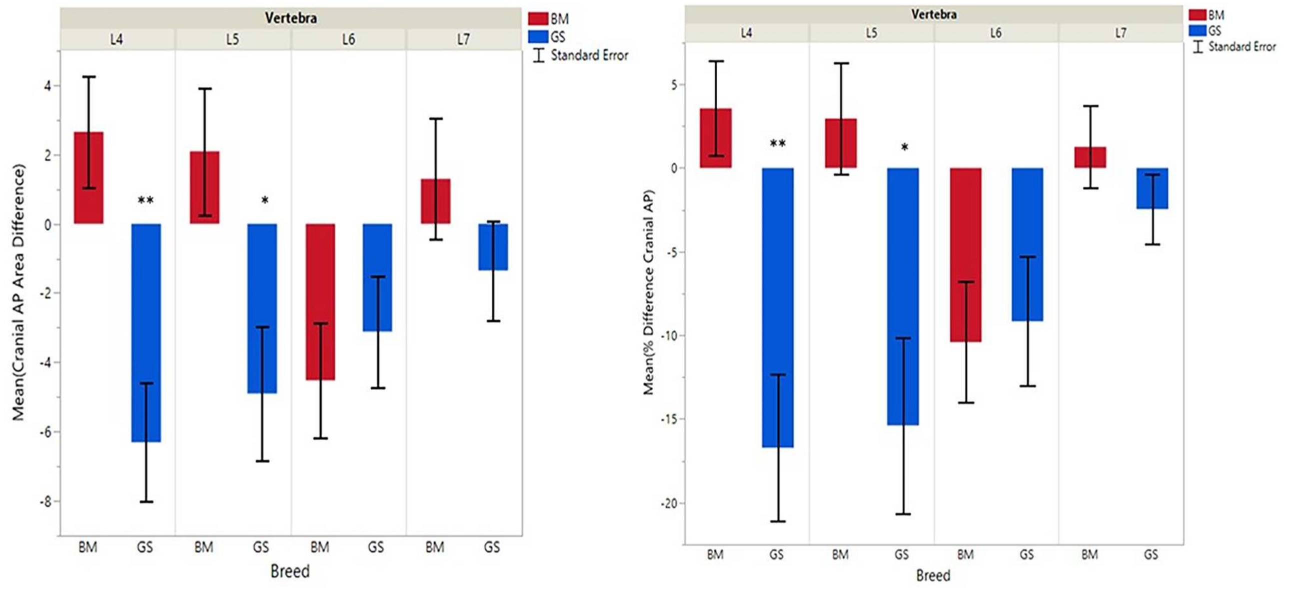


FIGURE 3 | Graph illustrating differences for right vs. left cranial articular processes and mean transverse area ratios (%) (CrAPd and CrAPp). The * indicates a *P*-value of <0.05, and ** indicates $P < 0.01$.

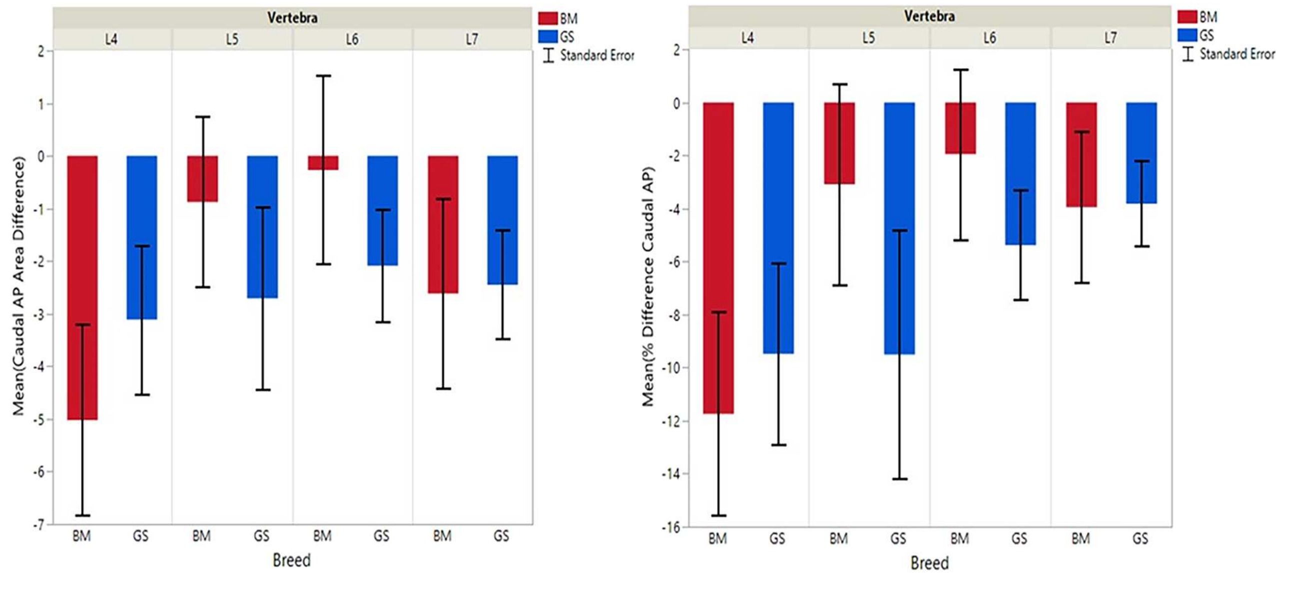


FIGURE 4 | Graph illustrating differences for right vs. left caudal articular processes and mean transverse area ratios (%) (CdAPd and CdAPp).

comparisons. The means of CrAPd and CrAPp at L4 ($P < 0.005$) and L5 ($P < 0.05$), VFD at L7 ($P < 0.01$), and VFP at L7 ($P < 0.0001$) varied between breed groups. When breed was the only factor, there was no significant difference in means between German shepherds and Belgian Malinois for either CT measure of caudal articular process dysplasia (CdAPd and CdAPp) at any of the vertebral locations ($P > 0.05$). The CrAPd and CrAPp means were greater for German shepherds than Belgian Malinois at L4 ($P < 0.01$) and L5 ($P < 0.05$), and there was no significant difference in these measures between breeds at L6 and L7. The VFP and VFD means were greater in German shepherds than Belgian Malinois at L7 ($P < 0.0001$ and $P < 0.01$, respectively).

Breed was significantly affected by the covariate weight for CrAPd at L7 ($P < 0.05$), CrAPp at L5 and L7 ($P < 0.05$), VFP at L7 ($P < 0.05$), and for CdAPd and CdAPp at L4 and L7 ($P < 0.05$). The mean CrAPd and CrAPp values were subjectively greater in German shepherds at L4 and L5 in the lower weight class, and at L4 in the upper weight class. The mean CrAPd and

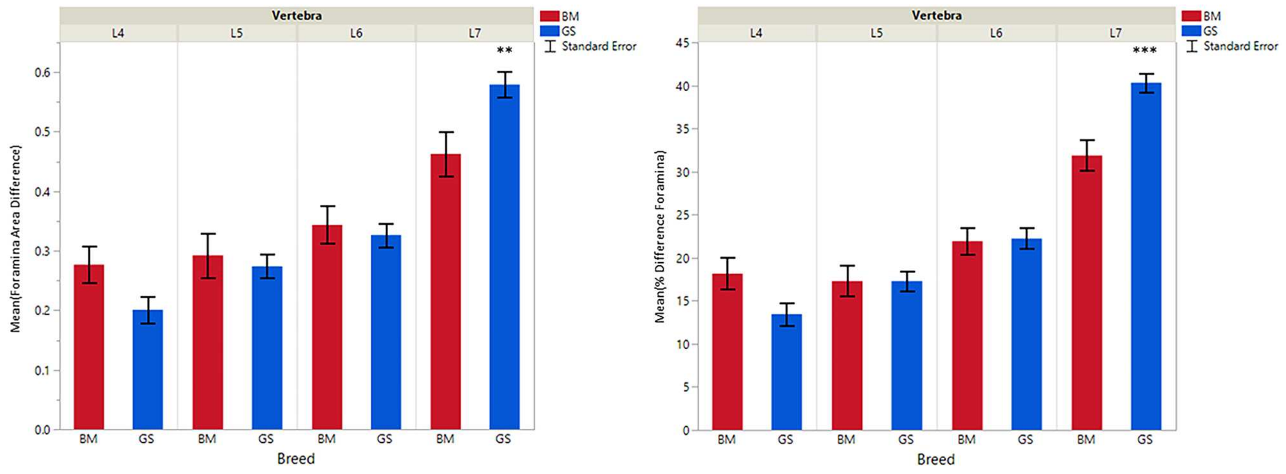


FIGURE 5 | Graph illustrating differences for caudal vs. cranial vertebral foramina and mean transverse area ratios (%) (VFD and VFP). The ** indicates a *P*-value of <0.01, and *** indicates $P < 0.0001$.

CrAPp values were subjectively greater in Belgian Malinois at L6 in the lower weight class and at L5 and L6 in the upper weight class. Mean CdAPd values were subjectively larger in German shepherds at all vertebral locations in the lower weight class, and were subjectively larger at L4, L6, and L7 for Belgian Malinois in the upper weight class. As a function of weight, German shepherds had subjectively greater mean VFP and VFD values at L7 vs. Belgian Malinois. However, none of these subjective differences were statistically significant.

DISCUSSION

Intentions of this preliminary study were to provide background for use in future research studies evaluating the selection criteria for military working dogs. Findings supported the study hypothesis in that CT measures of articular process dysplasia and funnel-shaped vertebral malformation phenotypes were greater in German shepherd vs. Belgian Malinois breed groups. However, study findings also indicated that these malformations were present in both breeds. In examining the variable of breed alone, German shepherds had significantly greater right-left articular process dysplasia at L4 and L5, and a significantly more frequent occurrence of funnel-shaped vertebral foramina at L7 than Belgian Malinois. Subjectively, as German shepherds got heavier, the degree of cranial and caudal articular process dysplasia worsened. Among Belgian Malinois, as weight increased, the degree of cranial and caudal articular process dysplasia improved minutely.

For both breeds, mean VFP and VFD values subjectively increased at each individual vertebral location, with the highest subjective differences at L7 (excepting the VFP value for Belgian Malinois at L4). In every dog measured, the cranial vertebral foramen area of L7 was smaller than the caudal foramen area, and for all but three dogs some degree of a funnel-shaped vertebral foramen malformation was also identified at L4, L5, and L6 locations. Based on our review of the literature, this vertebral malformation has not been previously reported in German shepherd or Belgian Malinois dogs. A funnel-shaped vertebral foramen is considered to be a malformation because standard anatomic reference texts describe vertebral foramina as having cranial portions nearly equal in size to caudal portions (18). Primary stenosis has been previously defined in dogs and humans as an abnormally narrow vertebral foramen caused by a congenital or developmental error in vertebral bone formation (19, 20). This primary stenosis can reduce the functional reserve capacity of the vertebral foramen and exacerbate effects of subsequent acquired stenosis due to degenerative disease. Authors of the current study propose that a funnel-shaped vertebral foramen is a form of primary stenosis in dogs. Experimental studies in pigs and dogs have demonstrated evidence that even mild cauda equina compression at more than one level results in impaired venous outflow, increased interstitial pressure in cauda equina nerve tissues located between the levels of compression, and reduced arterial perfusion of these cauda equina nerve tissues (21–23). Human studies have described an increase in these effects during exercise due to higher oxygen demand by nerve tissues and increased congestion of vertebral venous plexus vessels (24).

Dysplasia of the cranial articular processes has been considered uncommon, while caudal articular process dysplasia has been well-documented (11, 12, 15, 25). Current study findings therefore differed from previous studies in that cranial articular process dysplasia was found to be present in both German Shepherd and Belgian Malinois breeds. We did not identify a significant effect of breed alone on CrAPd comparisons, contrary to existing publications describing caudal articular process dysplasia (11, 12, 15). In both breeds and at all locations sampled in the current study, the average left caudal articular process transverse area values were larger than the right. The reason for this finding remains unknown. Overall et al. proposed that repetitive, obsessive-compulsive behaviors, such as spinning and tail-chasing could be exacerbated by excessive confinement (26). Vertebral structures are formed by endochondral ossification during skeletal growth and this process is affected by both internal and external forces (18, 27). It is therefore possible that spinning behaviors in kenneled, military working dogs could be a risk factor for articular process malformations affecting one side more than the other.

The clinical relevance of asymmetrical articular process morphology has been described in several previous publications (2, 13–16). A case report described a young German shepherd with neurological symptoms caused by dysplastic lumbar articular process joints (15). Another study described a common finding of straight-edged articular process joints among German shepherds, whereas other breeds featured rounded articular processes which were more desirable for vertebral stability (14). Three studies associated articular process dysplasia with lumbosacral stenosis and reduction in spine strength and range of movement (13, 15, 16). Three studies have associated lumbar pain with enlarged (hypertrophied) articular processes of the caudal lumbar spine in both dogs and humans (2, 15).

Findings from the current study indicated that weight had a significant effect on breed comparisons for at least one location for all variables except VFD. This finding supported other studies reporting evidence that heavier dogs were at a higher risk for lumbosacral stenosis (7, 8, 28). A surprising finding in this study was a lack of evidence of sex effects on comparisons. This finding did not support previous reports describing evidence that male dogs were at greater risk for lumbosacral lesions (8, 28–30). It is possible that these discordant results occurred because females were lower weight than males on average in our sample. It is also possible that a lack of significance for sex as a covariate in this study could have been caused by unequal group sizes. There was also no significant effect of age in the current study. This is likely due to sample bias in that we chose to include only dogs aged ≤6 years. This choice was based on our intention to focus on dogs that were in the age group previously reported to be more likely to recover following treatment for lumbosacral stenosis (2) and also our intention to prioritize early detection.

Findings from the current study also indicated that funnel-shaped vertebral foramen and articular process dysplasia malformations were present in both breeds at L4. It is currently not known whether lumbar vertebrae cranial to L4 may also have exhibited these malformations. One study reported high prevalence of articular process dysplasia in the entire spine in dogs that were euthanized for reasons unrelated to the study (31). Though the author concluded that this high prevalence discounted the clinical significance of these lesions, more recent studies have linked articular process dysplasia to clinical signs of spinal disease in dogs and horses (10, 11, 15, 16, 32). Multi-level bony foramen stenosis has also been reported as a cause of clinical signs of spinal disease in dogs (7, 10, 30, 33, 34). One study identified a moderate to high heritability of several lumbosacral disease-related anatomical features in the lumbar spine in a large sample of German shepherds (34). Another study identified potential candidate genes for lumbosacral stenosis in a sample of 8 military working Labrador retrievers, and recommended that this candidate gene be further evaluated in larger samples and other breeds (35). Repeating the present study with a larger sample size of dogs would allow a "normal" vs. "affected" population to be established. This would be helpful background information to further explore related candidate genes and to more definitively assess the clinical relevance of the phenotypes described in our study.

One of the limitations for the current study was the fact that dog positioning was not standardized. Some dogs were positioned with maximal extension (OFA position) and other dogs were positioned with a relaxed, frog leg extension position at the discretion of the presiding veterinarian. Because of these variations in patient positioning, we chose to focus on measurements of lumbosacral spinal components that were least likely to be affected by positioning, i.e., the vertebral foramina and articular processes. We also used MPR to correct for positioning obliquity to help ensure that true transverse planar slices were used for all measurements. Slice thicknesses for all dogs were either 0.625 or 1.000 mm, therefore effects of slice thickness variation were considered to be negligible. Reader bias was minimized by having one observer who was unaware of clinical findings perform all measurements, by selecting dogs in random order based on research numbers, and by randomizing measurement order for each dog with a coin flip. Intra-observer variability was also minimized by averaging triplicate measurements for all statistical comparisons. Interobserver repeatability was not assessed because one observer made all measurements. Group sizes for sampled dogs were also unequal. The statistician therefore selected analyses that minimized effects of unequal group sizes for this study.

In conclusion, findings from this preliminary study indicated that articular process dysplasia and funnel-shaped vertebral foramen phenotypic traits were present in German Shepherd and Belgian Malinois military working dogs, with the frequency and quantitative measures of these traits being greater in German Shepherd dogs and heavier dogs. Lower weight dogs had a lesser degree of the funnel-shaped foramen phenotype at all sampled vertebral locations. A consistent relationship between articular process dysplasia measures and body weight was not seen. Quantitative CT phenotyping characteristics of these two vertebral malformations may be helpful background for future research studies evaluating the criteria for selection and breeding programs in military working dogs and other high-performance canine athletes. Future longitudinal studies are needed to test associations between these CT phenotypic measures and later development of clinical disease.

DATA AVAILABILITY STATEMENT

The datasets for this article are not publicly available because of patient confidentiality requirements detailed in the study's IACUC protocol. Requests to access the datasets should be directed to Army Public Health Center ATTN: MCHB-IP-V 8252 Blackhawk Rd. Aberdeen Proving Ground, MD 21010-5403; email usarmy.apg.medcom-aphc.mbx.iph-vet@mail.mil.

ETHICS STATEMENT

The animal study was reviewed and approved by the LTC Daniel E Holland Military Working Dog Hospital at Lackland Joint Base, San Antonio; IACUC Exempt Protocol No. 2019-04.

AUTHOR CONTRIBUTIONS

CD, JJ, WB, and HD: conception and design, revising article for intellectual content, and final approval of the completed article. CD and JJ: acquisition of data and drafting the article. CD, JJ, and WB: analysis and interpretation of data.

FUNDING

This project was supported by the Clemson University Department of Animal and Veterinary Sciences (supported the graduate research assistant), Clemson Creative Inquiry Program (supported the undergraduate research assistant), and the South Carolina Center of Biomedical Research Excellence for Translational Research Improving Musculoskeletal Health (SC TRIMH) NIH P20GM121342 (supported open access publishing fees and portions of the salaries for Dr. Jones and Dr. Bridges). All other faculty salaries were supported by Clemson University.

ACKNOWLEDGMENTS

The authors would like to thank undergraduate research students Emma Rovitz and Amanda Biddlecome for their assistance in recording and managing medical record data. The authors also acknowledge the veterinarians and staff at the LTC Daniel E. Holland Military Working Dog Hospital for their assistance in retrieving dog data for the current study.

REFERENCES

1. Moore GE, Burkman KD, Carter MN, Peterson MR. Causes of death or reasons for euthanasia in military working dogs: 927 cases (1993–1996). *J Am Vet Med Assoc.* (2001) 219:209–14. doi: 10.2460/javma.2001.219.209
2. Linn LL, Bartels KE, Rochat MC, Payton ME, Moore GE. Lumbosacral stenosis in 29 military working dogs: epidemiologic findings and outcome after surgical intervention (1990–1999). *Vet Surg.* (2003) 32:21–9.
3. Moody JA, Clark LA, Murphy KE. Working dogs: history and applications. In: Ostrander EA, Giger U, Lindblad-Toh K, editors. *The Dog and Its Genome*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (2006). p. 1–18.
4. Evans RI, Herbold JR, Bradshaw BS, Moore GE. Causes for discharge of military working dogs from service: 268 cases (2000–2004). *J Am Vet Med Assoc.* (2007) 231:1215–20. doi: 10.2460/javma.231.8.1215
5. Worth AJ, Sandford M, Gibson B, Stratton R, Erceg V, Bridges J, et al. Causes of loss or retirement from active duty for New Zealand police German shepherd dogs. *Anim Welfare.* (2013) 22:167–74. doi: 10.7120/09627286.22.2.167
6. Cobb M, Branson N, McGreevy P, Lill A, Bennett P. The advent of canine performance science: offering a sustainable future for working dogs. *Behav Process*. (2015) 110:96–104. doi: 10.1016/j.beproc.2014.10.012
7. Mukherjee M, Jones JC, Holaskov I, Raylman R, Meade J. Phenotyping of lumbosacral stenosis in Labrador retrievers using computed tomography. *Vet Radiol Ultrasound.* (2017) 58:565–80. doi: 10.1111/vru.12520
8. Pogorevc E, Lukanc B, Seliškar A, Pelc R, Zorko B. Radiological comparison of lumbosacral anatomy between German and Belgian Shepherd (Malinois) working dogs. *Sloven Vet Res.* (2016) 53:219–27. doi: 10.1053/jvet.2003.50001
9. Martin-Vaquero P, da Costa RC, Lima CGD. Cervical spondylomyelopathy in Great Danes: a magnetic resonance imaging morphometric study. *Vet J.* (2014) 201:64–71. doi: 10.1016/j.tvjl.2014.04.011
10. Martin-Vaquero P, Costa RC. Evaluation of traditional and novel radiographic vertebral ratios in Great Danes with versus without cervical spondylomyelopathy. *Vet Radiol Ultrasound.* (2014) 55:488–95. doi: 10.1111/vru.12159
11. Bouma JL. Congenital malformations of vertebral articular processes in dogs. *Vet Clin North Am Small Anim Pract.* (2016) 46:307–26. doi: 10.1016/j.cvsm.2015.10.006
12. Bertram S, ter Haar G, Decker S. Caudal articular process dysplasia of thoracic vertebrae in neurologically normal French bulldogs, English bulldogs, and Pugs: prevalence and characteristics. *Vet Radiol Ultrasound.* (2018) 59:396–404. doi: 10.1111/vru.12609
13. Jones JC, Sorjonen DC, Simpson ST, Coates JR, Lenz SD, Hathcock JT, et al. Comparison between computed tomographic and surgical findings in nine large-breed dogs with lumbosacral stenosis. *Vet Radiol Ultrasound.* (1996) 37:247–56. doi: 10.1111/j.1740-8261.1996.tb01226.x
14. Seiler GS, Häni H, Busato AR, Lang J. Facet joint geometry and intervertebral disk degeneration in the L5-S1 region of the vertebral column in German Shepherd Dogs. *Am J Vet Res.* (2002) 63:86–90. doi: 10.2460/AJVR.2002.63.86
15. Penderis J, Schwarz T, McConnell JF, Garosi LS, Thomson CE, Dennis R. Dysplasia of the caudal vertebral articular facets in four dogs: results of radiographic, myelographic and magnetic resonance imaging investigations. *Vet Rec.* (2005) 156:601–5. doi: 10.1136/vr.156.19.601
16. Breit S, Künzel W. Breed specific osteological features of the canine lumbosacral junction. *Ann Anat.* (2001) 183:151–7. doi: 10.1016/S0940-9602(01)80037-6
17. Jones JC, Wright JC, Bartels JE. Computed tomographic morphometry of the lumbosacral spine of dogs. *Am J Vet Res.* (1995) 56:1125–32.
18. Evans HE, de Lahunta A. *The Skeleton, Miller's Anatomy of the Dog*. St. Louis, MO: Elsevier Saunders (2013). p. 113–23.
19. Tarvin G, Prata RG. Lumbosacral stenosis in dogs. *J Am Vet Med Assoc.* (1980) 177:154–9.
20. Verbiest H. Developmental stenosis of the bony lumbar vertebral canal. *Acta Orthop Belg.* (1987) 53:373–87.
21. Yoshizawa H, Kobayashi S, Hachiya Y. Blood supply of nerve roots and dorsal root ganglia. *Orthop Clin North Am.* (1991) 22:195–211.
22. Takahashi K, Olmarker K, Holm S, Porter RW, Rydevik B. Double-level cauda equina compression: an experimental study with continuous monitoring of intraneural blood flow in the porcine cauda equina. *J Orthop Res.* (1993) 11:104–9. doi: 10.1002/jor.1100110112
23. Kobayashi S, Takeno K, Miyazaki T, Kubota M, Shimada S, Yayama T, et al. Effects of arterial ischemia and venous congestion on the lumbar nerve root in dogs. *J Orthop Res.* (2008) 26:1533–40. doi: 10.1002/jor.20696
24. Kobayashi S. Pathophysiology, diagnosis and treatment of intermittent claudication in patients with lumbar canal stenosis. *World J Orthop.* (2014) 5:134–45. doi: 10.5312/wjo.v5.i2.134
25. Damur-Djuric N, Steffen F, Hässig M, Morgan JP, Flückiger MA. Lumbosacral transitional vertebrae in dogs: classification, prevalence, and association with sacroiliac morphology. *Vet Radiol Ultrasound.* (2006) 47:32–8. doi: 10.1111/j.1740-8261.2005.00102.x
26. Overall KL, Dunham AE. Clinical features and outcome in dogs and cats with obsessive-compulsive disorder: 126 cases (1989–2000). *J Am Vet Med Assoc.* (2002) 221:1445–52. doi: 10.2460/javma.2002.221.1445
27. Khan IM, Redman SN, Williams R, Dowthwaite GP, Oldfield SF, Archer CW. The development of synovial joints. In: Schatten GP, editor. *Current Topics in Developmental Biology*. San Diego, CA: Elsevier, Inc. (2007). p. 1–36. doi: 10.1016/S0070-2153(06)79001-9
28. Danielsson F, Sjostrom L. Surgical treatment of degenerative lumbosacral stenosis in dogs. *Vet Surg.* (1999) 28:91–8. doi: 10.1053/jvet.1999.0091
29. Jones JC, Banfield CM, Ward DL. Association between postoperative outcome and results of magnetic resonance imaging and computed tomography in

working dogs with degenerative lumbosacral stenosis. *J Am Vet Med Assoc.* (2000) 216:1769–74. doi: 10.2460/javma.2000.216.1769
30. De Decker S, Saunders JH, Duchateau L, Pey P, Van Ham LML. Radiographic vertebral canal and vertebral body ratios in Doberman Pinschers with and without clinical signs of caudal cervical spondylomyelopathy. *Am J Vet Res.* (2011) 72:958–66. doi: 10.2460/ajvr.72.7.958
31. Morgan JP. Congenital anomalies of the vertebral column of the dog: a study of the incidence and significance based on a radiographic and morphologic study. *J Am Vet Radiol Soc.* (1968) 9:21–9. doi: 10.1111/j.1740-8261.1968.tb01082.x
32. VanderBroek A, Stubbs NC, Clayton HM. Osseous pathology of the synovial intervertebral articulations in the equine thoracolumbar spine. *J Equine Vet Sci.* (2016) 44:67–73. doi: 10.1016/j.jevs.2016.05.015
33. Johnson P, De Risio L, Sparkes A, McConnell F, Holloway A. Clinical, morphologic, and morphometric features of cranial thoracic spinal stenosis in large and giant breed dogs. *Vet Radiol Ultrasound.* (2012) 53:524–34. doi: 10.1111/j.1740-8261.2012.01953.x
34. Ondreka N, Amort KH, Stock KF, Tellhelm B, Klumpp SW, Kramer M, et al. Skeletal morphology and morphometry of the lumbosacral junction in German shepherd dogs and an evaluation of the possible genetic basis for radiographic findings. *Vet J.* (2013) 196:64–70. doi: 10.1016/j.tvjl.2012.07.015
35. Mukherjee M, Jones JC, Yao J. Lumbosacral stenosis in Labrador retriever military working dogs–an exomic exploratory study. *Canine Genet Epidemiol.* (2017) 4:12–4. doi: 10.1186/s40575-017-0052-6

Disclaimer: The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Army, Department of Defense, nor the United States Government.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 27 May 2020
doi: 10.3389/fvets.2020.00250

Effect of Handler Knowledge of the Detection Task on Canine Search Behavior and Performance

Mallory T. DeChant[1]*, *Cameron Ford*[2] and *Nathaniel J. Hall*[1]

[1] Department of Animal and Food Sciences, Texas Tech University, Lubbock, TX, United States, [2] Ford K9, Las Vegas, NV, United States

Detection dogs are commonly trained and tested under conditions in which the handler or the evaluator knows the true presence or absence of a target odor. Previous research has demonstrated that when handlers are deceived and led to believe that a target odor is present, more false alerts occur. However, many detection teams operate under *unknown* conditions, and it remains unclear how handler knowledge (or lack thereof) of odor presence/absence influences the dog's behavior. The aim of this study was to evaluate if knowing the number of hides placed influenced detection dog performance in an applied search environment. Professional ($n = 20$) and sport ($n = 39$) detection handler-dog teams were asked to search three separate areas (area 1 had one hide, area 2 had one hide, area 3 was blank). Handlers in the Unknown Group were not told any information on the number of hides whereas the Known Group were told there was a total of two hides in the three areas. The sport Unknown Group spent a longer duration (69.04 s) searching in area 3 compared to the sport Known Group ($p = 0.004$). Further, sport dogs in the Unknown group looked back to the handler more frequently. When a miss did occur, dogs of both sport and professional handlers showed an increase interest in the location of the target odor compared to a comparison location. Critically, however, there was no difference in false alerts between the Known Group and Unknown Group for sport or professional handlers. In a second experiment, fourteen professional, and thirty-nine sport teams from Experiment 1 conducted an additional search double-blind and an additional search single-blind. Both sport and professional-handler dog teams had statistically similar accuracy rate under single and double blind conditions. Overall, when handlers knew the number of hides, it led to significant changes in search behavior of the detection team but did not influence the overall false alert rates.

Keywords: detection dog, handler bias, behavior, olfaction, double-blind

OPEN ACCESS

Edited by:
Erik Hofmeister,
Auburn University, United States

Reviewed by:
Helen Zulch,
Dogs Trust, United Kingdom
Lowell Paul Waggoner,
Auburn University, United States

***Correspondence:**
Mallory T. DeChant
mallory.dechant@ttu.edu

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences, a section of the journal Frontiers in Veterinary Science

Received: *15 February 2020*
Accepted: *16 April 2020*
Published: *27 May 2020*

Citation:
DeChant MT, Ford C and Hall NJ (2020) Effect of Handler Knowledge of the Detection Task on Canine Search Behavior and Performance. Front. Vet. Sci. 7:250. doi: 10.3389/fvets.2020.00250

INTRODUCTION

Dogs have been utilized for a myriad of professional detection jobs for items such as explosive devices (1–3), narcotics (4, 5), semen stains for crime scenes (6), human odor (7–11), cancer (12–14), and conservation (15–19). Dogs are also utilized for non-professional detection jobs such as bird hunting (20) and sport detection in the United States and in growing sport detection dog organizations across the world. Sport detection dogs typically detect essential oils and compete in sanctioned events through the American Kennel Club, The National Association of Canine Scent Work, and United Kennel Club. Whether the dog is utilized for professional or non-professional detection work, a handler always works with the dog.

The focus of detection work is typically on the detection dog itself; however, an undeniable bond between the handler and the dog could influence performance (21). Handlers have the responsibility to recognize and call the dog's change in behavior toward an odor, or trained alert, to locate the target source. Failure to call an alert could result in a missed target source which can have detrimental implications for certain professional (e.g., explosive or live find search and rescue) detection dogs. Further, calling an alert when a dog is not showing the appropriate alert behavior or unintentionally cuing a dog to alert could lead to unnecessary emergency (e.g., explosive dog) or improper search (e.g., narcotics dog).

Maintaining a strong trained alert behavior in an operational setting is critical because detection dogs are often subjected to stressful situations in which they work (15). To prevent deterioration of the alert behavior and to maintain olfactory performance, handler-dog teams train on a regular basis, frequently referred to as "maintenance training" (22). Moser and McCulloch (23) reviewed various training details from publications in which dogs were trained to detect cancer. Training regimens range from a frequency of 15–30 times per day (24, 25) to a duration of 1–2 hours per day (26). The ranges used by typical professional detection dogs have not yet been clearly reported but are likely variable depending on the type of work. Non-professionals such as sport detection handler-dog teams commonly train once a week but may vary depending on opportunities available in their city. Frequent training sessions could potentially maintain a strong alert behavior and increase accuracy in odor discrimination; however, there is limited research on how different maintenance training regimens influence detection performance.

Typically, a dog remains with the same handler, or handlers, throughout their working life; however, there are some circumstances (e.g., retirement or death) where a new handler could take possession of the dog. For teams that have a consistent one-handler to one-dog working relationship, maintaining the same handler-dog team is preferred because changing handlers impacts performance by increasing response time (27), the dog becoming distracted more often, and potentially less accurate (28).

Handlers, however, may potentially negatively impact working performance by unintentionally cuing the dog (29–32). A classic example of unintentional postural and facial cues is the famous "Clever Hans" example, in which a horse's incredible skills was later demonstrated to be remarkably controlled by unintentional cues (33). A similar phenomenon could occur with handler-dog teams where the dog responds to unintentional handler cues (34). Dogs, in particular, are quite adept at reading human communicative cues (35, 36), and perhaps may learn to utilize these cues during a search, even when unintentional by the handler.

In a critical study, Lit et al. (31) found that police canine handler belief that an odor source was present led to significant rates of false alerts by the dogs. In this study, handlers were deceived by informing them that a target odor was placed in a certain location, when marked. Importantly, no target was actually placed, but instead, sometimes a non-target distractor odor was presented (31). This was contrasted to areas in which no obvious maker was placed for the handler. Overall, more false alerts occurred when the handler was led to believe a target odor was present (31).

Importantly, no follow-up to this study has been conducted. There remain several limitations that require further study. First, Lit et al. (31) deceived handlers as to the presence of an odor which caused handlers to influence the dog's indication; however, a double blind or unknown condition is perhaps more realistic of an operational setting, when handlers are unsure of target odor presence, rather than being told by a researcher that target odors are present. Second, although more false alerts were called (i.e., the handler is the person "calling" the alerts), there was no direct behavioral observation of the dog to investigate the dog's behavior and to what degree they showed a true alert (i.e., the dog displayed the behavior trained for indicating a target odor at a specific location such that the handler can recognize).

The objectives of this study were to determine whether knowing the number of odor sources prior to a real-life scenario search influenced detection dog outcomes and to evaluate if single-blind or double-blind searches influenced detection dog outcomes. In addition to professional handlers, sport handlers were also utilized to increase the sample size of detection dog handlers. Dog behavior during each study was coded by video, and certain team descriptor covariates (i.e., handler training experience, frequency of double-blind training, and frequency of blank training) were utilized as covariates for data analysis.

MATERIALS AND METHODS

This study was conducted at various field locations and was approved as an observational Institutional Animal Care and Use protocol. The Institutional Review Board at Texas Tech University approved this study (IRB2019-501). All participants provided written consent. Recruitment statements were sent out via email to various professional detection dog agencies and sport scent detection trainers following an approved script. Participants were volunteers and completed a survey following the search. There were two components to this study: Experiment 1) a three area search with varying levels of knowledge of the number of target odors present and Experiment 2) a single-blind and double-blind search. This study was conducted in six cities across the United States and at seven different facilities that permitted dog search teams during after-hours. Specific cities and facilities will remain confidential to maintain anonymity of professional and sport handler-dog teams. Search areas were chosen and secured for training by the local organizations (and was therefore not experimentally controlled). Size of the search areas were kept as consistent as possible across locations (e.g., typical search was in a standard classroom approximately 85 m^2). Twenty professional and 39 sport handler-dog teams were recruited for Experiment 1. Fourteen professional and 39 sport handler-dog teams from Experiment 1 participated in Experiment 2. All searches were video recorded for data collection.

Handler-Dog Teams

The handler-dog teams that participated were divided into professional and sport handler groups, based on self-report. All searches were typical of an operational search or a scent detection trial. All sport detection searches utilized birch essential oil. Professional detection searches utilized smokeless black powder, ammonium nitrate, trinitrotoluene dynamite, composition-4, cast booster, detonation cord, marijuana, methamphetamine, cocaine, heroin, bed bugs, kerosene, and 75% evaporated gasoline depending on the type of detection dog and the agencies' training aids provided. This variability in target odor was required given the different agencies and odor stimuli to which they had previously trained the dogs. Training aids were supplied by the respective departments. All distractors utilized in searches were cotton balls and plastic gloves. Dogs were allowed on or off-leash depending on handler preference. Each search was timed for duration but did not have a time limit.

Experiment 1: Knowledge of the Number of Hides Present

Three indoor and temperature regulated areas were utilized for three searches. The facility used was secured and organized by the professional or sport team organization, and therefore little control over specific facilities was available, but most were schools or business with similar sized rooms (~83 m^2). To control within a given study-site, the three areas were selected based on identifying three approximately similar sized rooms or parts of a room, as facility geometry allowed. The professional handler-dog teams searched the entire area for the target odor provided by the organization's trainer that was concealed in a small tin not visible to the handler or canine and was ~50 cm from the ground. The sport handler-dog teams searched containers provided by the trainer (small cardboard boxes and plastic shoe bins) placed on the floor spread across a similar area. The target odor was also provided by the trainer and was concealed in a small tin not visible to the handler or canine. Twenty professional handler-dog teams and 39 sport handler-dog teams were recruited. Handlers were instructed to search and clear each area before proceeding to the next area. Handlers therefore determined when to move from one area to the next. A handler was instructed to call an alert by placing their hand up and saying "alert." The experimenter would then provide feedback as to whether it was correct or a false alert. No feedback was given if a dog missed a target, until after the study completion. The primary experimenter video recorded the dog through the search from as far back as possible and was knowledgeable of the target odor placement. Handlers in the Unknown Group (Professional $n = 10$; Sport $n = 19$) were instructed from Script 1 (see below) and Known Group (Professional $n = 10$; Sport $n = 20$) were instructed from Script 2 (see below). Areas were set up as follows: area 1 contained one target odor, area 2 contained one target odor, area 3 had no target odor and plastic gloves as a distracting odor. Detailed data recorded during the searches are defined in **Table 1** (see below).

Script 1. Unknown Group instructions for three area searches.

You will be searching a total of three areas.

When your dog alerts, call out the location and I will immediately tell you if it is correct or not.

Script 2. Known Group instructions for three area searches.

You will be searching a total of three areas.

There are exactly 2 target odors in total, over the three areas.

One room has no target odor in it.

When your dog alerts, call out the location and I will immediately tell you if it is correct or not.

Experiment 2: Single-Blind and Double-Blind Comparison

Two indoor and temperature regulated areas were utilized for single-blind and double-blind searches. Fourteen professional handler-dog teams and 39 sport handler-dog teams were recruited from Experiment 1 immediately after (same day). All handlers were instructed from Script 3 (see below). Areas were set up as follows: single-blind contained one target odor and a cotton ball as a distractor odor, double-blind contained one target odor and a cotton ball as a distracting odor. Detailed data recorded during the single-blind and double-blind searches are defined in **Table 2** (see below). Accuracy was defined if the handler-dog team correctly called an alert to the target odor. Trials were videotaped via a tripod positioned to record the entire search area. All other experimental arrangements such as the facilities used were identical to Experiment 1.

Script 3. Handler instructions for single-blind and double-blind searches.

Mallory will be the judge and can help you through this experiment.

Mallory will know where the odor is located.

For the single-blind search, Mallory will be watching the search.

For the double-blind search, Mallory will be facing a wall and will not be watching the search.

When your dog alerts, call out the location and Mallory will immediately tell you if it is correct or not.

Survey

A custom survey was created using Qualtrics but distributed via paper copy (www.qualtrics.com; see **Supplementary Material** for complete survey). Dog handlers (de-identified via a participant ID number) answered questions about their dog, whether they were a professional or sport handler (do they receive money for detection services or not), detection training frequency (reported as 1: daily, 2: 4–6 times a week, 3: 2–3 times a week, 4: once a week, 5: 2–3 times a month, 6: once a month, 7: less than once a month), years of experience, listed the odors the dog is trained on, frequency of double-blind training (reported as 1: always, 2: most of the time, 3: about half the time, 4: sometimes, 5: never), frequency of conducting blank searches (reported as 1: multiple times a training session, 2: once a training session, 3: every other training session, 4: every 3–5 training session, 5: almost never, 6: never), the dog's alert behavior, handler belief that target odor was present in area 3 (reported as 1: strongly agree, 2: agree, 3: somewhat agree, 4: neither agree nor disagree, 5: somewhat disagree, 6: disagree, 7: strongly disagree), and survey measures of canine behavior which will be analyzed

TABLE 1 | Data recorded during three area searches.

Term	Definition	Area recorded	ICC
Total search duration	Start of search to when handler called "clear"	Area 1, 2, 3	0.99
Hit	Handler called alert when dog is at target odor source	Area 1, 2	NA
False	Handler called alert when dog is not at target odor source	Area 1, 2, 3	NA
Correct rejection	Handler correctly did not call an alert	Area 3	NA
Miss	Handler did not call an alert and dog did not locate target odor	Area 1, 2	NA
Target investigate duration	If dog "misses," duration of sniffing time at target odor source	Area 1, 2	0.83
Non-target investigate duration	If dog "miss," duration of sniffing time at selected non-target area along search path	Area 1, 2	0.92
False alert duration	Start of search to when handler called alert and dog is not at target odor source	Area 1, 2	0.92
Hit duration	Start of search to when handler called alert and dog is at target odor source	Area 1, 2	0.98
Lookback	Number of times the dog turned their head back to look at the handler	Area 1, 2, 3	0.97

later with a larger sample. Handlers were given multiple choice options (they could select more than 1) and an optional fill in the blank. The survey was administered to the handlers after completion of study 1 and study 2 via a paper copy. Handlers were not required to answer all of the questions. An ID was given to the survey to correspond to the video from study 1 and study 2.

TABLE 2 | Data recorded during both single-blind and double-blind searches.

Term	Definition
Hit	Handler called alert when dog is at target odor source
False	Handler called alert when dog is not at target odor source

Hypotheses

Based on previous research indicating important effects of handlers on dog search performance (27, 28) and that handler belief may factor into this (31) we hypothesized that individuals' knowledge regarding the number of odor hides present, or whether the moderator of the search knew the presence of the hides, would lead to differences in canine performance. We developed the below specific hypotheses in which we expected handler knowledge to influence performance and additional hypotheses related to the dog's search behavior:

Experiment 1

1. Sport and Professional handlers in the Unknown condition would report (via survey) a higher expectation to find an odor in area 3 than the Known group that were previously informed there was a blank room and only a total of 2 hides.
2. Sport and Professional handlers in the Unknown condition will search longer in the final area compared to handlers that Know there is a blank room.
3. When a dog's misses a target odor (in search areas 1 and 2), they would have shown more investigate behavior toward the target odor than a presumably equivalent comparison blank area along the search path.
4. Dogs will make a false alert later in a search compared to a hit in search areas 1 and 2.
5. Dog will look back toward the handler more frequently in search area 3 for handlers that do not know the number of hides.
6. There will be more false alerts when handlers do not know there is a blank room, compared to handlers that Know one room is clear.
7. Training practices, such as frequency of training under double blind conditions or frequency of training with blank searches will influence the effect of the above hypotheses.

Experiment 2

8. Accuracy will be higher in the single blind condition compared to double-blind condition indicating an effect of the presence of a judge or moderator of the search that is knowledgeable of target odor positioning.

Statistical Analysis

Data from the sport and professional handler teams were analyzed separately because the level of training is different between the teams, the target odors are vastly different in terms of volatility, and the motivation for working with the dog is different. We did, however, compare trends observed across both groups, but did not formally compare these distinct groups. Logistic regression was utilized to compare accuracy in the single-blind vs. double-blind conditions (Experiment 2). Linear mixed model was utilized for all other comparisons (Experiment 1). Analyses were conducted using R [R version 3.5.1, www.r.project.org; (37)] and the lme4 (38) and lmerTest (39) packages. P-values and Z-tests of the logistic regression model were obtained from the summary function of the lmerTest package. The following were covariates utilized from the survey: years of handler experience with detection dog, frequency of no target odor (blank) training runs, and frequency of double-blind training. Interobserver agreement was calculated for video-scored behavior for 10% of the total number of participants by a coder naïve to the hypotheses. All Interclass correlation coefficient for video coded variables are in **Table 1** and were obtained from the ICC function of the psych package using R (40).

To address hypothesis 3 that when a dog misses, they searched the target area longer than a comparison area, the duration of sniffing was recorded by video toward the target odor and a comparison area. The comparison area was chosen by selecting a location that could have equally of held the target odor (e.g., another container or comparable item where a target maybe hidden) was along the same search path (i.e., along the same wall or edge) and was 3 m from the target odor (to avoid cross-contamination). Sniffing duration to both locations was then scored by video.

RESULTS

Overall Performance

Table 3 shows an overview of overall performance for sport and professional dogs in each area. **Table 3** shows the percentage of handler teams that made each response (out of 20 professional teams and 39 sport teams). Hits and misses were coded as mutually exclusive given only one target odor was present per search area, but false alerts could occur in addition to a hit or miss. Total search duration was coded from the start of the search until the handler cleared the area. Overall, a majority of professional and sport handlers correctly identified the target odor and most searches were completed within 2 min. There were no clear systematic differences in performance across the search areas, and a two sample proportion test indicates that there was no overall statistical difference in the proportion of handlers false alerting between area 2 and area 3 (professional: $\chi^2 = 2.00$, df = 1, $p = 0.16$; sport: $\chi^2 = 2.60$, df = 1, $p = 0.11$) or area 1 and 3 (professional: $\chi^2 = 2.00$, df = 1, $p = 0.16$; sport: $\chi^2 = 0.83$, df = 1, $p = 0.36$).

Experiment 1: Knowledge of the Number of Hides Present

Hypothesis 1

We tested whether the handler's expectation for an odor being present in the final search area was higher for individuals in the Unknown group, given that prior to area 3, each room contained a target odor, and they had no knowledge of the number of hides compared to the Known group who were informed there was a blank room and only 2 hides. Overall, there was no difference in self-reported handler expectation for odor presence in area 3 ($t = -0.224$, $p = 0.82$).

Hypothesis 2

Although handler expectations did not seem to change, we next evaluated whether handler and canine search behavior changed based on knowledge of the test parameters. Sport-handler dog teams in the Unknown group searched for 69.04 s longer than the known group ($t = 3.056$, df = 33, $p = 0.004$; **Figure 1**), indicating knowledge that one room was blank reduced overall search time in the final area. In addition, search duration of sport handlers was associated with length of previous training, such that as length of previous experience increased, the dogs searched for a shorter period of time ($t = -2.268$, df = 33, $p = 0.048$). There was no effect of frequency of double-blind training ($t = 1.198$, df = 33, $p = 0.23$) or frequency of blank training runs ($t = 1.232$, df = 33, $p = 0.22$) detected. This analysis, however, includes all participants, whether the individuals accurately found the hides in the first two rooms or not. To evaluate whether this result was the same for participants that accurately identified the first two hides (thus the Known group should be fully knowledgeable that the final room is blank), we subset our data to only these participants. This left eight participants in the Known group and 10 in the Unknown group. In this subset, the trend remained similar, such that handlers in the Unknown group spent ∼91 s longer in search area three, although statistically this effect only reached the trend level ($t = 1.89$, df = 16, $p = 0.08$). Professional-handler dog teams in the Unknown group searched for 7.47 s longer than the Known group (similar direction of effect); however, this difference did not reach statistical significance ($F = 0.035$, $t = 0.189$, df = 14, $p = 0.85$; **Figure 1**). In addition, there was no effect of frequency of double-blind training ($t = 0.861$, df = 14, $p = 0.40$), frequency of blank training runs ($t = 0.466$, df = 14, $p = 0.64$), or years of experience ($t = 1.65$, df = 14, $p = 0.12$). The sample size for professional handlers that correctly identified both finds in the first two areas, however, was too limited to evaluate as a subset as was done for sport handlers.

Hypothesis 3

A miss was scored when the handler did not call an alert when a target odor was present in the search area. To quantify whether the dog showed significant interest to the target odor, but simply did not show a readable alert by the handler, we compared investigation time to the target odor area compared to an equally sized comparison non-target area along the search path. When a miss occurred, sport-handler dog teams investigated the target odor for 3.04 s longer than a comparison non-target location ($t = -3.11$, df = 25, $p < 0.01$; **Figure 2**). This was not influenced by the frequency of double-blind training ($t = -0.49$, df = 7, $p = 0.63$), frequency of blank training runs ($t = 0.32$, df = 17, $p = 0.75$), or years of experience ($t = 1.14$, df = 13, $p = 0.27$). Professional-handler dog teams investigated the target odor for 9.52 s longer than a comparison non-target location ($t = -2.52$, df = 14, $p = 0.02$; **Figure 2**). This was not related to reported frequency of double-blind training ($t = 0.339$, df = 8, $p = 0.74$), frequency of blank training runs ($t = -0.648$, df = 6, $p = 0.54$), or years of experience ($t = 0.85$, df = 4, $p = 0.44$). Both sport and professional-handler dogs spent a longer time investigating the target odor than a comparable area along the search path.

Hypothesis 4

To evaluate if false alerts occurred later in the search than hits, we scored the time from the start of each search in areas 1 and 2 until the first outcome of either a false alert or a hit. We then compared

TABLE 3 | Overview of Professional and Sport dog-handler team performance in each area. Each cell shows the percentage of handlers that made the respective response (out of 20 professionals and 39 sport handlers).

	Area 1		Area 2		Area 3	
	Professional	**Sport**	**Professional**	**Sport**	**Professional**	**Sport**
Hits	50%	79.48%	65%	51.28%	N/A	N/A
False alerts	15%	38.46%	15%	30.76%	40%	51.28%
Misses	50%	20.51%	35%	48.71%	N/A	N/A
Correct Rejections	N/A	N/A	N/A	N/A	60%	48.71%
Average search duration (mean (s) ± sd)	95.29 ± 49.48	59.51 ± 40.47	120.54 ± 134.421	67.27 ± 71.55	127.35 ± 79.60	96.04 ± 69.76

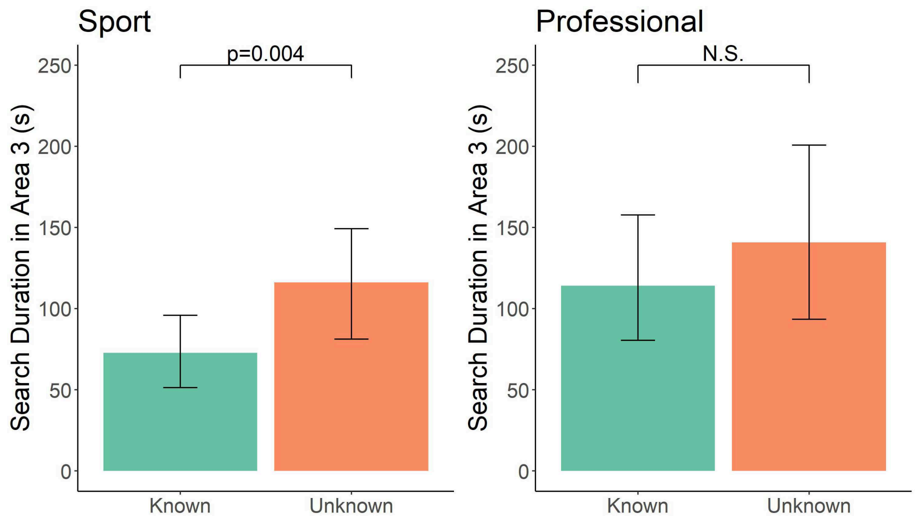


FIGURE 1 | Known vs. Unknown group in sport and professional handler dog teams for total search duration in area 3. Error bars represent 95% confidence intervals.

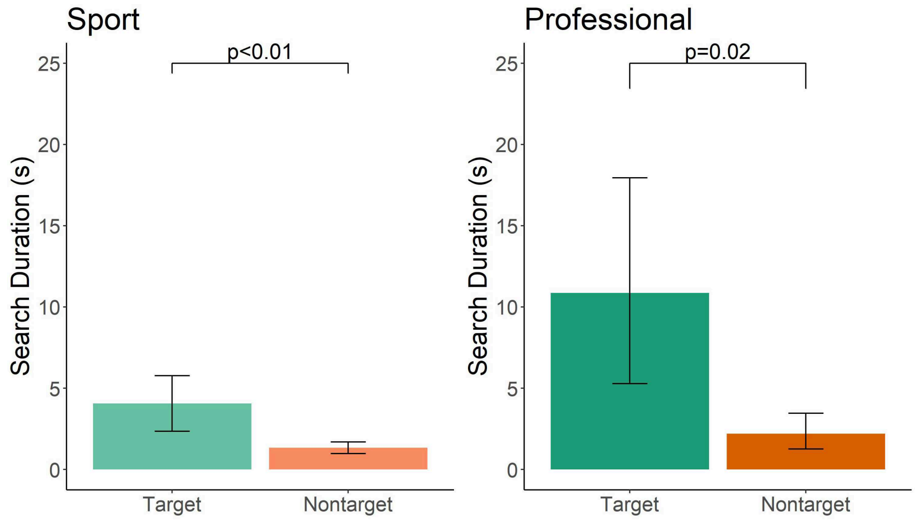


FIGURE 2 | Target vs. non-target investigation duration in sport and professional handler dog teams. Error bars represent 95% confidence intervals.

the duration of the search between hits and false alerts. Sport-handler dog teams that false alerted, had on average, searched for 19.70 s longer than when the average hit was called ($t = -2.682$, df = 55, $p = 0.009$; **Figure 3**). This effect was not associated with the frequency of double-blind training ($t = 0.201$, df = 32, $p = 0.84$), frequency of blank training runs ($t = 1.617$, df = 35,

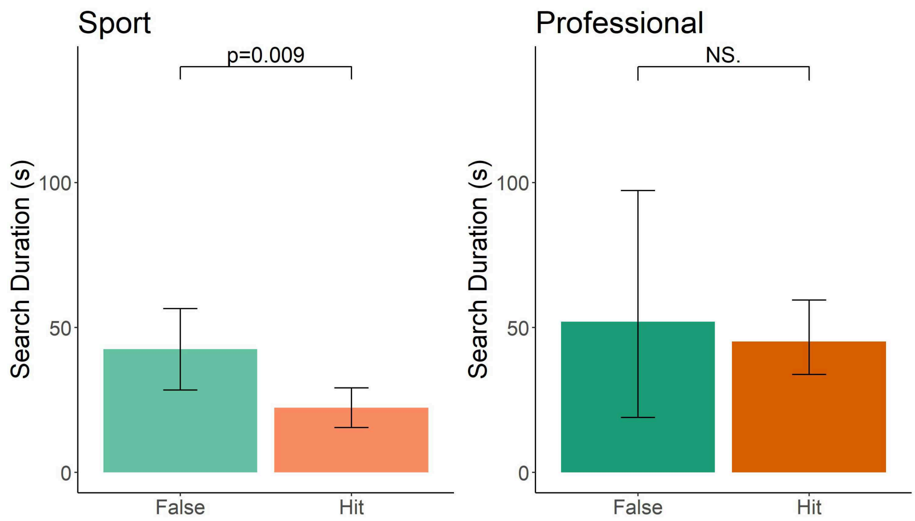


FIGURE 3 | Time of search until either a hit or false alert is called in sport and professional handler dog teams. Error bars represent 95% confidence intervals.

$p = 0.11$), or years of experience ($t = -0.397$, df = 37, $p = 0.69$). Professional-handler dog teams showed no difference between false alert and hit duration (**Figure 3**; $t = 0.572$, df = 9, $p = 0.58$), which may be due to the fact that only four false alerts were the first thing called in areas 1 and 2 for professionals. No covariates were analyzed for professional-handler teams because of small sample size for comparison.

Hypothesis 5

Sport-handler dog teams in the Unknown group had about three times more lookbacks than the Known group in search area 3 ($t = 2.522$, df = 33, $p = 0.01$; **Figure 4**). This was not influenced by the frequency of double-blind training ($t = 1.173$, df = 33, $p = 0.25$), frequency of blank training runs ($t = -0.515$, df = 33, $p = 0.64$), or years of experience ($t = -1.536$, df = 33, $p = 0.13$). Professional-dog teams did not have a difference between the number of lookbacks in the Unknown and Known group ($t = 0.507$, df = 18, $p = 0.50$; **Figure 4**). No covariates were analyzed for professional-handler teams because of small sample size for the number of dogs engaging in lookbacks (only six professional dogs looked back to the handler in area 3).

Hypothesis 6

Sport and professional-handler dog teams in the Unknown group did not false alert more in area 3 compared to the Known group (sport: $z = 0.484$, df = 37, $p = 0.62$; professional: $z = -0.711$, df = 18, $p = 0.47$; **Figure 5**). This was not associated with the frequency of double-blind training (sport: $z = 0.384$, df = 37, $p = 0.70$; professional: $z = 1.180$, df = 18, $p = 0.23$), frequency of blank training runs ($z = -1.640$, df = 37, $p = 0.10$; professional: $z = -0.434$, df = 18, $p = 0.66$), or years of experience (sport: $z = -0.454$, df = 37, $p = 0.65$; professional: $z = 0.705$, df = 18, $p = 0.48$). When restricting analysis to only sport-handler participants that correctly identified the first two targets, there remained no difference in the number of false alerts ($z = 0.979$, df = 17, $p = 0.33$). Interestingly, of the eight sport handlers that accurately identified the two targets in the first two room and knew there were only two targets total and that one area was blank, two of these eight handlers still called false alerts.

Hypothesis 8

In the double blind and single blind search, both sport and professional-handler dog teams had a statistically similar accuracy rate. Logistic regression relating accuracy in the search (1 or 0) to the condition (single blind vs. double blind), showed no significant effect (sport: $z = 0.295$, $p = 0.76$, professional: $z = 0.435$, $p = 0.43$; **Figure 6**).

DISCUSSION

Overall, the present results highlight that handler knowledge of the testing parameters influences search behavior, by increasing the search duration in a blank area and increasing the number of lookbacks to the handler by the sport dog. Importantly, however, this did not directly translate to increased rates of false alerts when the number of target odors was Unknown compared to Known. This highlights the need to consider handler knowledge in a search task, as it could lead to a handler limiting search time if the handler believes there is no odor present or extending a search because they believe something is present.

In addition, analysis of canine behavior revealed some interesting findings. First, when dogs did miss a target odor, both sport and professional dogs tended to investigate the location of the target odor more than a comparable area along the search path (from about 3.04 s for sport and 9.52 s for professional). This suggest the dogs did at least somewhat identify the target odor presence through a change in investigation behavior but did not quite show sufficient behavior for an alert. Under controlled conditions, canine investigation behavior does seem to be indicative of whether the response may be correct. Concha et al. (41) found that dogs showed reduced sniffing before a true negative response than before any other response

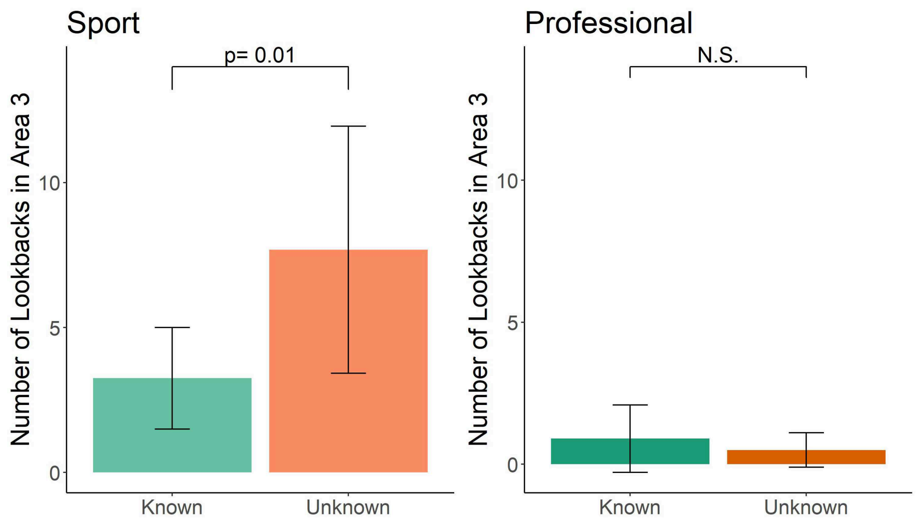


FIGURE 4 | Known vs. Unknown group in sport and professional handler dog teams for number of lookbacks in area 3. Error bars represent 95% confidence intervals.

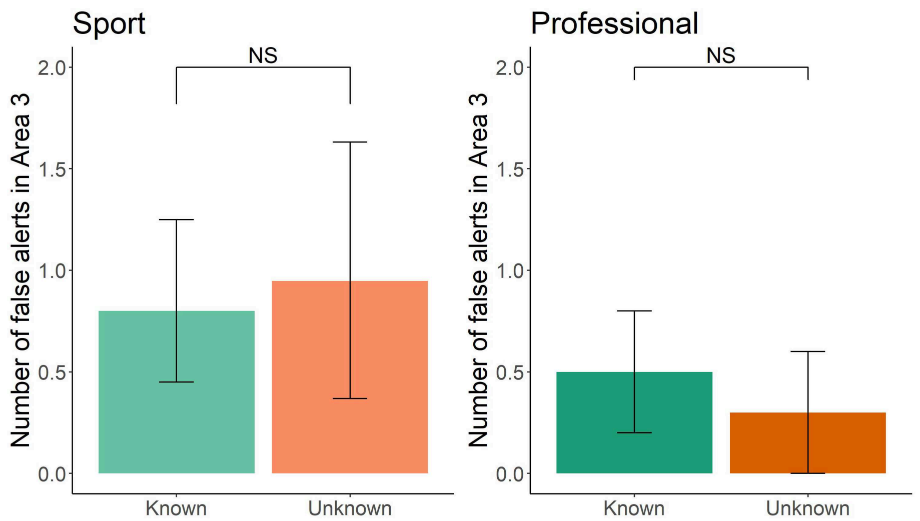


FIGURE 5 | Known vs. Unknown group in sport and professional handler dog teams for number of false alerts in area 3. Error bars represent 95% confidence intervals.

(e.g., false positive, false negative and true positives). Perhaps, together, these results indicate that misses could be limited further by careful observation of this investigation behavior by the handler. Further, increased investigation behavior could be used to indicate to a handler to manipulate the local environment to increase odor availability to the dog (e.g., perhaps by opening a drawer if it is not an explosives dog). It should be noted though, that we did not induce a strong bias in the handler to believe an odor was present. Perhaps, had we done so, any increased investigation time may have led to false alerts, as demonstrated by Lit et al. (31). Thus, interpretation of investigation behavior should be made cautiously, but our present results highlight that during a miss, the target odor was investigated longer than a non-target area.

In addition, analysis of the dog behavior indicated that for sport dogs only, false alerts tended to occur on average later in the search than did hits. This suggests that perhaps if a dog fails to find a target odor after a typical period of time, a false alert may become more likely. This may be an attempt to receive a reinforcer in scenarios in which the dog does not find a target odor, but further testing is required.

In addition, Sport dogs (but not the professional group) in the Unknown group looked back at the handler three times more compared to the Known group in area 3. Previous research has investigated dogs' propensity to look back at an owner when given an "unsolvable" task such as food trapped in a box that can't be opened (42–48). The underlying reasoning dogs engage in high rates of looking back to a human in these unsolvable conditions is still under discussion (44, 47, 48). Interestingly, however, the rate of looking back seems to be inversely related to persistence on the task (46), and positively related to a strong history of training and experience with humans (43). Water

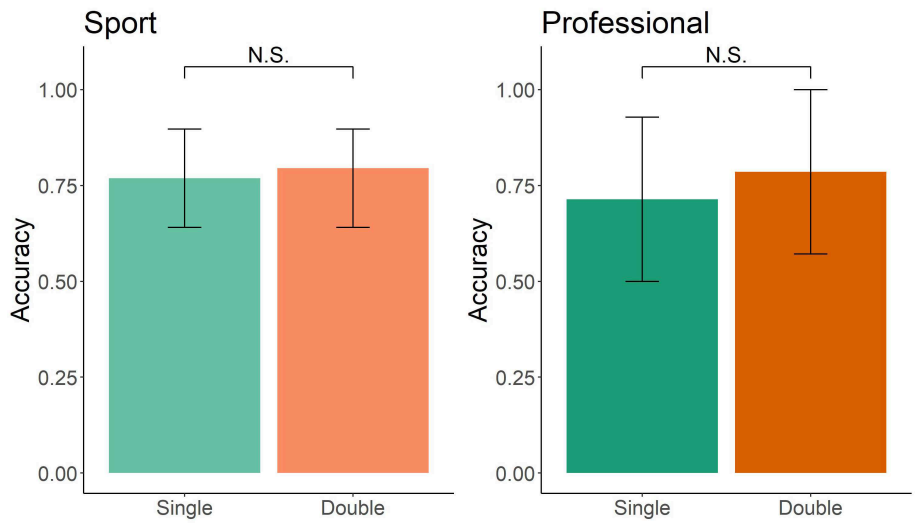


FIGURE 6 | Single-blind vs. double-blind search accuracy in sport and professional handler dog teams. Error bars represent 95% confidence intervals.

search dogs showed higher frequency of looking back compared to pets (42) and another study found that search and rescue, agility and pet dogs show differing patterns of looking back (45). Our blank area 3 for the Unknown group maybe a similar condition to the "unsolvable" task where a dog cannot find a target odor in the search, but the handler is still waiting for the dog to search. In contrast, handlers in the Known group may have been more apt to interpret a lack of alert suggesting that no target odor was present leading to fewer overall lookbacks. Interestingly, we did not see this effect in professional dogs, as few professional dogs engaged in many look backs. Given that the type and frequency of lookbacks to an owner or handler is related to ontogenetic experiences (43, 44, 46), training style differences between sport and professional handlers maybe related to the behavioral differences. However, this was not directly tested, but the present results suggest using blank area searches may be an interesting paradigm to explore dog-handler communication.

Lastly, we did not observe any differences between single-blind and double-blind testing. This suggests that an impartial judge or moderator of the trial may be present without directly influencing performance. Importantly, however, it is critical to note that the judge in the present experiment was a trained researcher familiar with phenomena associated with unintentional cuing. This was done to evaluate whether single-blind testing could be implemented impartially, which is important given that Pfungst (33) himself had trouble limiting unintentional cues given to Clever Hans. Under these conditions, we did not see a bias from the experimenter, but nonetheless, a less impartial judge, or a judge with strong motivations for the canine's performance may still unintentionally provide cues. Thus, it remains critical the impartiality of the judge of a detection dog trial remain under scrutiny and evaluation, but it nonetheless remains possible for an impartial judge to not provide cues to the team.

Across all analyses, we did not formally compare the performance of sport and professional teams given their significantly different backgrounds and variation in target odor volatility. Interestingly, although many comparisons did not quite reach the level of significance in our professional handlers, perhaps due to a smaller sample size, the direction of the effects all remained similar to the sport dogs. This suggests that perhaps sport canine teams may be a good model, where larger samples sizes can be reached quickly, to support research for professional dog teams.

Interestingly, we saw little effect of the years of experience training, reported frequency of double-blind training, or use of blanks in training on overall performance. All together, we only observed years of experience to reduce the duration of search in the blank area, with no other associations reaching the statistical criterion. To our knowledgeable, this was the first evaluation of how these different training methods (i.e., double-blind searchers or blank searches) influence a variety of performance measures. These results, however, may be limited due to sample size, handlers miss remembering reported training practices, or a reporting bias for procedures considered to be optimal. This suggests more rigorous and prospective experimental tests of the effect of these training methods should be conducted to evaluate their effects on operational performance.

There are several important limitations to the present study. First, we did not confirm an increased expectation of a target odor for the Known compared to the Unknown group (Hypothesis 1), although their actual search behavior did reflect this. There are several potential reasons for this finding. First, handlers may have simply mis-remembered their expectation when filling in the survey after the fact. Second, handlers may not have been aware of their changes in expectation for a target odor due to distraction during the search, or perhaps participants anticipated the Experimenters may have been trying to deceive them. Third, perhaps they simply wanted to report that their expectations were not influenced by the knowledge of the search task. Given that we did see behavioral changes between groups suggests that this lack of finding was not critical to the overall results but does suggest that future studies may need to do a better job clarifying the task parameters to participants in a known condition.

Another important limitation is that dogs did miss the target odor in the first two search areas. This likely would influence expectation of handlers in the Known group for search area 3 and introduced noise to the experiment. Nonetheless, we did still see changes in search behavior in area 3, and when we limited our relevant analyses to only the participants that correctly found the target odors in the first two areas, the direction and trend of our results remained similar. This suggests this was unlikely to be a critical limitation. This does suggests, however, that future studies could provide more explicit direction to handlers (e.g., "one odor is present in this area, one is present in this area, and nothing is present here") and see how that influences the results. For the present study, we opted not to do this as we thought this may be too explicit and would make handlers suspicious of the task. Further, such explicit knowledge rarely, if ever, occurs in the field. However, given that our script (i.e., script 2), did not generate the change in handler expectancy we expected (hypothesis 1), this would be a useful follow-up experiment.

Another important limitation was the relatively smaller sample size of professional handlers compared to sport handlers. This limited the power of some analyses and covariate analysis, but the direction of effects remained congruent with the sport groups. Future studies with increased power with professional groups and sport groups would be important to extend and replicate the present results.

In conclusion, the present results indicate that knowledge of the number of target odors present did lead to changes in behavior of the search team in a blank area. Teams searched the blank area longer, with the dog engaging in more lookbacks to the handler, when they did not know the number of target odors compared to when they did. Overall, however, we did not see handler knowledge about the presence of a blank area (no odor present) to change false alert rates compared to handlers that knew about the blank area. Lastly, we did not see any differences in performance in a single-blind and a double-blind search when an independent experimenter served the role as the trial moderator. Together, these results suggest that handler knowledge of test parameters influences team search behavior but did not lead to changes in false alert rates in a similar manner to previous work. More research is required, however, with varying levels of explicit handler knowledge on search parameters to evaluate its effect on the behavior of the team. Finally, we suggest that sport canine teams may be a good experimental model to evaluate these effects for professional handler teams.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board at TTU approved this study (IRB2019-501). The patients/participants provided their written informed consent to participate in this study. Ethical review and approval was not required for the animal study because Institutional animal care and use committee reviewed the protocol and it was determined to be an observational protocol. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

MD and NH contributed to the experiment development and analysis. CF contributed to recruiting participants and conducting the experiment. All authors were involved in writing the manuscript.

FUNDING

Funding was from internal Texas Tech University departmental source.

ACKNOWLEDGMENTS

We thank the handlers for their participation in this study and the trainers for their assistance on the project. We also thank the undergraduate students for coding dog behavior.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.00250/full#supplementary-material

REFERENCES

1. Furton K. The scientific foundation and efficacy of the use of canines as chemical detectors for explosives. *Talanta*. (2001) 54:487–500. doi: 10.1016/S0039-9140(00)00546-4
2. Gazit I, Terkel J. Explosives detection by sniffer dogs following strenuous physical activity. *Appl Anim Behav Sci*. (2003) 81:149–61. doi: 10.1016/S0168-1591(02)00274-5
3. Lazarowski L, Dorman DC. Explosives detection by military working dogs: olfactory generalization from components to mixtures. *Appl Anim Behav Sci*. (2014) 151:84–93. doi: 10.1016/j.applanim.2013.11.010
4. Marks A. Drug detection dogs and the growth of olfactory surveillance: beyond the rule of law? *Surveill Soc*. (2007) 4:868–73. doi: 10.24908/ss.v4i3.3450
5. Riva J, Marelli SP, Redaelli V, Bondiolotti GP, Sforzini E, Santoro MM, et al. The effects of drug detection training on behavioral reactivity and blood neurotransmitter levels in drug detection dogs: a preliminary study. *J Vet Behav*. (2012) 7:11–20. doi: 10.1016/j.jveb.2011.04.002
6. van Dam A, Schoon A, Wierda SF, Heeringa E, Aaldersa MCG. The use of crime scene detection dogs to locate semen stains on different types of fabric. *Forensic Sci Int*. (2019) 302:109907. doi: 10.1016/j.forsciint.2019.109907
7. Alexander MB, Friend T, Haug L. Obedience training effects on search dog performance. *Appl Anim Behav Sci*. (2011) 132:152–9. doi: 10.1016/j.applanim.2011.04.008
8. Gordon LE. Injuries and illnesses among urban search-and-rescue dogs deployed to haiti following the January 12, 2010, earthquake. *J Am Vet Med Assoc*. (2012) 240:396–403. doi: 10.2460/javma.240.4.396
9. Gordon LE. Injuries and illnesses among federal emergency management agency- certified search-and-recovery and search-and-rescue dogs deployed to oso, washington, following the march 22, 2014, state route 530 landslide. *J Am Vet Med Assoc*. (2015) 247:901–8. doi: 10.2460/javma.247.8.901

10. Greatbatch I, Gosling RJ, Allen S. Quantifying search dog effectiveness in a terrestrial search and rescue environment. *Wilderness Environ Med*. (2015) 26:327–34. doi: 10.1016/j.wem.2015.02.009
11. Slensky KA, Drobatz KJ, Downend AB, Otto CM. Deployment morbidity among search-and-rescue dogs used after the September 11, 2001, terrorist attacks. *J Am Vet Med Assoc*. (2004) 225:868–73. doi: 10.2460/javma.2004.225.868
12. Cornu J-N, Cancel-Tassin G, Ondet V, Girardet C, Cussenot O. Olfactory detection of prostate cancer by dogs sniffing urine: a step forward in early diagnosis. *Eur Urol*. (2011) 59:197–201. doi: 10.1016/j.eururo.2010.10.006
13. Pickel D, Manucy GP, Walker DB, Hall SB, Walker JC. Evidence for canine olfactory detection of melanoma. *App Anim Behav Sci*. (2004) 89:107–16. doi: 10.1016/j.applanim.2004.04.008
14. Walczak M, Jezierski T, Górecka-Bruzda A, Sobczynska M, Ensminger J. Impact of individual training parameters and manner of taking breath odor samples on the reliability of canines as cancer screeners. *J Vet Behav*. (2012) 7:283–94. doi: 10.1016/j.jveb.2012.01.001
15. Cablk ME, Heaton JS. Accuracy and reliability of dogs in surveying for desert tortoise (*Gopherus Agassizii*). *Ecol Appl*. (2006) 16:1926–35. doi: 10.1890/1051-0761(2006)016[1926:AARODI]2.0.CO;2
16. Cristescu RH, Foley E, Markula A, Jackson G, Jones D, Frère C. Accuracy and efficiency of detection dogs: a powerful new tool for koala conservation and management. *Sci Rep*. (2015) 5:8349. doi: 10.1038/srep08349
17. Dematteo KE, Rinas MA, Sede MM, Davenport B, Argüelles CF, Lovett K, et al. Detection dogs: an effective technique for bush dog surveys. *J Wildl Manag*. (2009) 73:1436–40. doi: 10.2193/2008-545
18. Reed SE, Bidlack AL, Hurt A, Getz WM. Detection distance and environmental factors in conservation detection dog surveys. *J Wildl Manage*. (2011) 75:243–51. doi: 10.1002/jwmg.8
19. Smith DA, Ralls K, Cypher BL, Maldonado JE. Assessment of scat-detection dog surveys to determine kit fox distribution. *Wildl Soc Bull*. (2005) 33:897–904. doi: 10.2193/0091-7648(2005)33[897:AOSDST]2.0.CO;2
20. Steen JB, Mohus I, Kvesetberg T, Walloe L. Olfaction in bird dogs during hunting. *Acta Physiol Scand*. (1996) 157:115–9. doi: 10.1046/j.1365-201X.1996.479227000.x
21. Zubedat S, Aga-Mizrachi S, Cymerblit-Sabba A, Shwartz J, Leon JF, Rozen S, et al. Human–animal interface: the effects of handler's stress on the performance of canines in an explosive detection task. *Appl Anim Behav Sci*. (2014) 158:69–75. doi: 10.1016/j.applanim.2014.05.004
22. Cooper R, Wang C, Singh N. Accuracy of trained canines for detecting bed bugs (*Hemiptera: Cimicidae*). *J Econ Entomol*. (2014) 107:2171–81. doi: 10.1603/EC14195
23. Moser E, McCulloch M. Canine scent detection of human cancers: a review of methods and accuracy. *J Vet Behav*. (2010) 5:145–52. doi: 10.1016/j.jveb.2010.01.002
24. Gordon RT, Schatz CB, Myers LJ, Kosty M, Gonczy C, Kroener J, et al. The use of canines in the detection of human cancers. *J Alternat Complement Med*. (2008) 14:61–7. doi: 10.1089/acm.2006.6408
25. McCulloch M, Jezierski T, Broffman M, Hubbard A, Turner K, Janecki T. Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. *Integr Cancer Ther*. (2006) 5:30–9. doi: 10.1177/1534735405285096
26. Willis CM, Church SM, Guest CM, Cook WA, McCarthy N, Bransbury AJ, et al. Olfactory detection of human bladder cancer by dogs: Proof of principle study. *BMJ*. (2004) 329:7126. doi: 10.1136/bmj.329.7468.712
27. Horn L, Range F, Huber L. Dogs' attention towards humans depends on their relationship, not only on social familiarity. *Anim Cogn*. (2013) 16:435–43. doi: 10.1007/s10071-012-0584-9
28. Jamieson L, Baxter G, Murray P. You Are Not My Handler! impact of changing handlers on dogs' behaviours and detection performance. *Animals*. (2018) 8:176. doi: 10.3390/ani8100176
29. Duranton C, Horowitz A. Let me sniff! nosework induces positive judgment bias in pet dogs. *App Ani Behav Sci*. (2019) 211:61–6. doi: 10.1016/j.applanim.2018.12.009
30. Gazit I, Goldblatt A, Terkel J. The role of context specificity in learning: the effects of training context on explosives detection in dogs. *Anim Cogn*. (2005) 8:143–50. doi: 10.1007/s10071-004-0236-9
31. Lit L, Schweitzer JB, Oberbauer AM. Handler beliefs affect scent detection dog outcomes. *Anim Cogn*. (2011) 14:387–94. doi: 10.1007/s10071-010-0373-2
32. Waggoner LP, Jones MH, Williams M, Johnston JM, Edge CC, Petrousky JA. Effects of extraneous odors on canine detection. In: DePersia AT, Pennella JJ, editors. *Proceedings SPIE 3575, Enforcement and Security Technologies*. Boston, MA. (1998). p. 355–62. doi: 10.1117/12.335008
33. Pfungst O. A contribution to experimental animal and human psychology. In: *Clever Hans (the Horse of Mr. Von Osten)*. New York, NY: Henry Holt and Company. (1911).
34. Wasser SK, Davenport B, Ramage ER, Hunt KE, Parker M, Clarke C, et al. Scat detection dogs in wildlife research and management: application to grizzly and black bears in the yellowhead ecosystem, alberta, Canada. *Can J Zool*. (2004) 82:475–92. doi: 10.1139/z04-020
35. Soproni K, Miklósi A, Topál J, Csányi V. Comprehension of human communicative signs in pet dogs (*Canis Familiaris*). *J Comp Psychol*. (2001) 115:122–6. doi: 10.1037/0735-7036.115.2.122
36. Virányi Z, Topál J, Gácsi M, Miklósi Á, Csányi V. Dogs respond appropriately to cues of humans' attentional focus. *Behav. Processes*. (2004) 66:161–72. doi: 10.1016/j.beproc.2004.01.012
37. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing (2020). Available online at: https://www.R-project.org/
38. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. (2015) 67:1–48. doi: 10.18637/jss.v067.i01
39. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. *J Stat Softw*. (2017) 82:1–26. doi: 10.18637/jss.v082.i13
40. Revelle W. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, IL: Northwestern University (2019). Available online at: https://CRAN.R-project.org/package=psych
41. Concha A, Mills DS, Feugier A, Zulch H, Guest C, Harris R, et al. (2014). Using sniffing behavior to differentiate true negative from false negative responses in trained scent detection dogs. *Chem senses*. 39:749–54. doi: 10.1093/chemse/bju045
42. D'Aniello B, Scandurra A, Prato-Previde E, Valsecchi P. Gazing toward humans: a study on water rescue dogs using the impossible task paradigm. *Behav Processes*. (2015) 110:68–73. doi: 10.1016/j.beproc.2014.09.022
43. D'Aniello B, Scandurra A. Ontogenetic effects on gazing behaviour: a case study of kennel dogs (labrador retrievers) in the impossible task paradigm. *Anim Cogn*. (2016) 19:565–70. doi: 10.1007/s10071-016-0958-5
44. Lazzaroni M, Marshall-Pescini S, Manzenreiter H, Gosch S, Pribilová L, Darc L, et al. Why do dogs look back at the human in an impossible task? Looking back behaviour may be over-interpreted. *Ani Cogn*. (2020) 23:427–41. doi: 10.1007/s10071-020-01345-8
45. Marshall-Pescini S, Passalacqua C, Barnard S, Valsecchi P, Prato-Previde E. Agility and search and rescue training differently affects pet dogs' behaviour in socio-cognitive tasks. *Behav Processes*. (2009) 81:416–22. doi: 10.1016/j.beproc.2009.03.015
46. Marshall-Pescini S, Rao A, Virányi Z, Range F. The role of domestication and experience in 'looking back' towards humans in an unsolvable task. *Sci Rep*. (2017) 7:1–7. doi: 10.1038/srep46636
47. Miklósi Á, Kubinyi E, Topál J, Gácsi M, Virányi Z, Csányi V. A simple reason for a big difference: wolves do not look back at humans, but dogs do. *Curr Biol*. (2003) 13:763–6. doi: 10.1016/S0960-9822(03)00263-X
48. Udell MAR. When dogs look back: inhibition of independent problem-solving behaviour in domestic dogs (*Canis Lupus* familiaris) compared with wolves (*Canis Lupus*). *Biol Lett*. (2015) 11:20150489. doi: 10.1098/rsbl.2015.0489

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 03 June 2020
doi: 10.3389/fvets.2020.00292

A Randomized Cross-Over Field Study of Pre-Hydration Strategies in Dogs Tracking in Hot Environments

Greta M. Niedermeyer[1,2], Elizabeth Hare[2,3], Leslie K. Brunker[2], Richard A. Berk[4], Kathleen M. Kelsey[2], Tracy A. Darling[1,2], Jess L. Nord[1,2], Kasey K. Schmidt[5] and Cynthia M. Otto[1,2]*

[1] Department of Clinical Sciences and Advanced Medicine, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, United States, [2] Penn Vet Working Dog Center, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, United States, [3] Dog Genetics, LLC, Astoria, NY, United States, [4] Department of Criminology and Statistics, University of Pennsylvania, Philadelphia, PA, United States, [5] College of Veterinary Medicine, Washington State University, Pullman, WA, United States

OPEN ACCESS

Edited by:
Mary M. Christopher,
University of California, Davis,
United States

Reviewed by:
Joseph Wakshlag,
Cornell University, United States
Mitsuaki Ohta,
Tokyo University of Agriculture, Japan
Yaron Bruchim,
The Hebrew University of Jerusalem, Israel

***Correspondence:**
Cynthia M. Otto
cmotto@vet.upenn.edu

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences, a section of the journal Frontiers in Veterinary Science

Received: 26 November 2019
Accepted: 29 April 2020
Published: 03 June 2020

Citation:
Niedermeyer GM, Hare E, Brunker LK, Berk RA, Kelsey KM, Darling TA, Nord JL, Schmidt KK and Otto CM (2020) A Randomized Cross-Over Field Study of Pre-Hydration Strategies in Dogs Tracking in Hot Environments. Front. Vet. Sci. 7:292. doi: 10.3389/fvets.2020.00292

The objective of this study was to evaluate 4 pre-exercise hydration strategies (oral water, chicken-flavored water, chicken-flavored oral electrolyte solution, and subcutaneous electrolyte solution) in working dogs conducting rigorous tracking operations in hot and arid conditions. In a randomized cross-over field study, 7 Border Patrol Search, Trauma, and Rescue (BORSTAR) Unit dogs working/training out of Fort Bliss in El Paso, Texas were randomly assigned to one of 4 different hydration strategy treatments each day for 4 days of study participation. Dogs were provided hydration treatment prior to running 2 separate one-mile tracks and were offered water while tracking. Body weight, blood, and urine were collected at the beginning of the study day and at the completion of each track. Core body temperatures were recorded using internal temperature sensing capsules. The impact of hydration strategy on change in weight, peak temperature, and serum chemical, hematological, and urinary parameters were analyzed using the COIN procedure in R[a]. Compared to the other 3 hydration strategies, dogs receiving chicken-flavored water had higher blood creatine kinase values at the end of the second track ($p = 0.0361$). Otherwise, hydration strategy had minimal effects on blood or urine parameters. Total fluid intake was lower with water only compared to the other three hydration strategies. Dogs developed elevated core body temperatures (median 41°C; 106°F) without signs of heat exhaustion or heat stroke. Alternate hydration strategies increased total fluid intake compared to water alone; however, chicken-flavored water resulted in increased markers of muscle injury suggesting electrolyte-enriched strategies may have an advantage as a hydration strategy. Additionally, electrolyte-enriched fluids before exercise may help these dogs maintain lower peak temperatures.

Keywords: sports medicine, electrolytes, thermoregulation, working dogs, field study

INTRODUCTION

Military and other working dogs are critical for U.S. security and aid in border control as well as natural disaster response. Exercise-induced hyperthermia limits the ability of dogs to perform physically (1) and is one of the few preventable causes of death or euthanasia in MWD (2, 3). In a study analyzing reasons for discharge in MWD, heat stroke was the most common non-behavioral reason for dogs <5 years old to be discharged (4). Additionally, heat stress followed gunshot wounds and explosion/blast wounds as the third most common cause of death in MWD deployed in Iraq and Afghanistan (5).

The ability to thermoregulate and avert heat stress and its progression to heat stroke are influenced by work, ambient conditions, acclimatization, and hydration (3, 6–8). Since working dogs are often required to perform physically challenging tasks in adverse environmental conditions (e.g., high temperatures), efforts to mitigate the impact through improved acclimatization and hydration are beneficial.

Acclimatization or adaptation to environmental conditions allows dogs to better withstand hyperthermia, whether environmental or exercise-induced (8). A heat-acclimated animal (adapted under artificial conditions) is better able to tolerate longer exposures to heat as well as more extreme heat (9). Acclimation occurs as the animal adapts to better dissipate heat, produce less heat, and under certain conditions expand the safe range of core body temperatures (9). A heat-acclimated state results in lower core body temperature, reduced heart rate, elevated cardiovascular reserve, and increased evaporative cooling (9). Bruchim demonstrated that together with physical training, heat acclimatization resulted in a decreased rise in rectal temperature and heart rate in MWD following a physical performance test despite increased test intensity (10). While Bruchim showed dramatic changes, these results were measured after ∼6 and 18 months of acclimatization (10). Partial acclimatization is thought to occur within 10–20 days, but full acclimatization can require up to 2 months (11).

Acclimatization is not always an achievable goal since working dogs are sometimes called to different climates with no time for acclimatization [e.g., When search and rescue (SAR) dogs from cooler climates were sent to Haiti following the 2010 earthquake (12)]. Additionally, hypohydration decreases heat tolerance regardless of acclimatization (13).

Hydration management provides a promising approach to reduce the risk of heat stress and heat stroke both in an acute setting when acclimatization is not possible and in concert with acclimatization. Yet, hydration management can pose a challenge in working dogs during demanding situations; dehydration was the most common health issue in dogs that responded to the earthquake in Haiti and was reported by handlers of SAR dogs that responded to 9/11 (12, 14).

Despite the importance of working dogs and the significance of heat stroke and dehydration, few studies have compared the safety and efficacy of various hydration strategies used in the field (15, 16). In a previous study comparing oral water, an oral electrolyte solution, and subcutaneous fluids, hydration strategy had only minor effects on physiological parameters and no detectable effect on behavioral parameters in vehicle-screening dogs working at the Sarita, Texas checkpoint, although dogs did increase their fluid consumption and hydration when offered a chicken-flavored oral electrolyte solution (16). It is unknown whether these same trends would hold up in more extreme conditions (i.e., high heat, no shade, and rigorous activity). To our knowledge, there are no studies evaluating hydration strategies in working dogs under these conditions or comparing the effects of a flavored oral electrolyte solution to flavored water.

The purpose of this study was to investigate the impact of four pre-exercise hydration strategies—oral water (W), chicken flavored water (CHK), chicken-flavored oral electrolyte solution (OES), and subcutaneous electrolyte solution (SCE)—on dogs tracking at the border in El Paso, Texas as part of the Border Patrol Search, Trauma, and Rescue (BORSTAR) unit. We hypothesized that hydration method would neither affect clinical parameters, including core temperature, in the dogs nor be associated with any adverse effects.

Abbreviations: CHK, chicken-flavored water; MWD, Military Working Dog; OES, chicken-flavored oral electrolyte solution; SAR, search and rescue; SCE, subcutaneous electrolyte solution; W, oral water.

MATERIALS AND METHODS

Ethics Statements

The protocols used in this study were reviewed and approved by both the University of Pennsylvania and US Army Medical Research and Materiel Command Institutional Animal Care and Use Committees (USAMRCM #SO120002, UPenn IACUC protocol #804293).

Animals

Seven Border Patrol dogs working for the Customs and Border Protection BORSTAR Unit based out of Fort Bliss in El Paso, Texas were invited based on canine and handler availability during the experimental period. All dogs were trained to track and trail humans traversing the open desert. All dogs worked, trained, and were kenneled at the same facility, and all lived with their handlers. All dogs were deemed healthy and in good condition based on physical examination by a veterinarian (CMO) prior to starting the study. The age, breed, sex, neuter status, body condition, physical examination parameters, and medical and work history were collected from the handlers on all dogs. Exclusion criteria included any canine illness or injury, request of the handler not to participate, or any adverse events.

Experimental Protocol

In this cross-over design, dogs were randomly assigned to each one of 4 treatment protocols over the 4 days of study participation. Study days were limited to alternating days to allow handlers time to attend mandatory training classes on base: Tuesday and Thursday during week 1, Monday and Wednesday during week 2. In a remote wilderness area, one-mile tracks were laid down by handlers or trainers ∼30 min prior to the start of each dog's tracking session (each dog followed a fresh track). (**Figure 1**) Tracking order was randomly assigned and

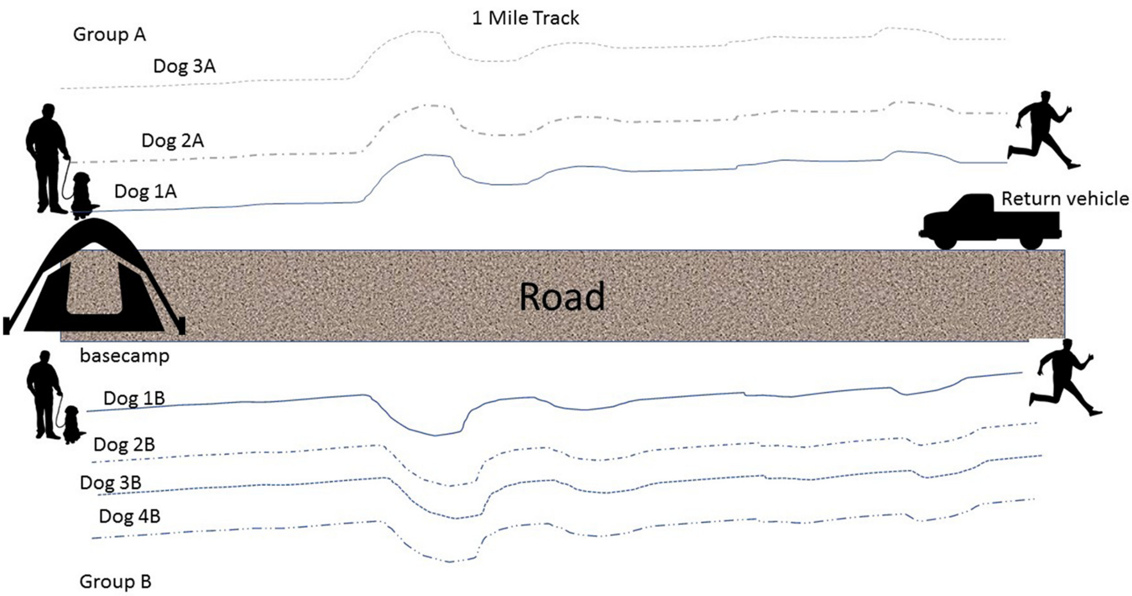


FIGURE 1 | An illustration of the tracks that were laid for each dog during each of the tracking sessions (Track 1 and Track 2).

each dog completed a one-mile track (Track 1), rested, and then completed a second one-mile track (Track 2). Handlers took breaks and doused (i.e., wet down for cooling) their dogs with water during Track 1 and Track 2 as they would normally during a working session.

The treatment protocols were assigned randomly, and each pair of dogs had their access to water restricted either 1 h (first 2 dogs) or 2 tracks prior to their turn and received their assigned hydration strategy during the track run preceding theirs or 30 min prior to tracking (first 2 dogs) (**Figure 2**). The dogs were given either W, OES (Hydrolyte, Advanced Nutritional Support, Elka Park, NY, USA.), CHK (Chicken flavoring used in Hydrolyte, Advanced Nutritional Support, Elka Park, NY, USA.), or SCE (Plasmalyte A, Abbott Laboratories, North Chicago, IL, USA). The volume of assigned fluid given was 15 mL/kg for SCE and 10 mL/kg for W, OES, and CHK (See **Table 1** for the composition of each fluid). After administration of the assigned fluid, water was restricted prior to launch. Each handler carried a water (H_2O) source on the track that was measured before and after the track was finished. Handlers offered water to their dogs as they would normally while working. During tracking, the water was poured into a portable bowl and the handlers were instructed to hold the bowl and minimize any loss. Residual water was replaced into a wide mouth water bottle for measurement in a graduated cylinder at the end of the track. Water consumption was quantified during each dog's rest period until the second round of tracks began wherein the same schedule was implemented: removing access to water two tracks prior to launch and offering the same treatment protocol one track prior to launch. Dogs on the SCE protocol only received subcutaneous fluids before their first track and were offered oral water by weight (10 mL/kg) one track prior to launch of their second track.

Data Collected

Dog Fluid and Food Intake

The fluid volumes offered and consumed were documented at the beginning of each track and rest cycle. The total fluid volume was recorded for each dog for each study day: fluids administered as well as water consumed for the SCE treatment, water intake for the W treatment, water plus OES for the OES treatment, and water plus chicken-flavored water for the CHK treatment. The fluids were measured using a graduated cylinder and offered in a bowl under supervision to limit any loss from spillage. Any remaining fluid was then again measured with a graduated cylinder to determine how much was consumed. Any food consumed by the dogs during the work day was recorded. Dogs were maintained on their normal feeding schedule, except for a small amount of canned dog food associated with the ingestion of the internal core temperature sensing capsules (CorTemp system, HQInc Wireless Sensing Systems & Design; Palmetto, FL, USA.).

Dog Physiologic Parameters

A physical examination was performed on each dog at the beginning of each day. Temperature, pulse or heart rate, and respiratory rate were obtained at the end of each track. If any dog exhibited signs of physical distress and/or was unable to maintain adequate hydration as evidenced by weakness, persistent tachycardia, poor pulse quality, and prolonged capillary refill time, it was to be removed from the study and treated appropriately.

Dog Activity

Each dog was assigned its own activity monitor that used omni-directional accelerometers (version 3.1, Actical®, Respironics, Koninklijke Philips Electronics, Bend, OR, USA) to collect quantitative activity data. This monitor has been validated in

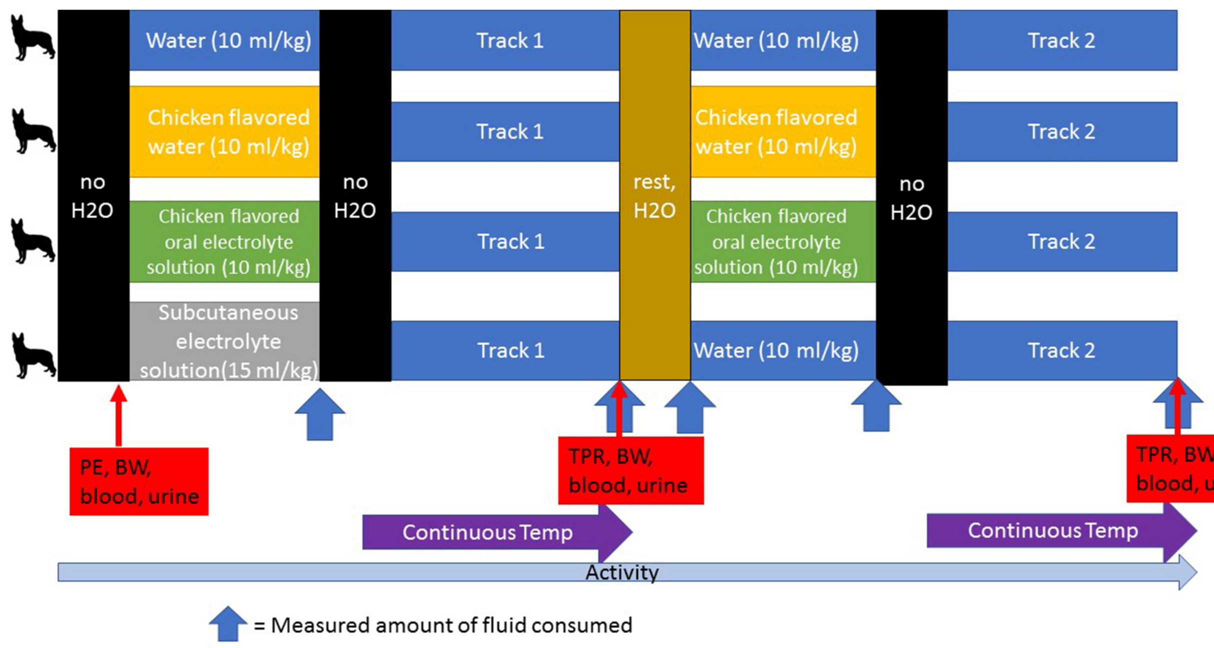


FIGURE 2 | A schematic of the study design. Water was restricted at designated times (no H_2O). An initial physical examination (PE), body weight measurement (BW) and blood and urine collection were performed at the start of each day. Dogs were randomly assigned to the treatment groups for each day and in this cross-over design, all dogs participated in each treatment group. Fluid consumption was measured after each dog was given its assigned treatment, after each track during which the handler kept track of the amount of water consumed, and during the rest period. Temperature was monitored continuously during the tracking exercise. Temperature (T), pulse (P), respiration (R), and body weight (BW) were measured and blood and urine were collected after each track. Activity was monitored continuously during each study day.

TABLE 1 | Measured and reported electrolyte composition of OES, SCE, and CHK.

Ingredient	OES measured	SCE reported	CHK measured
Sodium (mmol/L)	87	140	unmeasurable
Potassium (mmol/L)	7.4	5	2.7
Chloride (mmol/L)	67	98	unmeasurable
Buffer	Bicarbonate 8 mmol/L	Acetate 27 mEq/L Gluconate 23 mEq/L	Bicarbonate < 5 mmol/L
Magnesium (mmol/L)	4.4	1.5	1.4
Glucose (mmol/L)	18	0	<0.6
Osmolality (mOsm/L)	206*	294	unmeasurable
Effective strong ion difference (mEq/L)†	27.4	47	unmeasurable

**Osmolality was calculated as 2([Na$^+$] + [K$^+$]) + (glucose), where brackets represent concentration.*
†Effective strong ion difference = [Na$^+$] + [K$^+$] – [Cl$^-$].

dogs (17) and used in a several prior exercise studies (16, 18). The activity monitors were secured to standard flat buckle collars and placed on all dogs upon arrival onsite each morning. The monitors were worn throughout the day and removed following the final track of the day.

Environmental Parameters

Ambient temperature and percent humidity were measured every 15 min with a wireless weather station.

Core Body Temperature

Core body temperature (gastrointestinal) was monitored using internal temperature sensing capsules. Capsules were administered in a small amount of food and, as long as the original capsule was present and transmitting appropriately, a new capsule was not administered until the old capsule passed. Temperature-recording monitors for continuous temperature readings were attached to each dog's harness just prior to launch and removed after the track ended. Intermittent readings were taken following each track as necessary.

Body Weight, Serum Chemistry, Hematology, and Urine Measurements

The following samples were collected at the beginning of the study day and at the end of both tracking sessions: body weight, blood, and urine. Weight in kilograms was obtained using a walk-on electronic scale (Jorvet J0825PM, JorVet Walk on Scale 36"; Jorgensen Labs, Loveland, Colorado) that was calibrated twice daily. Peripheral venous blood samples from the saphenous or cephalic veins (3 mL) were anticoagulated with

Li heparin for use in a point of care blood analyzer (Abaxis veterinary research laboratories, Union City, CA; CG8+ ISTAT cartridge, Abaxis, Union City, California) to measure pH, partial pressure of carbon dioxide (pCO_2), sodium (Na), potassium (K), ionized calcium (iCa), glucose (Glu), hematocrit (Hct), and bicarbonate (HCO_3). Lactate blood levels were analyzed using a handheld lactate meter (Lactate Scout, EKF Diagnostics, Penarth Cardiff). All remaining blood was centrifuged at 3,150 rpm for 5 min; the plasma was drawn off in two samples, placed in cryotubes and frozen on dry ice for future analysis. Blood urea nitrogen (BUN), creatinine (Creat), creatine kinase (CK), and chloride (Cl) were measured from blood samples collected at baseline (morning exam) and after completion of Track 2 at a veterinary clinical laboratory (Clinical Laboratory, M. J. Ryan Veterinary Hospital, University of Pennsylvania, Philadelphia PA) after study completion. Urine specific gravity was measured onsite with a handheld refractometer on free catch, midstream urine samples and the remaining sample was refrigerated for future analysis. Urinary sodium and urinary creatinine were measured from baseline urine collection and urine collected after completion of Track 2 samples at a veterinary clinical laboratory (Clinical Laboratory, M. J. Ryan Veterinary Hospital, University of Pennsylvania, Philadelphia PA) after study completion to calculate the fractional excretion of sodium (FeNa) at those two timepoints. Fractional excretion of sodium was calculated as described by Hinchcliff et al. (19) and used in our previous study (16).

$FeNa = (UNa^*Screat)/(SNa^*Ucreat)\ ^*100$

Where UNa and SNa are the concentrations of sodium in urine and serum, respectively, and Screat and Ucreat are the concentrations of creatinine in serum and urine, respectively.

Statistical Analyses

This study was conducted with a randomized crossover design, and the 4 interventions were assigned at random, once to each of 7 dogs; the random assignment by itself justified the permutation tests we employed, assuming "no interference" (20). For this study, no interference means that the intervention assigned at random to one dog had no effect on any other dog.

To address the within dog dependence, treatment sums of squares were computed within dogs, and the residual sum of squares had the between dog sum of squares removed. For a distribution free approach, we applied permutation ANOVA tests approximated by Monte Carlo methods (21). The very large number possible permutations effectively precluded exact tests (10,000 permutations were used for each test).

For each response variable, we evaluated all possible contrasts between our 4 interventions, searching for the greatest contrast. Beyond usual concerns about statistical tests based on multiple comparisons, "cherry picking" the greatest contrasts required statistical adjustments for post-selection statistical inference. We applied a special case of generalized maximally selected statistics that when coupled with permutation tests, provided proper statistical inference. The key requirement was that all possible contrasts were computed (22). To this end, we used the procedure COIN (23) available in the programming language R (R: a Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.) to analyze the effect of hydration strategy on the following parameters: total fluid intake (sum of consumption before Track 1, during Track 1, after Track 1, before Track 2, during Track 2, and after Track 2), peak core body temperature reached over the study day, and the difference between parameters after Track 2 compared to baseline — weight, Na, Cl, K, iCa, pH, HCO_3, pCO_2, Glu, lactate, BUN, Creat, CK, Hct, USG, and FeNa.

Descriptive data were reported as median and range. To determine if there was a significant difference in absolute (rather than change) blood or urine parameters between baseline and Track 2, a two-way repeated measures ANOVA or Wilcoxon Signed Rank Test was run on values for the following parameters: weight, peak core body temperature, BUN, CK, FeNa, and USG.

To address whether the environmental conditions affected peak body temperature, we plotted the peak core body temperature measured for Track 1 and Track 2 against both ambient humidity and ambient temperature and analyzed the regression coefficient.

RESULTS

Four intact males and 3 spayed females participated in this study. Two of the females were Belgian Malinois, 2 of the males were German Shepherds, one female and one male were Labrador Retrievers, and one male was a Belgian Malinois-German Shepherd cross. All dogs were between 2 and 7 years old: median age 6 years. The dogs' diets were one of 2 commercial dry diets (Science Diet Active Adult or Science Diet Advanced Fitness, Hill's Pet Nutrition, Topeka, KS, USA). All dogs were assessed as a 5 or 5.5 on the 1–9 body condition scoring system (24). All dogs except one who arrived from Harlingen, Texas were located in El Paso, Texas prior to the start of the study. As a specialized unit, the BORSTAR dogs do not work on a routine schedule. To account for differences in acclimation to activity, we documented the last time the dog had been deployed to track or perform any other duty related work. At the start of the study it had been a mean of 3.5 ± 2.3 weeks since the dogs last worked. All dogs appeared normal on physical examination at the start of each study day, and no dogs demonstrated signs of physical distress or illness that prevented them from participating for the full study day. Dogs received no food while working each study day except for that associated with placement of the internal temperature sensing capsules. Throughout the entire duration of the study, handlers reported 1 case of diarrhea (W), 2 cases of sore paws (two different dogs on OES), 3 episodes of decreased appetite all in the same dog (W, OES, SCE), and 1 case of decreased urination (CHK).

Total activity counts for each dog for each track were divided by the minutes spent tracking, for Track 1 activity counts ranged from 1,390–4,198 (median 2,386) counts per minute. For Track 2, activity ranged from 1,217–4,641 (median 2,373) counts per minute. Average track time for Track 1 was 24.3 min and ranged from 12–38 min. Average time for Track 2 was 25.2 min and ranged from 8–72 min.

Median baseline core body temperature across all study days was within normal limits but Track 1 and Track 2 peak temperatures were significantly elevated from baseline (**Table 2**) ($p < 0.05$). Peak core body temperatures while tracking ranged from 39.95 to−42.88°C (103.91–109.18°F). During the study days, ambient temperature and humidity ranged from 24.7–30.6°C (76.5–87.0°F) and 37.0–63.3%, respectively for Track 1. For Track 2, ambient temperature and humidity ranged from 28.3–37.1°C (83.0–98.8°F) and 23.3–60.0%; see **Table 2** for mean environmental parameters and baseline and treatment and track specific examination parameters. Accurate respiratory rates were not obtained due to the dogs' panting.

Effect of Hydration Strategy on Fluid Intake

Total fluid intake was influenced by hydration strategy ($p = 0.0208$) when controlling for individual dog. Contrast analysis revealed that fluid intake for the W treatment, which is considered the baseline fluid consumption, was significantly different from intake with the other three strategies. Mean total fluid intake by hydration protocol was 49.1 (41.1, 50.5) mL/kg for W, 59.7 (52.4, 74.4) mL/kg for CHK, 62.9 (56.2, 74.4) mL/kg for SCE, and 83.3 (56.8, 85.6) mL/kg for OES. All dogs drank the full pre-tracking fluid volume for both Track 1 and Track 2 when they were offered OES. When offered CHK, all dogs drank the full pre-hydration amount for Track 2, and all but one dog drank the full pre-hydration amount for Track 1. Additional Na load for OES was 1.74 mEq/kg and 2.1 mEq/kg for SCE. No measurable amounts of Na were provided by W or CHK.

Influence of Hydration Strategy on Peak Temperature and Weight Loss

The dogs' median peak temperature from both tracks and each treatment group are reported in **Table 2**. Although we were unable to detect an overall effect of hydration strategy on peak temperature ($p = 0.1316$), when we compared the means of electrolyte enriched (i.e., OES and SCE) vs. electrolyte free (i.e., CHK and W) solutions to address whether the presence of electrolytes influenced peak core body temperature, the p value was not significant at 0.07. When controlling for dog and hydration strategy, there was no significant difference in the dogs' peak temperature for Track 1 vs. Track 2. We were not able to detect an effect of hydration strategy on the change in weight when comparing baseline values to values post Track 2.

Effect of Hydration Strategy on Blood and Urine Parameters

Hydration strategy had a significant impact on the change in CK ($p = 0.0361$) with a significant contrast between CHK and other hydration strategies (**Table 3**) ($p = 0.0361$). We were not able to detect an effect of hydration strategy on change in Na, Cl, K, iCa, pH, pCO_2, Glu, lactate, BUN, Creat, Hct, USG, and FeNa. When we compared the average change in HCO_3 between the hydration strategies, the p value was not significant at $p = 0.0652$ with contrast between W and the other three hydration strategies. See **Table 3** for blood chemistry, hematology, and urinary parameters at baseline and after each tracking session for each hydration strategy.

TABLE 2 | Median core body temperature, pulse rate, and weight of all dogs over all study days as a function of time and treatment group.

	Baseline	Track1					Track2				
		Overall, N = 28	W, N = 7	OES, N = 7	CHK, N = 7	SCE, N = 7	Overall, N = 28	W, N = 7	OES, N = 7	CHK, N = 7	SCE, N = 7
Weight (kg)	31.8 (30.4–33.7)	32.1 (30.4, 33.4)	31.7 (30.3, 32.8)	32.0 (30.4, 33.2)	32.5 (30.6, 33.9)	32.6 (30.6, 33.4)	31.7 (30.0, 32.7)*	31.4 (29.6, 32.2)	32.0 (29.9, 32.6)	32.0 (30.6, 33.5)	30.6 (30.2, 33.1)
Pulse (bpm)	102 (96–120)	140 (120, 150)	130 (120, 140)	140 (118, 148)	150 (130, 175)	140 (119, 144)	140 (120, 150)	140 (122, 142)	142 (131, 154)	150 (145, 158)	120 (110, 142)
Peak core temperature (°C)	38.47 (38.12–38.83)	41.15 (40.77, 41.73)*	40.96 (40.61, 41.20)	41.22 (40.80, 41.67)	41.96 (41.37, 42.37)	40.79 (40.69, 41.23)	41.30 (40.87, 41.86)*	41.59 (41.02, 42.36)	41.16 (40.96, 41.82)	41.57 (40.77, 41.81)	41.30 (41.10, 41.44)
Peak core temperature (°F)	101.25 (100.62–101.90)	106.07 (105.39, 107.12)*	105.72 (105.10, 106.16)	106.20 (105.43, 107.00)	107.52 (106.47, 108.26)	105.42 (105.25, 106.22)	106.34 (105.56, 107.34)*	106.86 (105.83, 108.25)	106.08 (105.72, 107.28)	106.82 (105.38, 107.26)	106.34 (105.98, 106.59)
Track time (min)		24.0 (19.5, 29.0)	27.0 (18.5, 27.0)	22.0 (19.2, 29.2)	27.0 (21.0; 29.5)	21.0 (18.5; 29.2)	22.5 (13.5, 31.5)	20.0 (15.0, 27.8)	25.0 (19.5; 32.5)	25.0 (16.8, 41.0)	12.0 (12.0; 22.5)
Ambient temperature°C (°F)		28.0 ± 1.6 (82.5 ± 2.8°F)					32.4 ± 2.2 (90.3 ± 4.0°F)				
Humidity		50.2 ± 7.7 %					38.9 ± 8.8%				

*Core body temperature for tracks are presented as the median peak value reached within each tracking session for each treatment group. Pulse rate and body weight were measured at baseline and following completion of each track. Statistics, Median (IQR) Mean +/- SD (Ambient Temperature and Humidity); W, water; OES, chicken-flavored oral electrolyte solution; CHK, chicken-flavored water; SCE, subcutaneous electrolyte solution; *, significantly different from baseline.*

TABLE 3 | Blood chemistry, hematology, and urinary parameters of all dogs at baseline and after each tracking session.

Characteristic	**Track 1**					**Track 2**				
	Overall, $N = 28$	**W, $N = 7$[1]**	**OES, $N = 7$[1]**	**CHK, $N = 7$[1]**	**SCE, $N = 7$[1]**	**Overall, $N = 28$**	**W, $N = 7$[1]**	**OES, $N = 7$[1]**	**CHK, $N = 7$[1]**	**SCE, $N = 7$[1]**
Sodium (mmol/L) *RR 139–150* **B: 147 (145, 148)**	147 (146, 148)	146 (145, 148)	148 (146, 150)	146 (146, 147)	148 (148, 148)	144 (144, 146)	144 (144, 146)	145 (144, 148)	144 (144, 147)	144 (144, 146)
Potassium (mmol/L) *RR 3.4–4.9* **B: 4.1 (4.0, 4.3)**	3.85 (3.68, 4.10)	3.80 (3.75, 4.00)	3.80 (3.65, 3.85)	4.00 (3.75, 4.10)	4.10 (3.60, 4.15)	3.60 (3.48, 3.80)	3.70 (3.60, 3.95)	3.50 (3.40, 3.75)	3.50 (3.40, 3.60)	3.80 (3.65, 3.85)
Ionized Calcium (mmol/L) *RR 1.12–1.40* **B: 1.33 (1.32, 1.34)**	1.23 (1.21, 1.26)	1.27 (1.23, 1.27)	1.22 (1.17, 1.25)	1.22 (1.17, 1.25)	1.23 (1.21, 1.25)	1.19 (1.15, 1.22)	1.20 (1.16, 1.23)	1.16 (1.15, 1.21)	1.17 (1.13, 1.20)	1.21 (1.18, 1.24)
pH *RR 7.35–7.45* **B: 7.41 (7.38, 7.44)**	7.53 (7.51, 7.59)	7.52 (7.48, 7.56)	7.56 (7.52, 7.59)	7.54 (7.53, 7.61)	7.56 (7.47, 7.57)	7.63 (7.58, 7.70)	7.61 (7.53, 7.72)	7.63 (7.59, 7.69)	7.65 (7.59, 7.72)	7.61 (7.59, 7.65)
Bicarbonate (mmol/L) *RR 15–23* **B: 19 (18, 20)**	13.25 (11.43, 15.33)	14.80 (13.10, 15.15)	11.70 (11.15, 14.60)	11.20 (10.00, 13.85)	15.30 (12.80, 15.65)	11.60 (10.50, 12.20)	10.70 (10.55, 11.70)	11.20 (10.45, 11.80)	12.10 (11.80, 12.25)	11.50 (10.45, 12.30)
Partial pressure of CO_2 (mmHg) *RR 35-38* **B: 31 (28, 33)**	14.9 (12.3, 20.1)	17.2 (13.6, 20.5)	13.1 (11.4, 17.9)	13.3 (9.7, 16.8)	16.5 (14.1, 21.6)	10.7 (9.3, 12.4)	10.5 (8.9, 13.3)	10.0 (8.9, 11.2)	10.2 (9.6, 12.1)	11.0 (10.1, 12.6)
Glucose (mmol/L) *RR 3.3–6.4* **B: 5.0 (4.7, 5.5)**	5.2 (4.8, 5.7)	5.1 (4.8, 5.5)	5.2 (4.7, 5.9)	5.1 (4.6, 5.8)	5.2 (5.0, 5.5)	4.9 (4.6, 5.4)	5.4 (4.5, 5.5)	4.9 (4.6, 4.9)	5.0 (4.5, 5.3)	4.9 (4.8, 5.5)
Lactate (mmol/L) *RR < 2.0* **B: 1.4 (1.0, 1.6)**	2.65 (1.90, 3.32)	2.50 (2.15, 3.05)	2.60 (1.65, 3.80)	4.40 (2.80, 4.55)	2.30 (1.85, 2.90)	2.95 (2.48, 3.70)	3.00 (2.80, 4.00)	2.70 (2.30, 3.30)	3.40 (1.80, 3.75)	3.00 (2.50, 3.20)
Hematocrit (%) *RR 35–50* **B: 47 (44, 49)**	47 (46, 50)	47 (45, 47)	47 (47, 49)	48 (48, 51)	47 (46, 50)	46 (43, 47)	46 (44, 48)	45 (43, 46)	46 (46, 50)	43 (43, 46)
Urine Specific Gravity **B: 1.065 (1.046, 1.071)**	1.055 (1.042, 1.068)	1.057 (1.046, 1.066)	1.055 (1.039, 1.069)	1.050 (1.022, 1.063)	1.054 (1.048, 1.069)	1.024* (1.014, 1.046)	1.041 (1.014, 1.045)	1.032 (1.018, 1.047)	1.016 (1.014, 1.017)	1.045 (1.018, 1.062)
Chloride (mmol/L) *RR 109–120* **B: 118 (115, 119)**						119 (117, 121)	120 (118, 120)	119 (117, 121)	118 (116, 122)	118 (117, 122)
Blood Urea Nitrogen(mmol/L) *RR 3.6–9.3* **B: 6.4 (5.7, 7.1)**						6.4 (5.7, 7.1)	6.8 (5.7, 6.8)	6.4 (5.3, 7.1)	6.8 (6.4, 7.1)	6.1 (5.7, 7.1)
Creatinine (umol/L) *RR 62–159* **B: 115 (97, 115)**						114.9 (113.1, 123.8)	123.8 (114.9, 123.8)	114.9 (106.1, 123.8)	123.8 (114.9, 132.6)	114.9 (110.5, 119.3)
Creatine Kinase (U/L) *RR 46–467* **B: 66 (50, 88)**						158* (128, 246)	124 (114, 217)	172 (147, 302)	322[a] (152, 562)	150 (136, 183)
Fractional Excretion of Sodium (%) **B: 0.26 (0.11, 0.37)**						0.70* (0.39, 1.28)	0.67 (0.28, 1.11)	1.04 (0.53, 1.51)	0.69 (0.40, 0.96)	0.67 (0.45, 1.15)

[1]*Statistics presented: median (IQR) W, water; OES, chicken-flavored oral electrolyte solution; CHK, chicken-flavored water; SCE, subcutaneous electrolyte solution; RR, reference range; B, baseline (pretracking value); a, significantly different from the other treatment strategies ($p = 0.0361$); *, significantly different from baseline.*

All values are reported in international units Reference range provided for point of care blood analyzer cartridge, (36) (CG8+ ISTAT cartridge, Abaxis, Union City, California; Na, K, iCa, pH, HCO3, pCO2, Glu, Hct) and veterinary clinical laboratory (Clinical Laboratory, M. J. Ryan Veterinary Hospital, University of Pennsylvania, Philadelphia PA, USA; Cl, BUN, Creat, CK).

Effect of Activity and Environmental Conditions

Median body weight was significantly lower after Track 2 compared to baseline when controlling for hydration strategy ($p = 0.041$). USG, FeNa, and CK also varied significantly between baseline and Track 2 ($p = 0.016$, $p = 0.005$, $p < 0.001$, respectively). We were unable to detect a difference in BUN after Track 2 compared to baseline BUN. Peak core body temperature was not associated with either ambient humidity or ambient temperature.

DISCUSSION

The dogs in this study faced harsher conditions and exhibited more rigorous activity than dogs in a previous study (16). They tracked in an open arid environment with no access to shade and displayed higher heart or pulse rates, higher body temperatures, and higher activity counts. The highest core body temperature recorded in the previous study was 40.3°C (104.5°F) (16) while the dogs in the current study had average peak temperatures above 41.1°C (106°F). Accelerometer data has been categorized into sedentary, walking, and trotting activity in pet dogs, although the cutoff between trotting and more rigorous activity was not defined (17). The average speed of tracking was 2.4 miles/hr with activity counts for most of the tracking sessions above the 1,751 counts per minute threshold indicative of trotting (17). In comparison, the dogs in the previously published hydration study had a mean activity of 750 counts/min although this did include rest intervals, but they were rarely observed trotting (16). Based on activity counts and visual observation, the tracking activity of the dogs in this study could be classified as moderate intensity.

In this study of dogs undergoing moderate activity while tracking in hot and arid conditions at the border in El Paso, Texas, the 4 hydration strategies had negligible effects on blood and urinary parameters. Of all the hematological, blood serological, and urinary parameters measured, hydration strategy only had a statistically significant impact on the change in CK. The increase in CK in the dogs given chicken-flavored water, suggests that this strategy is associated with a higher degree of muscle injury. Small, but significant changes in CK were also reported after a 4-h search and rescue exercise (Spoo, 2015). The post track CK for all but 2 dogs in the study was within the normal reference range and no dog showed signs of rhabdomyolysis. Bruchim et al. (2019) reported that military working dogs with a history of heat stroke, had higher CK after indoor (but not outdoor) exercise compared to military working dogs without a history of heat stroke. None of the dog handlers in the current study reported a history of heat injury or heat stroke, and since it was a cross over design, any individual dog effects are accounted for. Although the mechanism of muscle injury cannot be determined in this study, hydration strategy did have a significant impact on total fluid intake. The higher fluid consumption without electrolytes in the CHK strategy may be a contributing factor.

Dogs consumed less fluids with the W treatment suggesting the benefit of pre-treatment with an alternate hydration strategy. Dogs on the SCE treatment were guaranteed to receive 15 mL/kg of fluids prior to the start of Track 1 while dogs voluntarily consumed all the pre-tracking fluids before both sessions when offered OES or CHK in all but one case. While flavoring likely accounted for the willingness to consume CHK and OES pre-tracking fluid, the added sodium and chloride in the OES and SCE fluids likely drove subsequent water intake as studies have demonstrated that dogs will increase their water consumption following increased salt intake (25–27).

Serum Na did not vary by hydration strategy so despite the additional Na load from the OES and SCE protocols, dogs were able to maintain normal serum Na levels. The FeNa did not vary by hydration strategy and was increased after Track 2 compared to baseline regardless of hydration strategy. Similarly, USG did not vary by hydration strategy but was lower after Track 2 compared to baseline. This finding was in contrast to the expectation that the USG would increase to conserve water while dogs were exercising in a hot environment. The decrease in USG following repeated tracking exercises suggests that the dogs were either more hydrated or that there was a cumulative effect of exercise on the dogs' ability to concentrate their urine. The average weight after Track 2 was lower than baseline which suggests that the dogs were losing water not becoming more hydrated. Both the decrease in USG and the increase in FeNa likely represent a cumulative effect of exercise. We have previously documented an increase in FeNa in Border Patrol dogs following mild to moderate exercise (16). Increased excretion of Na may contribute to increased water loss due to osmotic drag which would cause a subsequent decrease in USG, a loss in weight, and potentially a decrease in serum Na.

It has been suggested that electrolyte supplementation is unwarranted for working dogs given that they pant rather than sweat to cool themselves (28). However, this study found some potential benefit and no electrolyte abnormalities associated with electrolyte supplementation (OES or SCE) and is consistent with previous study (16) which found no adverse effects associated with either oral electrolytes or subcutaneous fluids. In a field study of racing sled dogs by Hinchcliff et al. (19), prolonged exercise did lead to hyponatremia. In contrast to our dogs, the sled dogs had decreased urinary sodium and no change in urine osmolality. It is possible that the combination of prior fitness training, unique diet, endurance activity and cold temperatures may elicit a different physiologic response. We did not measure aldosterone or vasopressin; therefore, we cannot directly compare the results. During exercise, sodium can also be lost through salivation (29) and urine. The prior studies of Hinchcliff et al. (19) and Otto et al. (16) suggest that electrolyte supplementation may have value in exercising dogs. The results of this study and Otto et al. (16) suggests electrolyte supplementation is not only safe but may also be beneficial by providing sodium to counter endogenous losses and helping to drive consumption of fluids necessary to enhance evaporative cooling.

We did not detect a correlation between ambient temperature or ambient humidity and peak core body temperature which is consistent with reports from a previous study in working dogs (18). The dogs were all from warm climates so the independence of core body temperature from ambient conditions may reflect acclimation to the environment and work. The dogs' average peak core body temperatures during tracking were above 41°C (105.8°F) which is conventionally considered indicative of heat stroke and potential for permanent brain damage (30–32) although this temperature-based criterion has been criticized (33). Despite the high core body temperatures recorded throughout this study, only one dog showed signs of heat stress. On the first study day, one dog reached a core body temperature above 42.8°C (109°F) and displayed unsteadiness and shade-seeking behavior upon completion of Track 1. He was wet down and closely monitored but recovered uneventfully and was able to complete the second tracking session that day although his Track 2 time was slow (over an hour to complete the one-mile track). Later analysis of his blood CK level (3,391 U/L) suggested muscle injury, although there was no evidence of pigmenturia. Another dog that same day reached an even higher temperature of 42.88°C (109.18°F), yet did not have any behavioral or blood value changes indicative of heat stress. Similarly, during many of the other tracking sessions throughout the study, dogs reached core body temperatures above 41°C (105.8°F) with no signs of heat stress. Other studies on working dogs have also reported core body temperatures (measured with internal core temperature sensing capsules) above 41.1°C (106°F) with no clinical signs of heat stress (18, 34). Heat-acclimation has been shown to elevate the core body temperature threshold for thermal injury in rats (35), and increased transcription of heat shock proteins has been documented in MWD following acclimatization and physical training (10). Likely, heat-acclimation, acclimatization, and exercise conditioning accounts for the dramatically elevated temperatures observed in working dogs without subsequent heat stroke, but further studies are needed to sort out the role and underlying mechanisms of these factors in regard to this phenomenon.

This study looked at various hydration methods strictly as pre-treatments, and only water was offered to the dogs while they tracked. Some SAR dog handlers report administering SCE prophylactically prior to a shift (12), but repeated SCE administration is impractical. Additionally, OES and CHK can spoil in the heat after preparation. Some individuals flavor water to increase palatability; however, the electrolyte content can be highly variable, and, as seen in the dogs in the CHK strategy, the absence of electrolytes may be associated with an increased risk of muscle injury.

There are several limitations to this study. Weight loss was used as a proxy for water loss through evaporation, urine, or saliva. However, these losses were not directly measured so weight loss could potentially have been due to other factors (e.g., defecation or scale fluctuations). The amount of water consumed may have been impacted by loss during drinking; however, since the study was a crossover design, dogs that were more likely to splash water would have done so in each treatment arm. This study relied on a small number of dogs that were all from Texas, and represented different ages and breeds. Our findings may not apply to dogs from cool climates. The variability of ages and breeds are representative of the working dog population, but the variability may have masked some outcomes that could have been significant in a more uniform population. We were not able to address carry over effects since we only analyzed changes that occurred over study days. It is possible that the small sample size limited our ability to detect a significant impact of electrolyte-enriched fluids on peak core body temperature. Our data suggests that pre-treatment with electrolyte-enriched fluids may contribute to lower peak core body temperatures, but further investigation is needed, especially given the number of other factors that could contribute to peak working temperature. Finally, this study analyzed only one OES formulation, and results cannot be extrapolated to other oral electrolyte formulations as composition could drastically alter palatability, safety, and efficacy. A previous study utilizing a different OES found no increase in fluid consumption, and some dogs even refused to drink the electrolyte solution (15). More studies are needed to address the long-term effects of OES and SCE hydration protocols.

In conclusion, while this study found that hydration strategy had limited effects on blood and urinary parameters, all three alternate hydration strategies (CHK, OES, SCE) increased total fluid intake compared to W. To achieve the goal of increasing fluid intake, without the associated risk of increased muscle injury, OES and SCE pre-treatment should be considered based on field conditions and availability. Furthermore, electrolyte-enriched hydration pre-treatments (OES and SCE) may additionally help dogs to maintain a lower peak temperature although further studies are required to explore this trend.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The animal protocols used in this study were reviewed and approved by both the University of Pennsylvania and US Army Medical Research and Materiel Command Institutional Animal Care and Use Committees (USAMRCM #SO120002, UPenn IACUC protocol #804293). Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR'S NOTE

A. COIN is a software library in R. Conditional inference procedures for the general independence problem including two-sample, K-sample (non-parametric ANOVA), correlation, censored, ordered and multivariate problems.

AUTHOR CONTRIBUTIONS

GN led the manuscript preparation and review. EH and RB participated in data analysis and manuscript review. LB, TD, JN, and KS participated in data collection, and manuscript review. KK participated in study design, data collection, and manuscript review. CO oversaw the design of the study, assisted with data collection, data analysis and manuscript preparation and review.

FUNDING

Supported by the Department of Defense-Special Operations Command USSOCOM W81XWH-13-1-0038.

REFERENCES

1. Kozlowski S, Brzezinska Z, Kruk B, Kaciuba-Uscilko H, Greenleaf JE, Nazar K. Exercise hyperthermia as a factor limiting physical performance: temperature effect on muscle metabolism. *J Appl Physiol.* (1985) 59:766–73. doi: 10.1152/jappl.1985.59.3.766
2. Moore GE, Burkman KD, Carter MN, Peterson MR. Causes of death or reasons for euthanasia in military working dogs: 927 cases (1993-1996). *J Am Vet Med Assoc.* (2001) 219:209–14. doi: 10.2460/javma.2001.219.209
3. Bruchim Y, Klement E, Saragusty J, Finkeilstein E, Kass P, Aroch I. Heat stroke in dogs: a retrospective study of 54 cases (1999–2004) and analysis of risk factors for death. *J Vet Intern Med.* (2006) 20:38. doi: 10.1892/0891-664020062038:HSIDAR2.0.CO;2
4. Evans RI, Herbold JR, Bradshaw BS, Moore GE. Causes for discharge of military working dogs from service: 268 cases (2000–2004). *J Am Vet Med Assoc.* (2007) 231:1215–20. doi: 10.2460/javma.231.8.1215
5. Miller L, Pacheco GJC, Janak J, Grimm RC, Dierschke NA, et al. Causes of death in military working dogs during operation iraqi freedom and operation enduring freedom, 2001–2013. *Mil. Med.* (2018) 183: e467–74. doi: 10.1093/milmed/usx235
6. Greenleaf J, Kolzowski S, Nazar K, Kaciuba-Uscilko H, Brzezinska Z, Ziemba A. Ion-osmotic hyperthermia during exercise in dogs. *Am J Physiol-Leg Content*. (1976) 230:74–9. doi: 10.1152/ajplegacy.1976.230.1.74
7. Horowitz M, Nadel ER. Effect of plasma volume on thermoregulation in the dog. *Pflugers Arch Eur J Physiol.* (1984) 400:211–3. doi: 10.1007/BF00585045
8. Hemmelgarn C, Gannon K. Heatstroke: thermoregulation, pathophysiology and predisposing factors. *Compend Contin Educ Vet.* (2013) 35:E4.
9. Horowitz M. Do cellular heat acclimation responses modulate central thermoregulatory activity? *Physiology*. (1998) 13:218–25. doi: 10.1152/physiologyonline.1998.13.5.218
10. Bruchim Y, Aroch I, Eliav A, Abbas A, Frank I, Kelmer E, et al. Two years of combined high-intensity physical training and heat acclimatization affect lymphocyte and serum HSP70 in purebred military working dogs. *J Appl Physiol.* (2014) 117:112–8. doi: 10.1152/japplphysiol.00090.2014
11. Johnson K. Pathophysiology of heatstroke. *Compend Contin Educ Pract Vet.* (1982) 4:141–4.
12. Gordon LE. Injuries and illnesses among urban search-and-rescue dogs deployed to Haiti following the January 12, 2010, earthquake. *J Am Vet Med Assoc.* (2012) 240:396–403. doi: 10.2460/javma.240.4.396
13. Cheung SS, McLellan TM. Heat acclimation, aerobic fitness, and hydration effects on tolerance during uncompensable heat stress. *J Appl Physiol.* (1998) 84:1731–9. doi: 10.1152/jappl.1998.84.5.1731
14. Slensky KA, Drobatz KJ, Downend AB, Otto CM. Deployment morbidity among search-and-rescue dogs used after the September 11, 2001, terrorist attacks. *J Am Vet Med Assoc.* (2004) 225:868–73. doi: 10.2460/javma.2004.225.868
15. Mazin RM, Fordyce HH, Otto CM. Electrolyte replacement in urban search and rescue dogs: a field study. *Vet Ther Res Appl Vet Med.* (2001) 2:140–7.
16. Otto CM, Hare E, Nord JL, Palermo SM, Kelsey KM, Darling TA, et al. Evaluation of three hydration strategies in detection dogs working in a hot environment. *Front Vet Sci.* (2017) 4:174. doi: 10.3389/fvets.2017.00174
17. Michel KE, Brown DC. Determination and application of cut points for accelerometer-based activity counts of activities with differing intensity in pet dogs. *Am J Vet Res.* (2011) 72:866–70. doi: 10.2460/ajvr.72.7.866
18. Robbins PJ, Ramos MT, Zanghi BM, Otto CM. Environmental and physiological factors associated with stamina in dogs exercising in high ambient temperatures. *Front Vet Sci.* (2017) 4:144. doi: 10.3389/fvets.2017.00144
19. Hinchcliff KW, Reinhart GA, Burr JR, Schreier CJ, Swenson RA. Effect of racing on serum sodium and potassium concentrations and acid-base status of alaskan sled dogs. *J Am Vet Med Assoc.* (1997) 210:1615–8.
20. Rosenbaum PR. Interference between units in randomized experiments. *J Am Stat Assoc.* (2007) 102:191–200. doi: 10.1198/016214506000001112
21. Edgington ES, Onghena P. Chapter 6. In: *Randomization Tests* Statistics: *A Series of Textbooks and Monographs.* New York, NY: Chapman and Hall/CRC (2007). doi: 10.1201/9781420011814
22. Hothorn T, Zeileis A. Generalized maximally selected statistics. *Biometrics*. (2008) 64:1263–9. doi: 10.1111/j.1541-0420.2008.00995.x
23. Hothorn T, Hornik K, van de Wiel MA, Zeileis A. Implementing a class of permutation tests: the coin package. *J Stat Softw.* (2008) 28:1–23. doi: 10.18637/jss.v028.i08
24. Laflamme D. Development and validation of a body condition score system for dogs. *Canine Pract.* (1997) 22:10–5.
25. Cizek LJ. Long-term observations on relationship between food and water ingestion in the dog. *Am J Physiol-Leg Content*. (1959) 197:342–6. doi: 10.1152/ajplegacy.1959.197.2.342
26. O'Connor WJ. Drinking by dogs during and after running. *J Physiol.* (1975) 250:247–59. doi: 10.1113/jphysiol.1975.sp011052
27. Stevenson AE, Hynds WK, Markwell PJ. Effect of dietary moisture and sodium content on urine composition and calcium oxalate relative supersaturation in healthy miniature schnauzers and labrador retrievers. *Res Vet Sci.* (2003) 74:145–51. doi: 10.1016/S0034-5288(02)00184-4
28. Wakshlag J, Shmalberg J. Nutrition for Working and Service Dogs. *Vet Clin North Am Small Anim Pract.* (2014) 44:719–40. doi: 10.1016/j.cvsm.2014.03.008
29. de Beer E, Wilson D. The inorganic composition of the parotid saliva of the dog and its relation to the composition of serum. *J Biol Chem.* (1932) 95:671–85.
30. Lewis S, Foster RC. Effect of heat on canines and felines. *Iowa State Univ Vet.* (1976) 38:117–21.
31. Flournoy WS, Wohl JS, Macintire DK. Heatstroke in dogs: pathophysiology and predisposing factors. *Small Anim Compend.* (2003) 25:410–8.
32. Bruchim Y. Canine Heatstroke. *Isr J Vet Med.* (2012) 67:92–5.
33. Baker JL, Hollier PJ, Miller L, Lacy WA. Rethinking heat injury in the SOF multipurpose canine: a critical review. *J Spec Oper Med Peer Rev J SOF Med Prof.* (2012) 12:8–15.
34. Angle TC, Gillette RL. Telemetric measurement of body core temperature in exercising unconditioned Labrador retrievers. *Can J Vet Res.* (2011) 75:157–9.
35. Horowitz M. From molecular and cellular to integrative heat defense during exposure to chronic heat. *Comp Biochem Physiol, Part A Mol Integr Physiol.* (2002) 131:475–83. doi: 10.1016/S1095-6433(01)00500-1
36. i-STAT Alinity v Utilization Guide ABAXIS Inc. Available online at: https://www.abaxis.com/sites/default/files/resource-brochures/887-0200%20Rev.%20E%20i-STAT%20Utilization%20Guide_no%20crop.pdf (accessed August 14, 2018).

Conflict of Interest: EH was employed by the company Dog Genetics, LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

REVIEW
published: 05 June 2020
doi: 10.3389/fvets.2020.00313

A Review of the Types of Training Aids Used for Canine Detection Training

Alison Simon[1]*, *Lucia Lazarowski*[2], *Melissa Singletary*[2], *Jason Barrow*[3], *Kelly Van Arsdale*[3], *Thomas Angle*[2], *Paul Waggoner*[2] and *Kathleen Giles*[4]

[1] AGS Forensics, LLC, Washington, DC, United States, [2] Canine Performance Sciences, College of Veterinary Medicine, Auburn University, Auburn, AL, United States, [3] Federal Bureau of Investigation Laboratory, Quantico, VA, United States, [4] Giles Consulting, LLC, Huntington Beach, CA, United States

OPEN ACCESS

Edited by:
*Cynthia M. Otto,
University of Pennsylvania,
United States*

Reviewed by:
*Kelly Alan Mann,
Colorado State University,
United States
Avi Avital,
Technion Israel Institute of
Technology, Israel*

***Correspondence:**
*Alison Simon
agsforensics@gmail.com*

Specialty section:
*This article was submitted to
Veterinary Humanities and Social
Sciences,
a section of the journal
Frontiers in Veterinary Science*

Received: *14 February 2020*
Accepted: *06 May 2020*
Published: *05 June 2020*

Citation:
Simon A, Lazarowski L, Singletary M, Barrow J, Van Arsdale K, Angle T, Waggoner P and Giles K (2020) A Review of the Types of Training Aids Used for Canine Detection Training. Front. Vet. Sci. 7:313. doi: 10.3389/fvets.2020.00313

The canine detection community is a diverse one, ranging from scientific fields such as behavior, genetics, veterinary medicine, chemistry, and biology to applications in law enforcement, military, medicine, and agricultural/environmental detection. This diversity has allowed for a flourishing and innovative community, yet it has also led to little acceptance and agreement on terminology. This is especially true when discussing the variety of training aids used in olfactory-based exercises. In general, authentic materials and pseudo-scents are the most commonly discussed, with the former accepted widely for training and certification, and the latter more often disregarded. However, as advances are made in the creation of training materials, alternative training aids are being introduced that do not fit into either of these categories. The misconceptions surrounding how these alternative training aids are manufactured has led to confusion on their classification, and therefore their reliance as an effective tool. This manuscript will review the existing language surrounding canine training aids, address relevant research revealing effectiveness, and clarify the different types based on their manufacture, chemical nature, and fundamental function.

Keywords: training aids, canine detection, pseudos, terminology, non-pseudo alternatives

INTRODUCTION

Target substances in canine detection are exceedingly varied, ranging from traditional materials such as narcotics, explosives, human scent, and human remains, to less common or emerging targets such as diseases, pests, and wildlife. This diversity of targets is mirrored by the professional community. Opinions, research, and experience from canine handlers and trainers, behavioral sciences, genetics, veterinary medicine, and analytical sciences, as well as various organizations and government agencies, influence the training methods and protocols of canine teams. This wealth of information has made the canine community inventive and successful. Yet, there is a considerable lack of agreement across the community regarding a standard terminology. Such disparity complicates effective transfer of knowledge across the canine industry, impeding advancements in technology and methodologies.

The variety of jargon specific to the canine community is especially apparent when referring to types of training aids. Training aids can be created by the onsite trainer, by an assisting specialist (such as a bomb technician), or in a laboratory. Such aids differ based on their manufacture, chemical nature, and fundamental function, yet this specific information is rarely discussed, is often proprietary, and has limited third-party evaluation or support. Thus, there is often confusion regarding what to call certain categories of training aids. The purpose of this article is to discuss current existing jargon, and to define them based on their function and chemical nature.

There are two assumptions made in this article regarding terms and definitions. First, the terminology of *odor* and *scent* used herein are derived from the Organization for Scientific Area Committees (OSAC) Dogs and Sensors subcommittee (1). OSAC is an organization administered by the US National Institute of Science and Technology (NIST), which replaced the Scientific Working Group for Dogs and Orthogonal detector Guidelines (2), and makes standards and guidelines for the canine detection community. *Odor* refers to the "volatile chemicals emitted from a substance that are able to be perceived by olfaction," while *scent* refers specifically to the "volatile chemicals emitted from a live human" (1).

Second, the article will focus specifically on sources of odor and scent rather than odor delivery systems or transport containers. The odor/scent source is the training aid itself, or the object providing the target odor/scent. Odor delivery systems are devices that contain the training aid, such as "scent" boxes, the Mixed Odor Delivery Device (MODD) (3), the Training Aid Material Delivery Device (TAMDD) (4), or the Training Aid Delivery Device (5), for example. Transport containers are used to "move training aids in compliance with storage and handling guidelines of the Federal, state, and/or local agencies' policy" (1).

Considering these assumptions, the following discusses three categories of training aids as determined by different methods of manufacture: true material, pseudo-odors, and non-pseudo alternatives (see **Figure 1**). True material is the actual target substance. Pseudo-odors are created in a way so that the true material has no direct part in their manufacture. Non-pseudo alternatives are made through utilization of the true material in their manufacture. Each of these sources of odor/scent are examined in detail below, along with discussions of their function. We will also examine any existing research which sheds light on their efficacy and accuracy as training aids.

TRUE MATERIAL

True material, also referred to as bulk material, actual material, genuine material, or parent material, refers to the target substance itself, whether it be an explosive, a narcotic, human remains, or any other target. True materials may be in a solid [e.g., composition-4 (C4) or cocaine], liquid [e.g., nitromethane or ethylene glycol dinitrate (EGDN)], or gaseous (e.g., human scent or certain chemical warfare agents) phase. For solid and liquid true materials, canines generally locate the source of the odor, whereas for gaseous true materials, they may simply be identifying the presence or absence of the odor/scent. These substances are currently what is recommended best practice for use in training and certification (2), though, as will be discussed, other types of training aids may serve as suitable training materials. However, it is unlikely that true material will be replaced in certifications, since those records are necessary

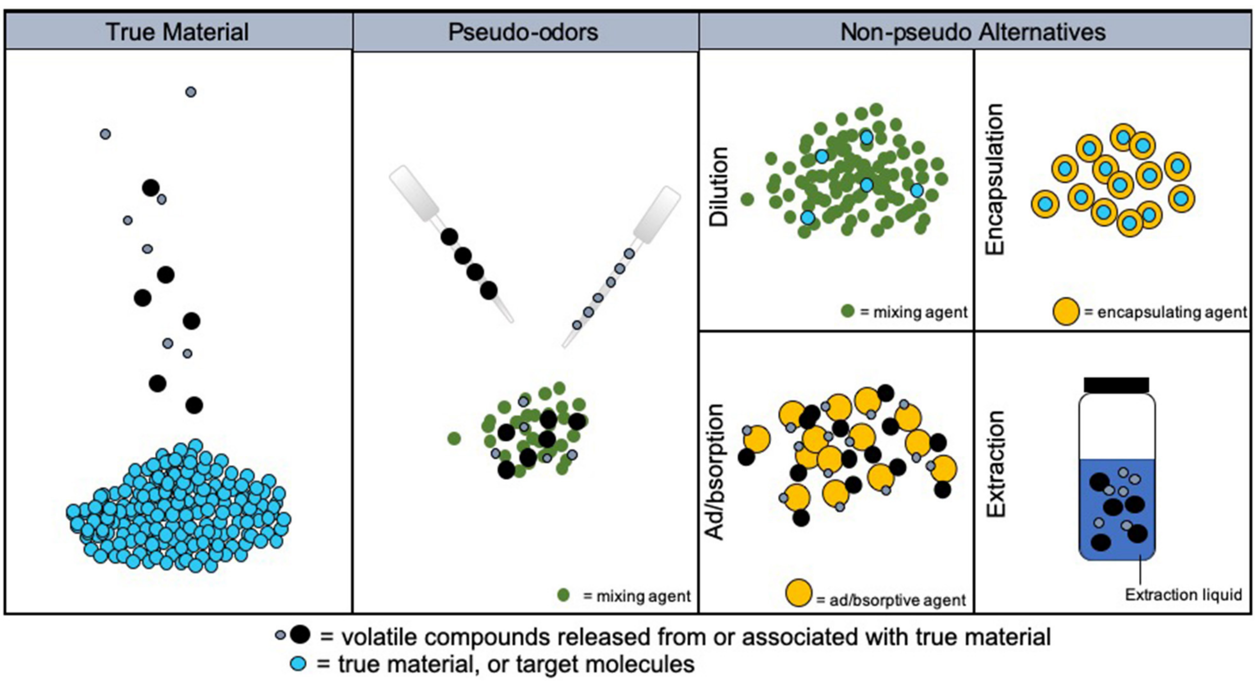


FIGURE 1 | Visual representation showing the chemical and manufacture differences between true material, pseudo-odors, and non-pseudo alternatives, given the same true material.

for most law enforcement and military organizations where canine searches may result in probable cause for entry or evidence collection.

While the use of true materials may seem straightforward, there are actually many considerations to be made in the selection of these materials. It is generally accepted that training dogs on the purest form of a substance is the best method for ensuring reliable detection. However, training on a single pure odor has been shown to produce a strong response and subsequent detection to that specific odor while narrowing the tendency to respond to variations of it (6). This degree of specificity presents a challenge due to the high variability in targets that a dog may encounter, such as homemade explosives or improvised explosive devices that often consist of mixtures of various compounds. Recent publications suggest that in addition to the pure odor, the inclusion of additional mixtures in training could improve a canine's ability to generalize in other contexts of background or conflicting odors. For example, studies by Lazarowski et al. (7) and Hall and Wynne (8) each found that canines trained to detect ammonium nitrate (AN) utilizing only a pure source did not proficiently [to the Military Working Dog criterion of 95% success (9)] locate mixtures containing AN. However, when mixtures were included in training scenarios, the canines' proficiency in detecting other AN-based mixtures increased (8). Even when considering AN apart from mixtures there is some debate as to whether industrial grade or laboratory grade should be considered purer. DeGreeff et al. (10) analyzed the volatiles from seven different sources of AN, and found different quantities of ammonia and contaminates were present, depending on the manufacture method and form of the AN (i.e., prill or ground). These studies speak to the difficulty of selecting a pure training material that will lead to the greatest success of the canine team. While it is folly to assume that these same patterns will be observed in all true target materials, there is existing literature that identifies differences in either canine behavior or the chemical odor or scent profiles for various pure materials: human remains (11–13), human scent (14–16), blood (17, 18), potassium chlorate mixtures (19–21), trinitrotoluene (TNT) (22, 23), hexamethylene triperoxide diamine (HMTD) (24, 25), single- and double-based smokeless powders (26, 27), marijuana (28, 29), synthetic cathinones (30), and methamphetamine (31), among others. However, there is also literature implying that for certain highly volatile compounds, canines readily perceive variants of a true material as the same, for example nitromethane (32), triacetone triperoxide (TATP) (33) and accelerants (34).

While the selection of a true material may seem complex based on this information, it is often directed by departmental, state, or federal regulations, or based on operational needs. Depending on the legality of obtaining, maintaining, and transporting a substance, certain forms of that substance may be regulated to a point that prohibit a team from training on certain forms. This issue is often encountered, for example, by human remains detection teams. It may not be possible to obtain visually identifiable remains in some states within the United States, and it is not possible to obtain any human remains for training purposes in many countries. Therefore, many handlers turn to human teeth or pig remains as surrogate training materials (35, 36). While human teeth may be a portion of human remains, there is ample chemical evidence that these odors differ from other types of human remains (11, 36).

As another example of limited choice for true materials, safety of an explosive, toxic, or infectious material can limit access. Most explosives teams within the United States train on peroxides provided by a single federal entity. The peroxide HMTD has been shown to have extremely variable odor profiles depending on synthesis and age (24, 37). Yet, for safety purposes, a single, high purity synthesis route is followed for the creation of HMTD canine training aids. This means that despite drastic differences in the spectrum of HMTD odor profiles, there is actually only a single option for a true HMTD training aid. Toxic hazards are also tightly regulated, often limiting certain narcotics training aids, such as fentanyl or synthetic cathinones, to seized material. Such materials can have high purity, depending on the seizure, but it does limit greatly the choice of synthesis route and manufacturing method that may be desired for certain operational requirements. For example, in one study of synthetic cathinones, canines trained to detect seized samples were successful in locating another bath salt [seized α-pyrrolidinopentophenone (α-PVP) and ethylone were each used to imprint a group of canines, (38)]. However, bath salt formulations vary greatly by region, so this may not hold true in all jurisdictions.

Viruses, bacteria, diseases, pests, agriculture, and wildlife/conservation, which are growing applications for canine detection, have other complications to consider. These target materials are exceedingly varied based on the age of the disease, infection, or agricultural product, leading to wide variability in the range of odor profiles that the dog needs to be able to detect. Some diseases have periods of time when a person, animal, or plant is infected with a virus or bacteria, but not yet ill or displaying symptoms. Plus, training aids are often presented in the context of extremely high background, for example in fecal samples, or background that could be very interesting to a dog's natural abilities and desires, such as urine. Such background odors or scents could confuse the actual target, and complicate generalization to other sources of the target odor. Each of these considerations are key when selecting which true material will be used for imprinting the odor and for training in operational scenarios, when even more extraneous background odors/scents may be encountered, such as an outdoor environment at a port or a grove, or even detection in a live species. There are a handful of manuscripts describing choices and chemical analysis of true material for such detection purposes: for cancers (39, 40) and other medical alerts (41), agriculture (42), and wildlife or pests (43, 44). Of course, safety and regulation can also limit these choices. Infectious materials can be tightly regulated to help stint the spread of diseases, limiting training with the true material to a specific geographic area. As another example, bed bug detection has become a common application of canines, yet the age of the bugs can influence true material choices. Further, keeping the bugs or insects from spreading is imperative when training, even though no specific regulations exist, which may influence the choice of age. One recent study evaluating training aids for the detection of an insect pest in Australia found that canines trained using dead insects successfully found the live counterpart on the majority of the test trials (45). However, the sample size was small with a limited number of observations and the authors noted that such generalization may not occur with other, non-specialized and less experienced canines.

In addition to selecting a true material with which to imprint, train, and certify, there are considerations to take into account in the maintenance of the material, which can alter its odor/scent profile and affect the way a canine perceives it. While the age of a substance is of particular importance for such living true materials, it is also key for other materials. For example, chemical analysis of HMTD over time has shown how unstable the compound is. The odor profile changes drastically due to natural decomposition of the molecule (24, 25, 37). This is also the case for other compounds, such as TNT, AN, or cocaine (10, 46, 47).

A related consideration is how often target materials should be varied in training. For example, human scent is unique to each individual (48). Yet, if a canine team does not alter their true material (in this case a human decoy or target), then the canine may learn to identify only that person. It has been shown that canines associate positively with familiar human scents, even when the familiar scent is not their handler (49). The tendency to associate a scent with an individual person creates a particular challenge in the training of disease detection. For example, dogs trained to detect prostate cancer apparently memorized the set of training samples rather than the common cancer odor profile, as their detection dropped when tested with samples from new individuals (50). This memorization of people as targets is a learned behavior, and an issue that can be remedied by varying the target material so that memorization becomes too difficult and learning the common odor class is a more efficient strategy for the canine. A similar issue may arise outside the subdiscipline of human scent and arises from handling. If only one person handles the material for all training sessions for a substance, the canine may associate that person's scent with the true material. While this is also a learned behavior, it relates more to a lack of experimental controls than to the target material itself. However, if the target is not varied, then that single person's scent will remain on the target and will continue to be associated with the target's odor profile. It is also worth noting that these inadvertent clues caused by familiar human scent are separate from other handler influences that may evolve in the course of canine training, such as handler stress or handler beliefs (51, 52).

As another example of varying the true material, contamination and cross-contamination can occur if materials are stored too closely together. In a study from 1997, Hallowell et al. (53), found that canines trained to detect nine separate explosives using true material could only identify the most volatile explosive species. The true materials were all stored in the same explosives bunker, and cross-contamination confused the odor profiles. These examples all help to demonstrate that even when using true materials, there are many considerations that must be made to ensure that the true material is actually representative of the desired target substance, and is therefore an effective training material.

PSEUDO-ODORS

True material, while generally considered best practice, is not always available due to cost, handling, storage, transportation, safety, and security challenges (54). Therefore, several types of alternative training aids have been developed, one of which is referred to as pseudo-odors, "pseudos," odor mimics, or simulants. "Pseudo" is a term that is often used in the canine community, yet is infrequently applied correctly or consistently. In general, a pseudo training aid is one in which the true material had no part in its manufacture. The most common method for making a pseudo training aid is to identify the major chemical components in the headspace of the true material, and use pure (or neat) compounds to create a physical mixture of those components intended to simulate the odor profile. One technical magazine review separated these training aids into four types based on which chemical components were included in the training aid: active odorant, byproduct or impurity, filler or additive, and a non-related volatile that attempts to mimic the perceived smell of the target (55). As will be discussed, the pseudos composed of active odorants tend to be more supported in the published literature; however, there is a lack of published information regarding many proprietary pseudos, and these remain unverified by an independent evaluator.

There are pros and cons to the pseudo approach in training. A clear advantage to using pseudos is they can be used as a training material when access to the true material is regulated, whether for safety or legal reasons. They can therefore provide access to substances to a greater number of canine teams. For example, a study that used sarcosine as a target molecule for canines trained to detect prostate cancer observed sensitivity and specificity to sarcosine that was comparable to the detection of prostate cancer in urine (56). The use of such a pseudo-odor could improve canine detection training for medical purposes by easing access to training materials and increasing their uniformity. However, there are many problems associated with pseudos as well. The most prominent disadvantage is associated with complex target materials, such as human remains and HMTD. As discussed above, such true materials change odor profiles dramatically with time, environment, and storage conditions. This makes creating accurate pseudo training aids very complicated. From the most basic perspective, the presence of varied odor profiles corresponding to the target means that one combination of neat chemicals does not truly represent the various odor profiles of the true material. Additionally, pseudos often produce much more odor than the original material, which could provide a chemical hazard to the canine. The amount of odor could also affect the canines' perception of that trained odor, influencing threshold. For example, dogs trained to detect a certain quantity of an explosive do not necessarily respond to larger or smaller quantities of the same material. Further, the material used to contain the chemicals, generally cellulose, diatomaceous earth, low-density polyethylene (LDPE) bags, or similar materials, also produce odors that could change perception of the odor, especially if blanks or negative controls are not provided. Blanks and negative controls in this instance refer to the use of packaging materials within a training session to help proof (or proof off) of extraneous materials. Many of these challenges are also applicable to non-pseudo alternative training aids, as will be discussed.

The unsuitability of multiple pseudo training aids has been shown both chemically and in canine behavioral studies. Studies of several narcotic and explosive pseudo training aids, plus human decomposition pseudos have demonstrated that these substances are not generally successful in canine trials. For example, one study examining the effectiveness of pseudo training aids for single-based smokeless powder, TNT, and C4 showed that the pseudos tested were poor simulations of the true material, despite containing previously-identified explosive-related odors in their headspace. Canines were trained on either the pseudo or the real explosive and then tested on the counterpart. Alert rates to the test odor were well below proficiency, ranging from 0 to 25%, indicating that the dogs did not perceive the real material and the pseudos interchangeably (57). Further evaluations of pseudos for pentaerythritol tetranitrate (PETN), hexogen (RDX), TNT, chlorate, and nitrate had similar results: none of the pseudos were effective in training dogs to detect the true material (26, 27). Heroin and marijuana pseudos have also shown ineffective. Only one out of 12 trained detection canines alerted to an acetic acid-based heroin pseudo, and no canines detected a marijuana pseudo, suggesting that the dogs perceived the pseudo as different than the true material they had been trained with (27). Rice and Koziel (58) supported these findings when they used odor activity values to examine one commercial brand's pseudo materials for heroin, marijuana, and cocaine, and determined that they do not adequately represent the odor impact of the true material. Human decomposition pseudos have been analyzed for chemical and behavioral response similarity to true material, and none of the existing pseudos (at the time of the research) were efficient odor mimics (59, 60). Such publications make the defense of pseudos in court difficult, which could invalidate an otherwise legal search or seizure.

There are several possibilities for why pseudos fail to perform in the same way as true material. Pseudos make significant changes to the original odor profile through the dilution, absence, or addition of chemical substances. These inconsistencies mean that commercial aids may not provide the same volatiles as the true material, or they may alter volatile ratios (27). Further, because pseudos are created using pure chemical compounds, these compounds must be held in place with some mixing agent, such as cellulose or diatomaceous earth, which could change the transportation properties associated with how the molecules enter the atmosphere. Without considering these physical properties, any complete comparison of a pseudo material to the material it is attempting to imitate is not possible. Finally, the pure chemicals are usually also combined in a way that provides a much larger quantity of odor than what would be encountered with the true material, which could affect canine perception of the material and could also be hazardous.

While the efficacy of pseudos has been largely disputed by scientific analysis, there are some cases where a pseudo is useful and scientifically shown to be accurate, both by chemical analysis and canine behavior. The best example of this is for cocaine. Cocaine, depending on the synthesis and manufacturer, can produce a variety of chemicals in the odor profile, such as methyl benzoate, benzoic acid, methyl cinnamate, anhydroecgonine methyl ester, trans-cinnamic acid, and ecgonine methyl ester. However, canine analyses from two independent research groups have shown that canines identify methyl benzoate as the active odorant of cocaine. In other words, canines alert to the presence of methyl benzoate, a decomposition product of cocaine, rather than the cocaine molecule itself (47, 61, 62). Further, canine alerts to cocaine have been upheld as proficient for probable cause in the Florida State Supreme Court, even given the evidence surrounding methyl benzoate as an active odorant suitable for training. In the case of *Florida v. Jardines*, the defense argued that because canines detect methyl benzoate, and methyl benzoate is also a product of snapdragon flowers and perfumes, that canines are not specific enough for an alert to serve as probable cause. Subsequent research and the final court decision ruled in favor of the canine alert (63, 64). While methyl benzoate is the most established pseudo in the detection community, there are others with published evidence of support. For example, piperanol as a pseudo for 3,4-methylenedioxymethamphetamine (MDMA) elicited a canine alert rate of 60% (65). There is equal support for the use of 2,4-dinitrotoluene (2,4-DNT) (50% alert rate) and 2-ethyl-1-hexanol (66.7% alert rate) for TNT and C4, respectively (46). While these are not operationally proficient success rates, they do show that more investigation is warranted.

NON-PSEUDO ALTERNATIVES

To address the disadvantages of pseudos, while still providing safe access to training aids, many types of non-pseudo alternative training aids have been manufactured. They vary based on the type of target as well as the chemical nature and fundamental function of the aid. Generally speaking, a non-pseudo alternative is a training aid in which the true material had a part in its manufacture, but is not present in bulk. In other words, the true material is utilized to render a safe target through various methods. These training aids are manufactured through four main methods: (1) dilution of the true material by simple mixing, (2) encapsulation of the true material inside another substance, (3) ad/bsorption of the odor of the true material, and (4) extraction of the odor from the true material. Dilution and encapsulation methods still contain trace or small amounts of the true material, while ad/bsorption and extraction of the odor do not.

Dilution

Dilution refers to taking small, or trace amounts of a target material, and mixing it with larger amounts of an inert solid or dissolving it in an inert liquid. This makes the true material safe by lowering the amount of material present, and separating the molecules from each other to remove shock sensitivity, for example. Dilution has been successful anecdotally, but no available reports have evaluated this method scientifically for accuracy or safety. For example, some HMTD training aids available on the market mix HMTD precursors with diatomaceous earth to create a final trace mixture that is non-detonable (66, 67). However, it has been shown that the synthesis method for HMTD has tremendous effect on the resulting odor profile. Without evaluation of the diatomaceous earth method to show any differences or similarities in odor profile to other methods, this method of synthesis may not be comparable. Other dilution matrices may include cellulose, glass beads, or any myriad of other materials.

The most common liquid dilution is water. This method is not commercially available, and is often done in-house. This means it is diverse, and its reproducibility has not been evaluated for each target material. The water-dilution method has been used for many years to help lower absolute threshold for hard-surface tracking of humans. A person's sweaty shirt or hat will be soaked in an un-specified amount of water. The water-scent mixture can then be continuously diluted and sprayed as a trail for the canine to follow. This method has had much anecdotal success; however, it does require a person to spray the trail, so there is generally a secondary, stronger scent trail to follow. The other most common liquid-dilution method is much more recent, and was developed by the Royal Canadian Mounted Police (RCMP) for fentanyl (68). A small amount of fentanyl is dissolved in water, dropped onto a cloth or cotton pad, and allowed to evaporate off. The canine is then trained on whatever fentanyl remains on the cloth. Again, while this method has anecdotal support, the danger exists that it has not been evaluated to determine how much fentanyl remains on the cloth for detection, which means it has not been examined for toxicity.

While the dilution method is common for human scent and many explosives and narcotics, matrix effects are not often considered. Generally, when a training aid manufacturer or canine trainer finds one dilution method that is successful, they tend to continuously use that matrix for all target materials. However, each target substance will have their own unique chemical properties that influence its interactions with that matrix. Such chemical properties as vapor pressure, diffusion rates with the material, polarity, molecular weight, and rates of evaporation or sublimation will affect how quickly a substance diffuses from the matrix. Further, these properties will determine how accurately the odor profile of the dilution training aid simulates that of the bulk material. Continuing with the HMTD example above, HMTD dissociates at room temperature. The diatomaceous earth-based training aids changed the way that the HMTD entered the atmosphere and resulted in poor representation of the bulk material (69). This same consideration must be given to liquid dilutions. For example, fentanyl is only moderately soluble in water. It is therefore plausible that very little material actually enters the liquid mixture, leaving an unknown amount in solution. Dilution training aids can provide easy-to-make non-detonable and non-toxic training aids, yet there is much research that needs to be done to prove their validity.

Encapsulation

Encapsulation of the target material is similar to dilution in that it places a trace amount of material within a matrix. The main difference is simply the mechanism (i.e., dilution or encapsulation) through which this is achieved. Existing encapsulation devices are mainly for explosives and narcotics, such as HMTD. One method for HMTD and TATP is to encapsulate the peroxide explosive inside microspheres. Heating the microspheres during canine training then releases the odor (70). This method does render the explosive safe; however, it lacks the same matrix consideration as the dilution method. Interactions with the matrix may change the odor profile so that it does not accurately reflect bulk material (69).

Ad/bsorption

Ad/bsorption of the odor of a target material onto a secondary material is another method of rendering safe a hazardous substance. Simply put, a secondary material, such as steel, cotton, or a polymer, is exposed to the headspace of a true material for a period of time to ad/bsorb the odor profile. The odor is subsequently released over time. This method has been used for many years informally to help lower absolute threshold or to make a hazardous substance easier to transport. For example, Dutch canine trainers have used steel tubes to collect human scent for scent-identification line-ups. This method does allow for easy cleaning of the collection matrix, but it may also provide uneven collection of the target material, depending on how well the steel adsorbs certain chemicals associated with human scent. In some way, this is the same method used in article search training, when a person leaves scent on an object by briefly touching the article.

Cotton and similar natural fibers are probably the most commonly used matrices to absorb odor. It is used in human scent and decomposition odor collection, in both static and dynamic ways. Static collection occurs when the cotton is left near the true material to collect the odor. Dynamic collection occurs when human scent or decomposition odor is pulled through the cotton, using a STU-100 or a similar device (71, 72). While cotton is the most common and reproducible fiber used to collect and release human scent, many other fibers such as cotton-blend, rayon, and wool are effective. Polyester has proven ineffective for the purpose (73, 74). While no canine trials were performed, investigations have been made into the static collection of target volatiles for C4 and single-based smokeless powder onto cotton gauze, with promising results for a potential training aid (75). Cotton pads have also been used to absorb the odor of a fungus to prevent the transportation of fungal spores in canine testing (76). Shelf-life or longevity of these training aids have not yet been evaluated. Human scent, decomposition, and fungal odor are extremely complicated targets, making evaluation of absorption aids similarly complex. The explosive TATP, on the other hand, has a much simpler odor, and cotton absorption training aids have been evaluated for this target. TATP-cotton training aids have been successfully deployed for canine training, but have a very short lifetime. They could only be used for about 20 min before the odor was depleted (77).

Polymer-based absorption training aids provide a more even and predictable matrix than natural fibers, which can be an advantage for creating reproducible odor profiles over time. While there are currently no polymer-based absorption training aids on the market, one has been evaluated in numerous studies. Designed at the National Institute of Science and Technology (NIST), a polydimeythlsiloxane (PDMS)-based training aid has been shown through chemical evaluations to accurately simulate odor profiles for 2-ethyl-1-hexanol, cyclohexane, DNT, and TATP (78, 79). Published studies have not yet evaluated the aid in canine trials. However, the inert and absorptive capabilities of the matrix, the chemical analyses, and the steady odor release rates over time show that the training aids deserve further investigations, which are underway.

The diversity of ad/bsorption materials that exist shows how valuable they can be, as they do not require the transportation and handling of true material, and are therefore non-detonable, non-toxic, and non-infectious. They also have a larger application than dilution and encapsulation, since they can be applied to any true substance, even diseases or infectious materials. Pathogens and infectious agents cannot necessarily be diluted or encapsulated to be rendered safe, given that these methods do not negate the infectious material in any way. There may be options for processing some pathogens by autoclave or radiation for example, but these methods should be tested for each true material to ensure safety, particularly with pathogens of highest concern to public health, safety, and economic impact.

Ad/bsorption matrices, just like the matrices used for pseudos, dilutions, and encapsulations, have considerations that cannot be overlooked. Each volatile compound will interact differently with a different matrix, so one matrix cannot be assumed to function for all target substances. Even if one matrix is applicable for multiple target materials, it will provide different saturation points and diffusion rates for each material. Such transport properties will define the rates at which various volatiles are released from the matrix, and may alter the odor profile in undesired ways if left uncontrolled.

Extraction

Liquid extraction of an odor from its source material is a process which removes chemical components of the odor from its true material using a solvent, such as the liquid pentane. Thus, the true material is removed and the remaining solvent is used for training. So far, this technique has a very limited application and has only been used to create training aids for insects and pests. This is advantageous as it omits the need to use live pests in cases where they are not native or may be undesirable. A recent publication by Moser et al. (45) demonstrated chemical similarity between the extract of one insect and the live insect. Further, they observed good canine selectivity and sensitivity (100%, $n = 2$) to the live insect following initial training on the extract. Extraction-based odors were also tested by dogs trained to detect bed bugs with a 100% response rate (80). Extraction has not been used very often in canine training, but these studies show promise in its application to create non-pseudo alternative training aids for insects.

CONSIDERATIONS FOR TRAINING AID PRODUCTION

There are several reasons why some pseudos and non-pseudo alternatives do not provide the desired success rate for canines. One is that the community does not yet fully understand the interaction of odor production, olfactory detection, and behavioral identification. It can therefore be difficult to determine exactly which targets to select or how a canine will perceive any given material. As Rice and Koziel (28, 58) have concluded, the abundance of a chemical in the headspace of a target is not equivalent to the odor impact. Chemical instrumentation and canine olfaction are different sensory systems with canines having a perceptual component to verification of the presence of a substance. Instruments and canines each have separate functional biases that do not attribute equivalent value to various volatile chemicals. While this challenge has been overcome for certain active odorants, it is difficult and full investigations are not always undertaken.

Another underlying cause for the lower than expected performance of alternative training aids is the tendency in olfactory work and research to underestimate matrix effects in order to focus on the odor/scent produced. While it is true that the odor is the target being identified or tracked, the matrix surrounding the material will influence that odor and determine how accurately an odor profile mimics that of the true material. Further, matrix effects can occur for each type of training aid: true material, pseudos, and non-pseudo alternatives. Whether the target material is merely wrapped for containment, or intentionally diluted, encapsulated, or ad/bsorbed, the matrix is important to consider for several reasons. First, many matrices, such as clear plastic bags, wood blocks, or nylon stockings, produce large volumes of odor which may compete with the target. Second, the target will interact with the matrix due to chemical properties of the materials. Vapor pressure, rates of evaporation and sublimation, polarity, diffusion rates, solubility, and molecular weight will each influence the transport properties of how a target moves through and is released from its matrix. There is no one matrix that is appropriate for all targets.

Finally, such information needs transparency. It has been previously noted [by Bradshaw (55), Simon and DeGreeff (69), among others] that the lack of third-party evaluations leads to confusion for practitioners. Training aids should be validated by chemical, biological, or physical analyses and behavioral evaluation in tandem in order to verify that what manufacturers see chemically is confirmed by the canine operationally. Third-party canine evaluations of training aids are essential to support chemical validations, and should clearly define the methods used. Assessments should be objective through such means as double-blind evaluations and distractor odors, plus should accurately measure and report sensitivity and selectivity [see for example (81, 82)]. Thorough reporting of methods for all types of analysis is significant for understanding the efficiency of training aids. Such due diligence is very important when considering operational teams that can encounter setbacks in training regimens without optimal equipment.

CONCLUSION

The goal of this article was to provide language and discussion about canine training aids based on the available knowledge.

Canine training aids can be categorized based on the level of contribution of the true material in their manufacture. True material is simply the actual target substance, whether solid, liquid, or gas. Pseudo-odors are manufactured without direct influence of the true material. Non-pseudo alternatives utilize the true material in their manufacture and render safe the true material in some manner. Within those categories, their chemical nature and function should be used for labeling. Specifically, dilution and encapsulation methods contain trace amounts of the true material that may be safely handled. Ad/bsorption and extraction methods, on the other hand, contain only the odor of the true material, rather than the material itself.

Further, the article discussed the available literature surrounding efficacy of these various forms of training aids. While true material is considered best practice for most situations, there may be challenges in the handling, transportation, storage, or the variety of available targets and odor profiles that can complicate the selection of true material. Because true material is not always available for logistical, safety, or regulatory reasons, alternative training aids can be of great value. However, these alternatives should be approached with care, as much of the discussed literature has cautioned.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

Funding for preparation of this review was provided by AGS Forensics, LLC and the Auburn University College of Veterinary Medicine Canine Performance Sciences Program through the generous support of Walt and Ginger Woltosz.

REFERENCES

1. *Crime Scene/Death Investigation – Dogs and Sensors Terms and Definitions*. 1st ed. ASB Technical Report 025, Dogs and Sensors Subcommittee of the Organization of Scientific Area Committee, Gaithersburg, MD (2017).
2. SWGDOG. *SWGDOG SC2 – General Guidelines. SWGDOG Approved Guidelines*. Miami, FL (2009).
3. DeGreeff LE, Malito MP, Brandon A, Katilie CJ. *Mixed Odor Delivery Device (MODD)*. US Patent No. US Government, US Secretary of Navy (2018).
4. Hughes TM. *Canine Scent Detection Training Device and Method*. US Patent No. US 10485219B1. US Government, US Secretary of Army (2019).
5. Maughn, MN. *Methods of Using Training Aid Delivery Devices (TADD)*. US Patent Pending US20170367298A1 (2017).
6. Moser AY, Bizo L, Brown WY. Olfactory generalization in detector dogs. *Animals.* (2019) 9:1–12. doi: 10.3390/ani9090702
7. Lazarowski L, Foster M, Gruen M, Sherman B, Fish R, Milgram N, et al. Olfactory discrimination and generalization of ammonium nitrate and structurally related odorants in Labrador retrievers. *Anim Cogn.* (2015) 18:1255–65. doi: 10.1007/s10071-015-0894-9
8. Hall NJ, Wynne CDL. Odor mixture training enhances dogs' olfactory detection of home-made explosive precursors. *Heliyon.* (2018)4:1–16. doi: 10.1016/j.heliyon.2018.e00947
9. Military Police. *Military Working Dogs. Army Regulation 190-12*. Washington, DC: Headquarters Department of the Army (2013).
10. DeGreeff LE, Peranich K, Simon A. *Detection of Ammonium Nitrate Variants by Canine: A Study of Generalization Between Like Substances*. NRL/MR/6181-−18-9791. Washington, DC: Naval Research Laboratory (2018).
11. Cablk ME, Szelagowski EE, Sagebiel JC. Characterization of the volatile organic compounds present in the headspace of the decomposing animal remains, and compared with human remains. *Forensic Sci Int.* (2012) 220:118–25. doi: 10.1016/j.forsciint.2012.02.007
12. Hoffman EM, Curran AM, Dulgerian N, Stockham RA, Eckenrode BA. Characterization of the volatile organic compounds present in the headspace of decomposing human remains. *Forensic Sci Int.* (2009) 186:6–13. doi: 10.1016/j.forsciint.2008.12.022
13. Swann LM, Forbes SL, Lewis SW. Analytical separations of mammalian decomposition products for forensic science: a review. *Anal Chim Acta.* (2010) 682:9–22. doi: 10.1016/j.aca.2010.09.052
14. Curran AM, Ramirez CF, Schoon AA, Furton KG. The frequency of occurance and discriminatory power of compounds found in human scent across a population determined by SPME-GC/MS. *J Chrom B.* (2007) 846:86–97. doi: 10.1016/j.jchromb.2006.08.039
15. Curran AM, Prada PA, Furton KG. Canine human scent identifications with post-blast debris collected from improvised explosive devices. *For Sci Int.* (2010) 199:103–8. doi: 10.1016/j.forsciint.2010.03.021
16. Prada PA, Curran AM, Furton KG. *Human Scent Evidence*, Boca Raton, FL: CRC Press (2015).
17. Forbes SL, Rust LT, Trebilcock K, Perrault KA, McGrath LT. Effect of age and storage conditions on the volatile organic compound profile of blood. *For Sci Med Pathol.* (2014) 10:570–82. doi: 10.1007/s12024-014-9610-3
18. Rendine M, Fiore C, Bertozzi G, De Carlo D, Filetti V, Fortarezza P, et al. Decomposing human blood: canine detection odor signature and volatile organic compounds. *J For Sci.* (2019) 64:587–92. doi: 10.1111/1556-4029.13901
19. Crespo Cajigas JM, Perez-Almodovar L, DeGreeff LE. Headspace analysis of potassium chlorate using on-fiber SPME derivatization coupled with GC/MS. *Talanta.* (2019) 205:120127. doi: 10.1016/j.talanta.2019.120127
20. Lazarowski L, Dorman D. Explosives detection by military working dogs: olfactory generalization from components to mixtures. *Appl Anim Behav Sci.* (2014) 151:84–93. doi: 10.1016/j.applanim.2013.11.010
21. Stevens NJ. Headspace methods and their application to inorganic explosives. In: *10th Annual Workshop on Trace Explosives Detection, Poster Presentation.* Ottawa, ON (2019).
22. Leggett DC, Jenkins TF, Murrman RP. *Composition of Vapors Evolved From Military TNT as Influence by Temperature, Solid Composition, Age, and Source*. CRREL Special Report (1997). p. 77–16.
23. Cerna K, Pinc L, Pachman J. Ability of explosives detector dogs to generalize odor of TNT. *New Trends Res Energ Mater Proc Semin.* (2011) 2:542–9.
24. DeGreeff LE, Cerreta MM, Katilie CJ. Variation in the headspace of bulk hexamethylene triperoxide diamine (HMTD) with time, environment, and formulation. *For Chem.* (2017) 4:41–40. doi: 10.1016/j.forc.2017.03.001
25. Steinkamp FL, DeGreeff LE, Collins GE, Rose-Pehrsson SL. Factors affecting the intramolecular decomposition of hexamethylene triperoxide diamine and implications for detection. *J Chrom A.* (2016) 1451:83–90. doi: 10.1016/j.chroma.2016.05.013
26. Harper RJ, Almirall JA, Furton KG. Identification of dominant odor chemicals emanating from explosives for use in developing optimal training aid combinations and mimics for canine detection. *Talanta.* (2005) 67:313–27. doi: 10.1016/j.talanta.2005.05.019
27. Macias MS, Harper RJ, Furton KG. A comparison of real versus simulated contraband VOCs for reliable detector dog training utilizing SPME-GC-MS. *Am Lab.* (2008) 40.1:16–9.

28. Rice S, Koziel J. Odor impact of volatiles emitted from marijuana, cocaine, heroin, and their surrogate scents. *Data Brief.* (2015) 5:653–706. doi: 10.1016/j.dib.2015.09.053
29. Huertas-Rivera AM. *Identification of Active Odors From Illicit Substances for the Development of Optimal Canine Training Aids* (FIU Electronic Theses and Dissertations). Florida International University, Miami, FL (2016).
30. Francis VS. *The Identification of Volatile Organic Compounds From Synthetic Cathinone Derivatives for the Development of Odor Mimic Training Aids* (FIU Electronic Theses and Dissertations) Florida International University, Miami, FL (2017).
31. Vu D. SPME/GC-MS characterization of volatiles associated with methamphetamine: Toward the development of a pseudomethamphetamine training material. *J For Sci.* (2011) 46.5:1014–24. doi: 10.1520/JFS15095J
32. Pinc L, Pachman J, Selesovsky J, Matyas R, Kloubek M. Ability of canines to detect sensitized nitro-methane. In: *International Conference on Forensic Research & Technology*, Vol. 3.8, Chicago, IL (2012).
33. Schoon A, Gotz S, Heuven M, Vogel M, Karst U. Training and testing explosive detection dogs in detecting triacetone triperoxide. *For Sci Commun*. (2006) 8.
34. Wright HF, Wilkinson A, Croxton RS, Graham DK, Harding RC, Hodkinson HL, et al. Animals can assign novel odors to a known category. *Sci Rep*. (2017) 7:9019. doi: 10.1038/s41598-017-09454-0
35. Jacobi KP. Cadaver detection in forensic anthropology and criminology: an overview with personal notes. In: Jezierski T, Ensminger J, Papet LE, editors. *Canine Olfaction Science and Law: Advances in Forensic Science, Medicine, Conservation, and Environmental Remediation.* Boca Raton, FL: CRC Press (2016). p. 245–57.
36. Knobel Z, Ueland M, Nizio KD, Patel D, Forbes SL. A comparison of human and pig decomposition rates and odour profiles in an Australian environment. *Aust J For Sci.* (2019) 51:557–72. doi: 10.1080/00450618.2018.1439100
37. Oxley JC, Smith JL, Porter M, McLennan L, Colizza K, Zeiri Y, et al. Synthesis and degradation of hexamethylene triperoxide diamine (HMTD). *Propell Explos Pyrotech*. (2016) 41:334–50. doi: 10.1002/prep.2015 00151
38. Francis VS, Holness HK, Furton KG. The ability of narcotic detection canines to detect illegal synthetic cathinones (bath salts). *Front Vet Sci*. (2019) 6:98. doi: 10.3389/fvets.2019.00098
39. Sonoda H, Konhoe S, Yamazato T, Satoh Y, Morizono G, Shikata K, et al. Colorectal cancer screening with odour material by canine scent detection. *Gut*. (2011) 60:814–9. doi: 10.1136/gut.2010.218305
40. Jezierski T. Detection of human cancer by dogs. In: Jezierski T, Ensminger J, Papet LE, editors. *Canine Olfaction Science and Law: Advances in Forensic Science, Medicine, Conservation, and Environmental Remediation.* Boca Raton, FL: CRC Press (2016).
41. Ensminger J. Medical alerting to seizures, glycemic changes, and migraines: significance of untrained behaviors in service dogs. In: Jezierski T, Ensminger J, Papet LE, editors. *Canine Olfaction Science and Law: Advances in Forensic Science, Medicine, Conservation, and Environmental Remediation*. Boca Raton, FL: CRC Press (2016). p. 389–402.
42. *United States Department of Agriculture's Animal and Plant Health Inspection Services*. National Detector Dog Training Center (2019). Available online at: https://www.aphis.usda.gov/aphis/ourfocus/planthealth/ppq-program-overview/nddtc (accessed January 1, 2020).
43. Lehnert MP, Weeks ENI. Trained dogs in insect detection. In: Jezierski T, Ensminger J, Papet LE, editors. *Canine Olfaction Science and Law: Advances in Forensic Science, Medicine, Conservation, and Environmental Remediation.* Boca Raton, FL: CRC Press (2016). p. 321–33.
44. Ensminger J. Canine biodetection in conservation, eradication, and border protection: a regulatory perspective. In: Jezierski T, Ensminger J, Papet LE, editors. *Canine Olfaction Science and Law: Advances in Forensic Science, Medicine, Conservation, and Environmental Remediation.* Boca Raton, FL: CRC Press (2016). p. 351–70.
45. Moser AY, Brown WY, Bizo LA, Andrew N, Taylor MK. Biosecurity dogs detect live insects after training with odour-proxy training aids: scent extract and dead specimens. *Chem Senses*. (2020) 45:179–86. doi: 10.1093/chemse/bjaa001
46. Lorenzo N, Wan TL, Harper RJ, Hsu YL, Chow M, Rose S, et al. Laboratory and field experiments used to identify Canis lupus var. familiaris active odor signature chemicals from drugs, explosives, and humans. *Anal Bioanal Chem*. (2003) 376:1212–24. doi: 10.1007/s00216-003-2018-7
47. Waggoner LP, Johnston JM, Williams M, Jackson J, Jones MH, Boussom T, et al. Canine olfactory sensitivity to cocaine hydrochloride and methyl benzoate. *Proc SPIE.* (1997) 2937:216–26. doi: 10.1117/12.266775
48. Curran AM, Rabin SI, Furton KG. Analysis of the uniqueness and persistence of human scent. *Forenisc Sci Commun.* (2005) 7.
49. Berns GS, Brooks AM, Spivak M. Scent of the familiar: An fMRI study of canine brain responses to familiar and unfamiliar human and dog odors. *Behav Process.* (2015) 110:37–46. doi: 10.1016/j.beproc.2014.02.011
50. Elliker KR, Sommerville BA, Broom DM, Neal DE, Armstrong S, Williams HC. Key considerations for the experimental training and evaluation of cancer odour detection dogs: lessons learnt from a double-blind, controlled trial of prostate cancer detection. *BMC Urol.* (2014) 14:22. doi: 10.1186/1471-2490-14-22
51. Zubedat S, Aga-Mizrachi S, Cymerblit-Sabba A, Shwartz J, Leon JF, Rozen S, et al. Human-animal interface: the effects of handler's stress on the performance of canines in an explosive detection task. *Appl Anim Behav Sci.* (2014) 158:69–75. doi: 10.1016/j.applanim.2014.05.004
52. Lit L, Schweitzer JB, Oberbauer AM. Handler beliefs affect scent detection dog outcomes. *Anim Cogn.* (2011) 14:387–94. doi: 10.1007/s10071-010-0373-2
53. Hallowell SF, Fischer DS, Brasher JD, Malone RL, Gresham GL, Rae C. Effectiveness of quality-control aids in verifying K-9 team explosive detection performance. *Proc SPIE*. (1997) 2937:227–34. doi: 10.1117/12.266782
54. Papet LE. Narcotic and explosive odors: volatile organic compounds as training aids for olfactory detection. In: Jezierski T, Ensminger J, Papet LE, editors. *Canine Olfaction Science and Law: Advances in Forensic Science, Medicine, Conservation, and Environmental Remediation.* Boca Raton, FL: CRC Press (2016). p. 265—78.
55. Bradshaw J. *K-9 Detection Training: Differences in the 4 Types Of Simulants*. (2013). Available online at: https://www.policeone.com/k-9/articles/k-9-detection-training-differences-in-the-4-types-of-simulants-XTtsL44atXAu4qhM/ (accessed January 10, 2020).
56. Pacik D, Plevova M, Urbanova L, Lackova Z, Strmiska V, Necas A, et al. Identification of sarcosine as a target molecule for the canine olfactory detection of prostate carcinoma. *Sci Rep*. (2018) 8. doi: 10.1038/s41598-018-23072-4
57. Kranz WD, Strange NA, Goodpaster JV. 'Fooling fido'—chemical and behavioral studies of pseudo-explosive canine training aids. *Anal Bioanal Chem*. (2014) 406:7818–25. doi: 10.1007/s00216-014-8240-7
58. Rice S, Koziel JA. The relationship between chemical concentration and odor activity value explains the inconsistency in making a comprehensive surrogate scent training tool representative of illicit drugs. *Forensic Sci Int.* (2015) 257:257–70. doi: 10.1016/j.forsciint.2015.08.027
59. Stadler S, Stefanuto P, Byer JD, Brokl M, Forbes S, Focant J. Analysis of synthetic canine training aids by comprehensive two-dimensional gas chromatography-time of flight mass spectrometry. *J Chrom A.* (2012) 1255:202–6. doi: 10.1016/j.chroma.2012.04.001
60. Tipple CA, Caldwell PT, Kile BM, Beussman DJ, Rushing B, Mitchell NJ, et al. Comprehensive characterization of commercially available canine training aids. *Forensic Sci Int.* (2014) 242:242–54. doi: 10.1016/j.forsciint.2014.06.033
61. Furton KG, Hsu Y, Luo T, Alvarez N, Lagos P. Novel sample preparation methods and field testing procedures used to determine the chemical basis of cocaine detection by canines. *Proc SPIE.* (1997) 2941:56–62. doi: 10.1117/12.266313
62. Furton KG, Hong Y, Hsu Y, Luo T, Rose S, Walton J. Identification of odor signature chemicals in cocaine using solid-phase microextraction-gas chromatography and detector-dog response to isolated compounds spiked on paper currency. *J Chrom Sci US.* (2002) 40:147–55. doi: 10.1093/chromsci/40.3.147
63. Cerreta MM, Furton KG. An assessment of detection canine alerts using flowers that release methyl benzoate, the cocaine odorant, and an evaluation of their behavior in terms of the VOCs produced. *For Sci Int.* (2015) 251:107–14. doi: 10.1016/j.forsciint.2015.03.021
64. Ensminger J, Ferguson MA. The practical and legal significance of the chemical analysis of odor in the relation to canine forensic and judicial evidence. In: Jezierski T, Ensminger J, Papet LE, editors. *Canine Olfaction*

Science and Law: Advances in Forensic Science, Medicine, Conservation, and Environmental Remediation. Boca Raton, FL: CRC Press (2016). p. 103–19.

65. Macias MS, Guerra-Diaz P, Almirall JA, Furton KG. Detection of piperanol emitted from polymer controlled odor mimic permeation systems utilizing *Canis familiaris* and solid phase microextraction-ion mobility spectrometry. *Forensic Sci Int*. (2010) 195:132–8. doi: 10.1016/j.forsciint.2009.12.006
66. Adebimpe DOBA. *Methods for Making Scent Simulants of Chemical Explosives, and Compositions Thereof*. US Patent No. 2009/0194744A1 (2009).
67. Vu DT. *Process for Producing Non-Detonable Training Aid Materials for Detecting Explosives*. US Patent No. 2014/0097551A1. Baltimore, MD: John Hopkins University (2014).
68. Creed G. What we should know about fentanyl. In: *Handler Instruction & Training Session*. Washington, DC (2018).
69. Simon AG, DeGreeff LE. Variation in the headspace of bulk hexamethylene triperoxide diamine (HMTD): Part II. Analysis of non-detonable canine training aids. *For Chem*. (2019) 13:1–7. doi: 10.1016/j.forc.2019.100155
70. Oxley JC, Smith JL, Canino JN. *Systems and Methods for Providing Non-Detonable Explosives or Explosive Stimulant Sources*. US Patent No. 9784723B1 (2017).
71. DeGreeff LE, Furton KG. Collection and identification of human remains volatiles by non-contact, dynamic airflow sampling and SPME-GC/MS using various sorbent materials. *Anal Bioanal Chem*. (2011) 401:1295–307. doi: 10.1007/s00216-011-5167-0
72. Eckenrode BA, Ramsey SA, Stockham RA, Van Berkel GJ, Asano KG, Wolf DA. Performance evaluation of the scent transfer unit™ (STU-100) for organic compound collection and release. *J For Sci*. (2006) 51:780–9. doi: 10.1111/j.1556-4029.2006.00178.x
73. DeGreeff LE, Curran AM, Furton KG. Evaluation of selected sorbent materials for the collection of volatile organic compounds related to human scent using non-contact sampling mode. *Forensic Sci Int*. (2011) 209:133–42. doi: 10.1016/j.forsciint.2011.01.010
74. Prada PA, Curran AM, Furton KG. The evaluation of human hand odor volatiles on various textiles: a comparison between contact and noncontact sampling methods. *J For Sci*. (2011) 56:866–81. doi: 10.1111/j.1556-4029.2011.01762.x
75. Sanchez CL, Huertas AM, Prada PA, Furton KG. A non-contact passive approach for the effective collection of target explosive volatiles for canine training aid development. *J For Sci Criminol*. (2016) 4:1–8. doi: 10.15744/2348-9804.4.205
76. Simon AG, Mills DK, Furton KG. Chemical and canine analysis as complimentary techniques for the identification of active odors of the invasive fungus, Raffaelea lauricola. *Talanta*. (2017) 168:320–8. doi: 10.1016/j.talanta.2017.03.015
77. Oxley JC, Smith JL, Moran J, Nelson K, Utley WE. Training dogs to detect triacetone triperoxide (TATP). *SPIE Proc*. (2004) 5403:349–53. doi: 10.1117/12.555791
78. MacCrehan W, Moore S, Schantz M. Reproducible vapor-time profiles using solid-phase microextraction with an externally sampled internal standard. *J Chrom A*. (2012) 1244:28–36. doi: 10.1016/j.chroma.2012.04.068
79. MacCrehan WA, Young M, Schantz MM. Measurements of vapor capture-and-release behavior of PDMS-based canine training aids for explosive odorants. *Forensic Chem*. (2018) 11:58–64. doi: 10.1016/j.forc.2018.09.002
80. Pfiester M, Koehler PG, Pereira RM, Ability of bed bug-detecting canines to locate live bed bugs and viable bed bug eggs. *J Econ Entomol*. (2008) 101:1389–96. doi: 10.1603/0022-0493(2008)101[1389:aobbct]2.0.co;2
81. Porritt F, Mansson R, Berry A, Cook N, Sibbald N, Nicklin S. Validation of a short odour discrimination test for working dogs. *Appl Anim Behav Sci*. (2015) 165:133–42. doi: 10.1016/j.applanim.2014.11.021
82. Johnen D, Heuwieser W, Fischer-Tenhagen C. An approach to identify bias in scent detection dog testing. *Appl Anim Behav Sci*. (2017) 189:1–12. doi: 10.1016/j.applanim.2017.01.001

Conflict of Interest: AS was employed by the company AGS Forensics, LLC. KG was employed by the company Giles Consulting, LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with one of the authors PW.

frontiers in Veterinary Science

REVIEW
published: 17 July 2020
doi: 10.3389/fvets.2020.00408

Methodological Considerations in Canine Olfactory Detection Research

Lucia Lazarowski[1]*, *Sarah Krichbaum*[1,2], *Lauryn E. DeGreeff*[3], *Alison Simon*[4], *Melissa Singletary*[1,5], *Craig Angle*[1] and *L. Paul Waggoner*[1]

[1] Canine Performance Sciences Program, College of Veterinary Medicine, Auburn University, Auburn, AL, United States, [2] Department of Psychological Sciences, College of Liberal Arts, Auburn University, Auburn, AL, United States, [3] Chemistry Division, U.S. Naval Research Laboratory, Washington, DC, United States, [4] AGS Forensics, LLC, Washington, DC, United States, [5] Department of Anatomy, Physiology, and Pharmacology, College of Veterinary Medicine, Auburn University, Auburn, AL, United States

Dogs are increasingly used in a wide range of detection tasks including explosives, narcotics, medical, and wildlife detection. Research on detection dog performance is important to understand olfactory capabilities, behavioral characteristics, improve training, expand deployment practices, and advance applied canine technologies. As such, it is important to understand the influence of specific variables on the quantification of detection dog performance such as test design, experimental controls, odor characteristics, and statistical analysis. Methods for testing canine scent detection vary influencing the outcome metrics of performance and the validity of results. Operators, management teams, policy makers, and law enforcement rely on scientific data to make decisions, design policies, and advance canine technologies. A lack of scientific information and standardized protocols in the detector dog industry adds difficulty and inaccuracies when making informed decisions about capability, vulnerability, and risk analysis. Therefore, the aim of this review is to highlight important methodological issues and expand on considerations for conducting scientifically valid detection dog research.

Keywords: detection dogs, olfaction, scent detection, canine olfaction testing, animal behavior

OPEN ACCESS

Edited by:
Nathaniel James Hall, Texas Tech University, United States

Reviewed by:
Astrid R. Concha, Army Research Office, United States
Janice Lauren Baker, Veterinary Tactical Group, United States

****Correspondence:***
Lucia Lazarowski
lucia.lazarowski@auburn.edu

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences, a section of the journal Frontiers in Veterinary Science

Received: *15 February 2020*
Accepted: *08 June 2020*
Published: *17 July 2020*

Citation:
Lazarowski L, Krichbaum S, DeGreeff LE, Simon A, Singletary M, Angle C and Waggoner LP (2020) Methodological Considerations in Canine Olfactory Detection Research. Front. Vet. Sci. 7:408. doi: 10.3389/fvets.2020.00408

INTRODUCTION

Dogs' superior olfactory abilities and high trainability are leveraged for a wide range of chemical and biological detection applications. As the scope of detection dog applications continues to grow, understanding detection dog olfactory capabilities and factors affecting performance is critical for improving training and deployment practices. However, methods for testing canine olfactory detection vary widely and such variation can influence the interpretation of results. Further, systematic reviews of canine olfactory detection literature have identified a major lack in reporting the information necessary to evaluate the validity of the results (1), as well as a prevalence of methodological confounds that could bias their interpretation (2). In contrast, analytical instrumentation undergoes rigorous validation standards prior to use in controlled and narrowly defined field operations. Here, we review the various critical features that should be included in the design and implementation of olfactory detection studies in order to ensure the quality and reproducibility of results. We expand upon the issues highlighted by Johnen et al. (2) to address considerations related to subject characteristics, experimental design, statistical analyses and reporting, and odor characteristics. Aspects of internal validity, or the extent to which results show evidence in support of what they claim, as well as external validity, referring to the ability of the results to generalize to populations, stimuli and environments other than those tested (3), will be discussed. The issues presented here should also be relevant for evaluating operational canine performance, for which there is also a lack of standardized protocols. While we argue for increased rigor in the examination of canine odor detection performance, the inherent variability of any biological system and technical challenges in its assessment across its wide operational field must be considered.

SUBJECT CHARACTERISTICS

Sensory Differences

Individual or breed-related differences in sensory and morphological traits can influence performance in olfactory detection tasks and thus the generalizability of results. These differences are especially important when examining performance capacities that are likely influenced by physical characteristics, such as threshold of sensitivity to low concentrations of odor. For example, genetic differences in olfactory receptor repertoire or anatomical variations in ear and nose shape may influence olfactory acuity (4). Bloodhounds possess 300 million odor receptor cells, more than any other breed (4), and are considered to have the greatest olfactory acuity (5). However, studies on canine olfactory threshold have been limited to single breeds (6–8) or few dogs from several breeds (9), with few studies examining breed as a factor in olfactory detection performance. In one study, differences in olfactory threshold were examined by comparing different breeds in a natural detection task in which food was hidden in containers of varying levels of permeability (10). Dogs from breeds selected for scenting abilities, such as hounds and beagles, exhibited a greater sensitivity than dogs from non-scenting breeds (e.g., grayhounds). Further, brachycephalic breeds showed the least sensitivity in detecting the odor compared to non-brachycephalic breeds. These findings suggest that differences in structure and function of the olfactory system may influence performance in an odor detection task, which could have implications for extrapolating results to other populations of dogs. However, whether the breed differences found in this study were specific to the nature of the target (i.e., raw meat) and would also be observed for a trained artificial target is unknown. Olfactory sensitivity and function is also influenced by a range of other factors including age, disease, medications, hydration, and diet [see (11) for review].

Behavioral Differences

Contrary to evidence that breeding for scenting abilities and elongated noses is associated with better scent detection performance (10, 12), Hall et al. (13) found that pugs outperformed German shepherds in learning and performing a simple odor discrimination across decreasing concentrations of an odorant (13). This result is surprising considering the anatomical differences between these breeds and the popularity of German shepherds in scent detection work. Thus, these findings imply that other factors may influence performance on an odor detection task such as athleticism and behavioral differences, which may vary depending on the nature of the task. Indeed, German shepherds were not originally bred for odor detection tasks, but for herding and guarding sheep. Rather, the use of German shepherds for contemporary roles in the security sector is due to a combination of attributes such as athleticism, desire to work, and trainability necessary for multi-purpose work. Thus, had a more complex or strenuous task such as a search in an operational environment been used, German shepherds may have performed better than pugs in the Hall et al. (13) study. The importance of behavioral traits is further reflected by the fact that, despite their superior olfactory acuity, scent hounds are rarely used in olfactory detection research due to their poor trainability (14, 15). Other differences in olfactory search patterns, such as the tendency for nose-to-ground tracking vs. air-scenting, can influence performance depending on the type of detection task. An advantage of air-scenting is the ability to cover a wider range of search area in a shorter amount of time and to more efficiently locate targets using air currents (16). It is also important to recognize individual differences in motivation of the dog, as a lack of motivation to learn or complete a task could negatively influence the results. However, it is imperative to select an appropriate reward by using one with which the dog has experience or by conducting a reward preference test prior to the experiment, and to consider potential effects of reward value (i.e., highly preferred vs. less preferred) on performance (17). In addition, an easy warm-up trial prior to the session will ensure that the dog is willing to work for the chosen reward (18).

Subject Selection

The sample of subjects selected may also influence the validity of the results obtained. In addition to differences in olfactory acuity or search behavior, training history can greatly influence detection performance. For example, experience with a particular odorant can affect sensitivity to that odor or generalization to other odors (5). Dogs specifically trained for scent detection are also more likely to perform better than novice dogs in search-based tasks. Thus, as acknowledged by the authors, the counterintuitive findings between pugs and German shepherds in the study by Hall et al. (13), which used privately owned pets, may have differed if purpose-bred or professionally trained detection dog German shepherds had been used. In applied research, it is sometimes imperative that the dogs used are representative of operational dogs for translation of the results to field applications. However, a potential concern when conducting research with operational detection dogs is that participation in the study could interfere with the dogs' operational performance. Recent studies have utilized privately owned pet dogs trained in sport detection (19, 20), which may represent a more practical model allowing for larger sample sizes and relevant experience. On the other hand, studies involving recruitment of pets may suffer from a sampling bias in which owners who volunteer their pets for behavioral studies may be more likely to engage their dogs in training and seek enriching activities. Similarly, studies using random source populations (e.g., shelter dogs) could introduce potential confounds related to the dog's experience, which is often unknown. Thus, the subject sample tested should always be taken into consideration when interpreting results, and efforts should be made to replicate and validate results in diverse populations and/or targeted to operationally relevant samples of dogs as laboratory derived results do not always directly correlate to the performance of operational teams as the subject population, behavioral requirements, target variables, environmental elements, and the canine handlers may not be the same.

EXPERIMENTAL DESIGN

Sample Size

A further limitation related to sample selection is the number of dog subjects tested. Due to challenges in availability and access to dogs for extended periods of time needed for training and testing, the majority of dog studies utilize a small number of subjects. Although adequate for proof-of-concept experiments aimed at identifying a given capability, studies with few subjects complicate data analysis in more sophisticated experimental designs and limits the external validity of the results.

A recent examination of published dog studies evaluated the influence of sample size on effect sizes (i.e., the strength of the relationship between variables) and statistical power (i.e., the ability of the chosen statistical test to identify possible relationships between variables) (21), and found that the majority of dog studies were vastly underpowered and had low effect sizes due to low sample sizes (22). For example, the median number of subjects was 16, and the power produced by these studies was nearly zero. Statistically, a larger number of subjects allows for outcome-sensitive testing, meaning that the results are externally valid and highly replicable (22). Increasing the sample size is particularly important for group or matched-pairs designs due to variability between groups that could affect statistical power. When group designs are used, and especially when group sizes are small, all attempts should be made to equate the groups in terms of dog experience and capability. In cases in which increasing the sample size is impossible, researchers can maintain some level of external validity by increasing the number of trials and emphasizing individual differences (22). *A priori* power analyses can be used to determine the number of subjects needed in order to produce a desired effect size with narrow confidence intervals.

Measuring Accuracy

A major goal of evaluating canine olfactory detection performance is to determine dogs' ability to correctly discriminate target odors from non-target odors. Accuracy of detection ability utilizes metrics utilized for medical diagnostics or analytical instruments, and is based on *sensitivity,* the probability of a response to a target odor when that target odor is present, and *specificity*, the proportion of non-targets correctly ignored (23). For example, studies of medical detection dogs' ability to detect certain diseases compare sensitivity and specificity between positive samples and controls, and can be used to compare dogs' performance to the best available gold standard for diagnostic technology [see (24) for review].

These metrics should be considered in tandem, as high sensitivity is meaningless if specificity is low (meaning dogs detect all targets, but also respond to non-targets), and a high level of specificity is not valuable if targets are also not detected. A low degree of sensitivity could be the difference between life and death for explosives or medical detection, and a low degree of specificity could lead to unnecessary response measures or anxiety (25, 26). A comprehensive assessment of canine olfactory detection accuracy then typically utilizes a signal detection theory approach, recording true positives (hits), true negatives (correct rejections), false positives (false alarms), and false negatives (misses) (27). The most commonly reported measures of performance include hit rates, calculated as the number of hits out of the total number of target exposures, and false alarm rate, calculated as the number of false alarms out of the total number of opportunities for a response (or conversely, correct rejection rate) (28, 29). Some metrics combine both sensitivity and specificity in order to measure overall accuracy, such as proportion of correct responses (hits and correction rejections) out of the sum of all responses (hits, correct rejections, false alarms, and misses) (30). A number of other metrics are also sometimes calculated depending on the measure of interest, such as positive predictive value (PPV) as a measure of how frequently a dog's alert is a correct one (20), false discovery rate (FDR; proportion of responses that are incorrect, or 1-PPV), and other variations.

Types of Tasks

Discrete trials

Initial validation of odor recognition is typically measured as a dogs' ability to discriminate target odors from non-target odors (23). This is often achieved using controlled set-ups in which dogs are presented with a fixed number of positions to sample from which may contain targets or non-targets. Common testing arrangements include radial arm carousels or odor sampling arrays arranged in a circle or line, and dogs are trained to sample from each position in the array. Because dogs are presented with samples one at a time, these types of tasks are analogous to the "go/no go" task widely used in behavioral research in which an independent decision (yes/no) is required for each stimulus sampled. Thus, these types of tasks are considered discrete trial procedures because they consist of isolated opportunities to make a single response to a given stimulus, whether a target or non-target odor (31). Because of the precise control over the presentation of both targets and non-targets and subsequent responses (or lack of) to each, comprehensive performance metrics can be calculated. For example, in order to calculate true false alarm rate, the proportion of correct rejections of non-targets encountered must be known. In a discrete trials fixed sampling task, the number of positions containing non-targets that the dog checks and does not respond to prior to encountering the target can be counted.

Such procedures are common in canine olfactory detection research as well as in accreditation or proficiency testing of trained detection dogs, as they are easy to standardize allowing for comparisons across dogs or groups. For example, the Bureau of Alcohol, Tobacco, Firearms, and Explosives (ATF) developed the National Odor Recognition Test for proficiency testing of explosives detection dogs using a fixed-sampling circular arrangement (i.e., paint cans arranged in a circle) with defined testing parameters, thus allowing for a uniform assessment of dogs across varying agencies and organizations. A tool developed by Porritt et al. (23) automatically generates a test design with balanced order and number of trials, plus number and placement of distractors allowing for standardized comparisons and is available online for free to practitioners and researchers (23, 32).

A challenge to procedures requiring a decision response for each sample is the response inhibition required for a correct rejection. Olfactory go/no go studies in rats suggest there is significant difficulty learning to inhibit a response when presented with a non-target odor (33). Response inhibition is considered a form of self-control, which may be reduced in populations of working dogs bred for high energy levels and exhibiting higher levels of impulsivity (34). Researchers have suggested that refraining from making any response introduces needless difficulty which can be mitigated by training differential responding (2, 18, 33). For example, Edwards (28) trained dogs to hold their nose in a port emitting a target odor to indicate a "yes" response, and to remove the nose and push a lever for a rejection response (28). In multiple-choice arrays, a rejection response could simply be leaving the positions and moving on to the next one.

Another obstacle in using multiple-alternative arrays is the tendency for dogs to develop positional biases. For example, when the same positions are used and re-used within a training session, dogs have been shown to defer to responding to a particular position that was more recently or more frequently rewarded (35). This type of bias is more common early in training when the dog is not proficient in detecting the target, and should minimize as dogs' confidence in detecting the odor increases. Thus, it is best to begin training with few positions, increasing number of positions as dogs' proficiency increases, so as to reduce the cognitive demands of the task (14, 36). The position in which the target odor is placed should then be randomized so that patterns in placement that dogs' could learn are minimized. However, researchers have cautioned that a risk of full randomization is that for dogs already exhibiting a positional bias, randomization could lead to targets being placed in positions that the dog already preferred, thus reinforcing the positional bias (37). Thus, Jezierski et al. (37) suggested quasi-randomization in which the position of the target on each trial is tailored to the dogs' training deficiencies, which should be corrected before testing (with full randomization) begins (37). Another strategy for reducing positional biases is counterbalancing, in which each position contains the target an equal number of times across the session. An advantage of counterbalancing is that potential positional biases can be detected because each position is used equally, and therefore false responses or misses should be equally distributed across all positions unless a bias has developed. A common positional bias that has been reported is the tendency for dogs to emit a false alert in the last position of the array due to an increasing likelihood that a given position contains a target as the number of positions checked increases (2), or because the last position represents that final potential opportunity for reinforcement (32) [though the opposite pattern is observed in scent line-up tasks due to the memory component and increased delay the further away the targets are placed (14)]. Johnen et al. (2) suggest several strategies to remedy this last-position bias such as using a circular array with no discernible start or end point, and making the number of potential targets per trial variable and unpredictable, either by training dogs that an array can contain more than one target (and thus each position has a 50% chance of containing a target) or that it may contain no targets (i.e., blank trials) (2). Rewarding a correct rejection of a blank trial, for example by training dogs to perform a specific response if no targets are detected, can also reduce the tendency for false alerts (38).

Search Tasks

Discrete trial procedures may not be applicable to answering questions about operational search performance. Higher level and more complex skills related to search technique and ability are required during operational searches. Therefore, in order to assess detection dogs' ability to conduct a safe and systematic search in an operationally relevant manner as well as to validate initial odor discrimination testing, free searches are typically also employed (23). Free search tasks do not have defined opportunities for responses as in discrete trial procedures, but are used to assess other aspects of performance beyond basic odor recognition, such as dogs' ability to detect target odors in a complex environment in the absence of ostensive cues and independently follow the odor trail to its source. However, the increased complexity of the task makes evaluating performance in a free search scenario challenging in regard to standardization (2). One major limitation of evaluating performance during a free search is the infinite number of potential odor sources, and thus the inability to accurately calculate correct rejections needed to calculate true false alarm rates or specificity. Instead of calculating false alarms as a proportion of total opportunities for a response, one could calculate the proportion of false alarms as a proportion of the number of distractors placed in the search area. However, it is not guaranteed that the dog will necessarily encounter each distractor placed. Alternatively, PPV could be calculated which is the proportion of a dog's total responses (both hits and false alerts combined) that are correct. The higher the number of false alerts, the lower the PPV will be.

An additional layer of variability in dogs' performance in a free search task often results from handler error rather than errors by the dog. One commonly reported source of handler error is handler-induced false alerts (39, 40). Handlers may also cause dogs to miss targets by failing to ensure or inhibit the dog from adequately searching the area containing the target, making it difficult to differentiate whether a failure to alert to an odor (i.e., a miss) is due to a detection failure by the dog or handler error. For example, impatient handlers may rush the dog through a search (31), or may conduct an inadequate search pattern preventing the dog from having the opportunity to locate the target (41). In this case, it can be argued that the dog was not presented with the odor, which should not count as a true miss (though the distinction may not be as important for operational certifications, for example, where the handler performance is an important aspect of the team's ability). In cases where the detectability of the odor by the dog is of interest, researchers have addressed this challenge in a number of ways. Engeman et al. (41) utilized an inconspicuous observer to record whether the handler positioned the dog in a way that it was likely to detect planted targets in order to categorize missed targets as handler error or failure by the dog to give an alert (41). Another strategy by Porritt et al. (42) utilized "vigilance points," in which specified locations unknown to handlers throughout a search area were identified to allow for data collection on whether or not the dog checked those locations (42).

There are several other ways in which handler factors can influence dogs' performance. Experience training and working with dogs as well as general familiarity with the dogs being tested can affect task acquisition (2) and interpretation of indication behaviors (12). Handler stress can also influence performance; for example, Jezierski et al. (12) reported dogs having longer detection times and more false alarms during formal certifications compared to informal examinations, which the authors attributed to handler anxiety due to the formality and pressure of the certification test (12). Schoon (18) reported that handlers' confidence was influenced by how the dog performed on a given trial, which could have influenced performance on subsequent trials (18). For example, if a handler calls an alert that turns out to be a false response, the handler may be hesitant to call subsequent responses by the dog. Dogs' performance has also been shown to be influenced by how familiar they are with the handler, with detection performance decreasing when working with an unfamiliar handler (43). On the other hand, Zubedat et al. (44) found that increased handler anxiety actually improved dogs' latency to detect targets (44). Interestingly, the authors suggested that handler stress led to a decrease in their control over the dog, thereby reducing handler influence and allowing dogs to work more independently. Further research is needed to directly examine handler effects, such as comparing dogs' performance when working on- and off-leash. Researchers should attempt to control for handler effects whenever possible by utilizing well-defined testing protocols, assessing both dog and handler performance to disentangle such variables, utilizing professional trainers, keeping the trainer/handler consistent, and keeping evaluators blind (discussed below).

Finally, detection of odors and search performance can be influenced by environmental factors such as temperature, humidity, air flow, and terrain (45). Hiding places for targets are also more variable in free searches, and odor availability can be influenced by depth, height, and containerization of the target odor. Placements of targets and non-targets should therefore be randomly distributed but matched in terms of level of difficulty or accessibility of target odors. Despite challenges in standardizing free searches, assessing performance in situations resembling real-world operations is critical for defining detection dog capabilities.

Human Bias

Types of Bias

When a human handler or observer partakes in the administration or evaluation of canine testing, there are several potential sources of bias that can affect performance. One well-established form of bias in behavioral testing is known as the experimenter expectancy effect, in which observer expectations influence the subjects' behavior. This effect was famously illustrated by the classic example of the horse *Clever Hans*, believed to be capable of counting by stomping his foot a certain number of times in response to mathematical queries. In reality, the horse had learned to respond to unintentional cues by the people observing him, who exhibited subtle changes in body language and facial expressions as he approached the correct answer. Dogs are especially skilled at detecting subtle and unconscious cues given by humans. For example, dogs will often follow human cues that contradict available perceptual information (46, 47), and handler expectations about the presence or absence of a target odor can influence the team's accuracy (40). Allowing dogs to work off-leash can minimize handler influence, but does not remove all cues. Often unintentionally, handlers or observers may move more quickly past a search area known or expected to not contain any targets (19), or linger longer or pay greater attention in an area known to contain targets, which can provide strong cues to the dog regarding the probability of encountering a target. Edwards et al. (24) suggest that other unintentional cues given by experimenters or evaluators could influence performance as dogs are highly sensitive to human hand signals (48), body orientation (49), and emotional content of facial expressions and speech (24, 50). Methods for reducing observer influence in a search task have included positioning the observer on a designated mark on the floor, as well as requiring the observer to score whether or not the dog searched each target and non-target placed in the area so that the observer's attention to targets and non-targets was equal (23).

Even when care is taken to minimize potential cueing, knowledge or expectations held by the observer can influence the interpretation of the dog's behavior. Forms of observer bias are widely acknowledged in animal behavior research, such as selectively attending to information that confirms hypotheses or being susceptible toward certain beliefs based on prior knowledge. For example, observers scoring videos of animal behavior scored the same video differently depending on false information they were given about the animals or context of the video (51). Observer bias readily occurs when observers have a vested interest in the hypotheses or outcomes, when the behavior under observation is ambiguous, and when the interpretation of the behavior is subjective (51). Thus, there is risk of observer bias both in canine olfactory detection research in which investigators have expectations based on hypotheses, and in operational canine assessments when handlers and trainers have a vested interest in the dogs' success, and often occurs unintentionally.

Observer bias is inherent in canine testing due to the subjectivity of the behavior under observation. Dogs are trained to indicate an alert using a variety of responses such as sitting, lying down, and freezing, all of which require a certain degree of subjective interpretation. For example, whether a dog fully sat, and the duration of the sit, can lead to ambiguity in interpreting whether a response was made or not. Observer bias is particularly confounding when the response is ambiguous, as assumptions about whether or not the response is correct can influence its interpretation. Further, handlers may differ in how conservative their interpretation of a response is (28). This can complicate scoring when the dog's response is in conflict with the handler's interpretation; for example, a handler may believe a response to be incorrect and calls it a false, when the dog was actually correct. In this case, a decision has to be made whether to side with the handler or the dog. Whenever possible, an operational definition of a response defining the topography and duration of the behavior is critical, and should allow for different observers to come to the same conclusion (28). However, reliable agreement between observers does not necessarily eliminate observer bias if the two observers hold the same beliefs, as demonstrated by Tuyttens et al. (51) finding that the highest inter-rater reliability also had the highest degree of observer bias (51).

Minimizing Bias

Due to the many sources of human-derived bias in canine performance, blinding of personnel is critical. Single-blind testing is the most common procedure used in canine testing, where the handler is unaware of the test conditions (e.g., presence or location of targets) and an observer sets up the test problem and informs the handler of the outcome. This requires an ability to prevent visual identification of the target, as handlers can visually identify targets and voids the purpose of blinding. This type of protocol is often preferred by operational teams so that the handler can deliver timely feedback to the dog (e.g., reinforcement for a correct response) in order to maintain performance. Because the handler is unaware of the test conditions, handler influence on the dog's behavior is reduced. However, blinding the handler does not remove all sources of potential influence, especially when the evaluator is present and is not blind (52). The evaluator's behavior can provide strong cues as to the presence and location of target odors not only to the dog, but to the handler as well. In true operational situations, no one will know where the target is located. Thus, double-blind testing in which none of the participants or observers present in the test area are aware of the trial conditions (i.e., presence or location of targets) is the only assessment that truly reflects real-world operations (53). These situations also need to be mimicked in scientific studies whenever possible so the data can be directly correlated to operational performance, and in training to better prepare dogs for real-world scenarios. Indeed, studies reporting a decline in dogs' performance once double-blind testing is implemented underscore the importance of applying these procedures in research practice as well as in operations (54, 55).

A common solution to minimizing human influence is to position the handler and other observers in a way that they can view the dog but the dog cannot see them, such as behind a screen or one-way mirror (9). However, Edwards et al. (24) caution that removing visual cues is not always sufficient and other cues (e.g., auditory) may still be available (24). For example, the dog may learn to associate the sound of a pocket opening in anticipation of delivering a reward, or observers becoming quiet or holding their breath as the dog approaches the target location.

Double-blind testing is considered the gold standard in animal behavior research as it can minimize both observer bias and observer influence; however, double-blind testing is less commonly used due to challenges in its implementation. For example, one approach to double-blind canine testing is for both the handler and evaluator to be blind, and the handler calls out the dog's responses while the observer records the information. Once the trial is completed, someone who was not present during testing but knows the details of the scenario reviews the recorded responses and scores the dog's performance. This approach is often undesirable because accurate feedback to the dog's responses is not possible and non-differential reinforcement must be used (i.e., the outcome for correct and incorrect responses is the same). One option is to reward all of the dog's responses so that correct responses do not go unrewarded, but the risk is that false alarms can increase if incorrect responses are rewarded; the alternative is to withhold reinforcement of all responses, with the risk of performance or motivation declining (e.g., extinction). To prepare dogs for this type of testing, intermittent reinforcement schedules are often introduced in which reinforcement of correct responses is gradually faded so that some but not all correct responses are rewarded, resulting in behavior that is highly resistant to extinction (56). Accustoming dogs to intermittent reinforcement is especially important for preparing dogs for operational conditions in which reinforcing indications is not possible, such as in medical detection when the status of a sample is unknown (24). Another approach utilizing double-blind testing that allows for reinforcement of correct responses is for the blinded handler to announce when the dog makes a response, which is then confirmed by a third-party who is removed from the test situation (24). For example, in a study by Johnen et al. (1) the handler called out the number of the position where the dog responded, and then an experimenter out of view confirmed the response (1). This can be achieved by having the observer behind a screen or one way mirror, watching on a monitor connected to a video camera, or using a mobile device to communicate with the handler. Other systems have been used that do not require this type of relaying from one individual to another, such as custom-made software in which the handler presses a key to reveal the result (9). In all of these instances, a slight delay will be imposed between the dog's response and its reward, which can be introduced in training until the dog is accustomed to the delay.

To minimize subjectivity and increase the accuracy and reliability of testing, some researchers have devised automated approaches to data collection. For example, sensors that automatically detect a response by requiring breaking an infrared beam for a pre-determined amount of time reduce subjectivity in interpreting and recording whether a response has been made. Edwards (28) built a carousel apparatus for canine scent detection testing that automated all aspects of testing related to stimulus presentation, response recording, and reinforcement delivery (28). Infrared beams detected whether the dog observed a sample by requiring a minimum sniff time, ensuring that each sample was observed. Breaking the infrared beam for a longer pre-determined amount of time recorded an indication response, and correct responses were automatically reinforced via a feeder. Dogs were also trained to give a "no" response by pushing a lever which advanced the carousel to the next sample given that the minimum observation response criteria had been met, allowing for precise calculations of correct rejections. Though this type of system is more costly and requires significant training to teach the dog to operate making it impractical for operational assessments, research requiring precise control over stimuli and observations may benefit from automated systems such as this.

An argument to be made for non-automated response detection is that canine behavioral responses, especially in challenging or ambiguous situations, can sometimes be nuanced requiring a subjective but expertise-based interpretation. An example of this is a characteristic "change of behavior" (COB) interpreted by the handler as indicative of recognition of a target odor (57, 58). COB is often considered essential by handlers in declaring whether a dog has detected a target as the COB is considered a reflexive-like response to the conditioned odor that is not as encumbered by ancillary influences as the trained alert response (e.g., not being able to identify or access the exact odor source, or training deficiencies in performing the operant response). Because of this, a COB is often enough for the handler to declare a target. Further, in actual operations a COB is typically considered enough to prompt a threat response because waiting for a trained final response could be costly. Due to the high degree of subjectivity required in interpreting a COB, this metric should only be used when the handler is blind as observers are more likely to identify a COB when they are aware of the target location and not identify a COB in response to a non-target odor. Accordingly, COBs called to non-targets should be documented as a type of false response. For COB to be a meaningful measure of the dogs' response to a given target, the rate should be higher than COB to non-targets. Further research is needed to examine the specific behaviors accompanying a COB which may aid in standardizing the response as an acceptable metric as well as training observers to identify the response. For example, if the COB is truly a conditioned response elicited by a conditioned stimulus (the odor, which has been previously paired with an unconditioned stimulus), we may predict that the COB resembles behavior that is anticipatory of a reward, such as an orienting response (e.g., raised ears, looking toward stimulus) or approach (59).

Odor Sample Controls

Positive Controls

Assessing canine olfactory detection performance requires constant scrutiny of extraneous variables by which olfactory behavior could be influenced. Positive controls are used to evaluate test validity, which, in the case of canine detection, ensure that dogs are responding to the target samples on the basis of the target odor. Positive controls involve the presentation of targets free from potential sources of contamination (60), such as new or refreshed samples, samples obtained from different sources (e.g., a different manufacturer or brand), preparation by a different person, or presenting the targets in new containers. A lack of responses to the positive controls suggests that dogs were responding to some other cue, such as contamination of the training sample that could occur from overuse, scent of the person that handled the odor, or the packaging material (2). Thus, positive controls are a necessary step for validating that dogs are capable of responding to the odor which they were trained to detect.

Positive controls are also useful to include during training to facilitate learning of the intended target. The use of a large number of positive controls has been shown to be especially critical in medical detection dog training as dogs have been shown to memorize the samples from individual people rather than the common odor (e.g., the disease). For example, Elliker et al. (55) found that the performance of dogs trained to detect cancer samples dropped when samples from new patients were introduced, indicating that the high accuracy observed to the training odors was due to memorization of the individual samples rather than the common odor profile (55). Training dogs to respond to odors based on some common classification essentially requires dogs to learn a concept, where the concept is the particular disease (e.g., cancer) or explosives class (e.g., chlorates) (24). Because dogs can readily learn and memorize a large number of individual odors within and across test sessions (61), it is recommended that training utilize a large set of samples. The larger the number of training samples, the more difficult it becomes to memorize individual samples and learning the concept common to all of the samples becomes a more efficient strategy. This "set-size effect" in which concept learning increases as a function of the number of training samples is a well-established phenomenon demonstrated in a range of species (62). The same principle applies to olfactory learning, where exclusively training with a particular odor or odor concentration tends to reduce the tendency to generalize to other variants or concentrations (63, 64). However, just as training with a fixed target can narrow the tendency to generalize to other variations of the target, training with a range of variants can enhance generalization (65). Research with detection dogs has demonstrated that the more that irrelevant factors are varied in training, such as source (66) and composition (29, 67–69), the more likely the dog is to generalize to other variants of the trained target. Thus, best practices for maximizing optimal generalization to potential variations of a trained target are to train with many exemplars of a target that vary by irrelevant dimensions (5, 70).

Negative Controls

Distractor odors used as *negative controls*, also referred to as interferents (23), consist of non-target odors and are equally critical in evaluating dogs' detection performance in terms of calculating specificity. The use of distractors is also important during training for teaching dogs to discriminate the target odors from non-target odors (e.g., discrimination training). For example, Elliker et al. (55) speculated that early training with only the target odors may have biased dogs toward memorization of the samples, and suggests that teaching dogs to disregard the controls by never presenting the target samples alone may be a better approach (55). Distractors should include odors that are similar to the target in terms of intensity, otherwise dogs could learn to differentiate the target odor based on its relative (higher or lower) strength, and should include odors from similar and differing odor categories (e.g., chemical, biological). Distractors should consist of odors commonly associated with training, the training environment (e.g., reward odors, handler/trainer odors), target containers (e.g., nylon bags), and preparation (e.g., gloves, pipettes), as these odors are likely to become associated with the training odor. For example, a preliminary study found that dogs trained to low levels of an explosive were actually detecting residue plasticizers from the pipettes used in preparation procedures (71). During search tests there are likely to be a variety of items present in the environment, but these items are likely acclimated to the environment which the dog may learn to ignore. Thus, other items should also be added and trainers and observers should touch various items in the area in order to introduce odors associated with human activity/disturbance similar to the disturbance that will be created by planting the target odor (72).

Items that systematically covary with the presentation of the target itself should also be presented as distractors. Sometimes referred to as *matched controls* or *matched blanks*, these distractors are designed to match the background odors that coincide with the target sample (73). In substance detection, matched controls should consist of empty and clean packaging materials and containers identical to those used to store or present the target odor, as well as gloves used to prepare and handle targets. In medical detection, matched controls consist of samples from patients of the same age and sex, as well as samples from patients with ailments different than the target disease but affecting the same organ so that samples are as comparable as possible and only differ by the specific disease status, eliminating other factors that covary with the disease (24). The use of matched controls during training helps dogs isolate the target odor, and during testing ensures dogs are responding to the target odor only.

Distractors should also include novel odors, particularly when testing for generalization to target odors that were not used in training. This is important because in such testing, the target odors will be novel to the dog. In order to ensure that any responses to the test odors are due to generalization based on the target odor and not due to responding to the anomalous odor, other odors that are novel should be present. Because dogs tend to be neophilic (74), disruption of performance during testing can be prevented with adequate discrimination training in which novel distractors are introduced early in training and are gradually faded in so that dogs learn that novel odors may be present at any time but do not learn any value associated with them (57).

Care should also be taken to remove visual cues that dogs could use to potentially identify targets. Although olfaction has been shown to be the dominant sense used by trained detection dogs to locate targets when compared to vision (75, 76), other studies have shown that in some contexts, such as when a human gesture conflicts with an olfactory cue, dogs may defer to visual cues (46, 47). Further, the use of distractors and controls requires a systematic approach of managing the materials, which often involves visually marking the materials. Dogs have dichromatic vision, expressing only two forms of light-sensitive photo pigments in the cells of the retina pertaining to color as compared to humans which express three forms and are trichromatic. Though this is generally considered to result in dogs exhibiting deutreranopia, a human-like red/green color blindness (77, 78), studies have demonstrated that dogs are capable of discriminating colors based on differences of brightness intensities (79, 80). Although color is thought to be predominant over brightness in canine visual processing, caution must be used if utilizing color coding in sample management as the colors may still be perceived differentially and could result in a visual cue being associated with the target.

Criteria Testing

Before formal testing occurs, it is important to validate training and establish that dogs are prepared for testing. For example, when testing whether dogs generalize from a trained odor to an untrained odor or whether a dog will be successful at detecting a trained odor in a different context, researchers often require that dogs meet some pre-determined performance criteria [e.g., (81, 82)]. The criteria often consist of a minimum hit rate to the trained target odor and a maximum false alarm rate. For example, Porritt et al. (23) developed a pass criterion based on signal detection theory in conjunction with subject matter experts, resulting in an acceptable pass criterion of at least a 70% higher hit rate than the false alarm rate. The direct comparison between hit and false rate requires that individual dogs respond to their trained target significantly more often than they commit a false alarm in order to meet the criteria (23). If dogs' ability to meet a performance criterion prior to testing is not demonstrated, test performance will be unclear. Furthermore, conducting criterion tests with all controls in place provides a baseline measure of performance and provides dogs experience with the experimental design that will occur in testing so that performance is not disrupted when test protocols are implemented.

Acceptable accuracy rates vary across researchers and organizations, and should be pre-determined based on the goals of the testing. Ideally, a training criterion should enable researchers to be confident that the dog is prepared for testing and allow meaningful comparisons to test performance (discussed below). More stringent criteria may be required for explosives detection dogs being trained for operational deployment with greater risks associated with errors, or for drug detection or forensics dogs for which training records may serve as probable cause or evidence in court. It has been argued that true detection accuracy should approach 100%, but such expectations may be unreasonable considering the variety of factors related to odor presentation, odor source, and other test parameters (12). In some circumstances, purposefully tailoring training toward a liberal bias in responding when target odor is present (e.g., aviation explosives detection) or toward a conservative bias in not responding when target odor is absent (e.g., drug detection) is warranted.

Test Parameters

In order to ensure validity of the results, specific session and trial parameters should be considered when evaluating performance. Most notably, both target and blank trials (i.e., no target odor present) should be included and should be randomized across the test session. In addition to reducing false alerts and positional biases as discussed above, blank trials are useful in keeping the probability of encountering a target on each search unpredictable to both the dog and handler. For example, if a target is placed in position five on a six-position wheel, every time a dog samples an empty position there is an increasing chance that the next position contains a target. Thus, detection rates for targets in later positions could be artificially inflated. By inserting blank trials, the dog cannot determine if a later position is likely to contain a target or if it is a blank (2), and the handler will be unaware of whether a lack of response was a miss or a blank trial. As mentioned above, varying the number of targets present on each trial is an alternative to inserting blank runs, though is arguably less practical.

This design is sufficient when dogs are trained to sample systematically and are quite accurate, however, adjustments are sometimes required. For example, it may be necessary to allow dogs to rerun the trial or search if an area is missed, the dog displays a COB, or shows interest but doesn't respond (9). Critically, the decision of whether the dog sufficiently searched the area or not should be made by a blind handler or evaluator before any feedback of the trial outcome is given. It is also important to note that allowing a dog to resample positions or re-run trials complicates calculating the correct rejection rate, and thus a priori decisions should be made regarding which run will be counted toward data analysis.

The number of test trials performed is another important consideration and should be determined based on statistical validity. As discussed above, statistical power will be influenced by the number of subjects which can be determined by an a priori power analysis. When the number of subjects is difficult to control, a priori analysis can also be used to determine number of test trials to determine a specific effect (24). However, the effect of repeating test trials for an individual subject should be considered. For example, rapid within-session learning can occur after repeated exposures if responses on test trials are reinforced (83). Alternatively, withholding reinforcement for responses on test trials can lead to within-session extinction. One option to reduce learning or extinction across trials is to implement intermittent reinforcement prior to testing so that performance is maintained in the absence of reinforcement, or to non-differentially reinforce correct and incorrect responses (29). Controlling for within-session changes in responding is especially critical in generalization studies when the goal is to assess spontaneous responses to an untrained odor, given that dogs are capable of learning to respond to a new odor in as little as 2–3 exposures (81). Thus when possible, the number of test trials should be limited in order to give a more accurate representation of initial response to the odor. When sample sizes are low and repeated test trials are needed to obtain sufficient data, first-trial performance or changes in responding across multiple exposures to the test odor should always be analyzed.

Within-session changes in motivation can also occur if testing is too difficult or too many non-reinforced trials occur. In order to maintain motivation during these testing sessions, reinforced baseline target trials are often dispersed throughout the session or search (24, 84). The inclusion of baseline trials during a test session also allows for a comparison between hit rates on baseline and test odors. For example, in generalization testing, comparing responses to trained and untrained targets is necessary for determining whether generalization occurred (85). Specifically, if the number of hits to the test odors is not significantly different than the number of hits to the trained target than it can be concluded that the dogs successfully generalized. In addition, comparing hit rate on test odors to hit rate on non-target odors as well as to random chance allows for an assessment of the degree of generalization. Responses to test odors that are significantly below baseline hit rate, but significantly above chance, could indicate that some degree of generalization occurred. Responses to test odors that are not significantly different from the false alarm rate indicates a lack of specificity which likely inflated hit rate. Borrowing from studies of animal concept learning, generalization that is equivalent to baseline and significantly above random chance could be considered *full transfer*, generalization that is below baseline but above chance could be considered *partial transfer,* and generalization that is not statistically different from chance could be considered a failure to transfer (62).

CONSIDERATIONS RELATED TO ODOR

Characterization of Odor Samples

In order to properly conduct olfactory detection research, it is imperative to have a clear understanding of the odorants that make up the odor of the substance to be detected and approximately how much of it is being presented during olfactory testing. Without an understanding of the odor being delivered, one risks testing or training the dog on a set of odorants or quantity of odorants different than intended. There are several factors to consider regarding odor characterization, perception, and availability, discussed in the following section.

Qualitative Characterization of Odor

Generally, if a target material is not in the gas phase, it cannot be detected through olfaction (it is possible that dogs are capable of detecting very small particles that enter the nasal passageway, but should this be the case, odorant molecules on the particle are likely volatilized in the nasal cavity and ultimately detected in the gas phase, or broken down within the mucous layer and delivered to the olfactory receptors by transport proteins). Often times the molecules making up the target material are too large to be readily available in the gas phase. Instead, the animal will detect an associated odorant or collection of odorants that are unique to that target. These odorants are often referred to as the active odor (odorants) (60), and have been studied for many substances relevant to canine detection (86–93). This is of particular importance when considering detection of a target material with a very low vapor pressure, such as many narcotics or explosives. For example, cocaine is a large molecule with an accordingly low vapor pressure [303 g/mol; 3 x 10^{-7} Torr at 20°C (94)], and is not readily available in the vapor phase. However, methyl benzoate, a degradation product of cocaine is smaller in size (136 g/mol) with a higher vapor pressure (3 x 10^{-3} Torr at 20°C). In regard to testing detection thresholds of a low volatility substance such as cocaine, it is imperative to understand that this threshold is related to the amount of methyl benzoate present and not the amount of solid cocaine.

In most circumstances, the recognizable odor of a target material is not made up of a single odorant, but of a mixture of odorants, referred to as the odor profile. For instance, cadaver odor consists of hundreds of individual volatile and semi-volatile analytes that together create a unique odor profile recognizable by trained human remains search dogs (95, 96). For complex odor profiles such as living human and cadaver scent, it can be quite difficult to delineate which compounds the dog uses for detection, or if extraneous odors from contamination have added to or altered the odor profile.

Quantitative Characterization of Odor

The quantity of odor available is as equal a concern as the quality. Returning to the cocaine example, research has also shown that the quantity of methyl benzoate present from cocaine is dependent on the type of cocaine with pharmaceutical-grade cocaine yielding a significantly lower amount of methyl benzoate than that from street cocaine (88). Furthermore, research with ammonium nitrate, another low volatility substance, has shown that variations in the source and purity of ammonium nitrate as well as in the amount of ammonia influences the detection of ammonium nitrate (85, 97), demonstrating the importance of being mindful of possible variations in odorant concentration between related substances.

There is a common misconception that the amount of odor available can be easily altered by increasing the mass (or volume, in the case of a liquid) of the material (i.e., 10 g of a given material will yield 10 times as much odor as 1 g of the same material) (98). While mass or volume of a given substance is correlated to odor availability, increasing (or decreasing) the amount of a solid or liquid does not generate an equivalent change in the vaporous components (10 g does *not* indeed yield a 10 times increase in odor over 1 g). This is because the amount of odorant emitted from a given substance is also related to the substance's vapor pressure, the rate of evaporation or sublimation of the odorant(s), the total available surface area, and environment factors, such as ambient temperature, humidity, and air movement (99–101). Although operational and scientific communities frequently overlook the effect of surface area, altering surface area is a highly efficient way of altering odor availability in both testing and training scenarios. An odorant can only be released into the gaseous phase from the outer surfaces of a material, whether a solid or a liquid. For instance, a single square of C-4, a plastic explosive, will have less surface area and thus less odor availability than the same mass of C-4 spread out in a thin layer or cut up into many smaller cubes. Container opening size will have a similar effect—for a given volume of liquid, more odor will be available from a container or opening with a larger diameter. Thus, filling a container to the top is not necessarily an effective way to increase the amount of odor. Although, in an open container, increasing the size of the mouth or opening is indeed an effective way of increasing odor availability for the same volume of material, where a pin-sized hole will release a very low amount of odor compared to an open wide-mouthed jar. This can be an effective way of increasing or decreasing odor availability during testing. Likewise, in a closed container, once the headspace above the sample in the container is saturated with odor (i.e., equilibrium has been reached), a further increase in amount of material will not result in a greater concentration of odor (98, 102). For example, researchers placed 10 mg of triacetone triperoxide (TATP) in the bottom of a vial, and the crystals only covered ∼10% of the bottom of the vial. The resulting vapor concentration from the vial was measured to be 80 ng/L at equilibrium. When 200 mg of TATP was placed in the vial, now covering 100% of the bottom, the resulting vapor concentration at equilibrium doubled to 160 ng/L. Finally, when the amount of material was further increased to 1,000 mg, which just increased the *volume* of TATP but not the *surface area* (still 100% coverage), the vapor concentration only increased by 18% to 190 ng/L (103).

There are many ways of characterizing and quantitating the odor profile of a given substance. The most common technique for measuring trace vapor components in the headspace is by solid phase microextraction (SPME) to extract the vapor molecules, paired with analysis by gas chromatography/mass spectrometry (GC/MS) for analysis of the extractant (104, 105). Unless a rigorous quantification method is used, which can be particularly arduous in SPME-GC/MS, each step in the analysis lends some amount of bias in the ratio of analytes measured. Meaning, the SPME fiber adsorbs some analytes preferentially to others, and the resulting data will yield a greater abundance of those analytes compared to others that may be present in the sample headspace in the same quantity. The gas chromatography column and mass spectrometer will also influence the ratio of analytes in the resulting data. It is thus important for researchers to understand that, with this or other headspace analysis methods, the ratios of measured odorants are not necessarily entirely reflective of the ratio that exists in nature.

Furthermore, the compounds that are in the highest abundance in the headspace, as determined by instrumental analysis, are not necessarily the same compounds that are *perceived* as having the greatest impact by dogs (106). Returning again to the cocaine example, Furton et al. (88) examined the headspace of multiple cocaine samples and found a number of volatile compounds present, to include methyl benzoate. Though methyl benzoate was not the dominant volatile species in the headspace, it was shown to be the active odorant of cocaine (88). Rice and Koziel (106) highlight that this discrepancy between what is measured instrumentally and what is perceived by the olfactory system has important implications in the creation and testing of mimic or surrogate training aids (106). The researchers compared instrumentally measured odorants from illicit drug samples and surrogate training aids to reported perceived olfactory intensity using both human and canine subjects. The results demonstrated that there was not a direct relationship between odorant concentration and perceived odor intensity, and that surrogates made using the compounds dominant in the instrumentally-determined odor profile, and *not* the *perceived* active odorants, would not elicit the same response.

Odor Delivery

In the specific instance of olfactory detection threshold (ODT) testing, it is particularly important to maintain a known and constant source of odor at a given concentration throughout testing, and be able to deliver that odor at adjustable and accurate concentrations as the testing requires. This task can be quite challenging as evidenced by the high variability in published values of ODTs for dogs even when evaluating the same odorant (7, 8, 107). Factors such as previous training and familiarity with the odorant, individual differences between dogs, and testing protocols are potential sources of variation; however, differences in odor delivery methods are large contributors to such discrepancies (106). The two greatest factors are ab/dsorption to surrounding surfaces and dilution as odorants move away from the source. Whenever an odor source is contained or passes through a material, such as tubing in an olfactometer, some amount of the odorant is potentially lost due to ab/dsorption resulting in the delivery of a vapor concentration lower than intended. Though some amount of loss is likely to all materials, when delivering odor with an olfactometer or the like it is recommended to use Teflon or passivated materials (such as coated stainless steel) for all tubing through which the odor passes, and it is additionally recommended to heat these materials and to remove all possible cold spots from the airflow pathways to minimize losses to adsorption. If the odorant being tested has a high vapor pressure, these means should alleviate the majority of adsorption to the wetted portions of the flow path. Should the material being tested be of higher molecular weight/lower vapor pressure, quantitative measurements of the vapor concentration should be conducted to account for loss to adsorption and calculate the final concentration delivered. Finally, as soon as the vapor exits the port of an olfactometer or diffuses into the environment beyond the odor source or containment, the vapor plume or stream is diluted by surrounding air. Furthermore, air flow in the testing location may carry the odorants away from source further diluting the concentration. Designing the experiment in such a way that the dog has to place its muzzle into a portal or deep container with a smaller opening and ensuring the dogs are trained to bring their muzzle close to source will begin to alleviate this issue. Again, using quantitative measurements of the vapor concentration at the point where the dog samples is the best way to confirm the dog is experiencing the intended odor concentration (108).

Contamination

Contamination and storage of target and non-target materials are essential and often inter-related considerations in maintaining the integrity of canine olfactory detection research. Contamination occurs when odor or scent is inadvertently transferred between materials or odor sources. A major source of contamination is the introduction of human scent to a target material. Mishandling targets can cause human scent to become associated with a given target, either confusing the odor profile or providing a secondary odor that dogs may learn to identify instead of the target odor. Further, scent trails of the people placing targets can contaminate testing areas, and provide dogs with a trail to follow toward a hidden target material (109–111). Contamination form saliva deposited on a target location can also provide inadvertent odor cues, which can occur when carousel setups are used if the positions are rotated but containers are not replaced (37, 112). As discussed above, the use of controls is important for minimizing the risk of dogs learning to respond to contaminating odorants rather than the target odor itself.

Cross-contamination occurs when the odor of one target is unintentionally transferred to another target, which can have varying effects on olfactory tests. For instance, dogs may incorrectly learn the target odor as a mixture of the contaminating odor and the target odor, and may fail to identify the pure trained material in a testing scenario. Cross-contamination most commonly occurs when different target materials are stored in close proximity to one another, otherwise known as "unit scent," and is most prevalent when those materials have a large disparity in vapor pressure. For example, Hallowell et al. (113) found that likely cross-contamination of explosives stored together led to a preventable fault in canine training (113). The dogs were only able to detect compounds with the highest vapor pressures, and could not identify lower vapor pressure explosives that had been co-stored.

In a study of cross-contamination between co-stored training materials (birch, clove, and anise essential oils), the relative amounts of cross-contamination apparent were compared for three types of containment (114). In this experiment, 5 μL of each oil was placed on separate cotton swabs, stored inside one of three common primary containment systems (20 mL glass vials, 4 oz canning jars, or Mylar bags), and placed within a single outer jar. Cross-contamination, monitored over a 24-week period, was noted as early as week 1. Methyl salicylate, a volatile component of birch oil, was identified in the clove and anise samples of each primary containment system. Such cross-contamination between segregated materials has the potential to alter the odor profiles of target aids and affect the integrity of testing materials. Proper handling and storage of testing materials including the use of both primary and secondary containment can be very important as barriers for odor containment and protection of target materials, especially when materials must be stored in close proximity with other testing materials. The primary, or inner layer of containment, should not impart odor to the training material or react with it. A non-corrosive metal or glass containment is suggested for this layer, as plastics emit chemicals that can cause contamination. The secondary, or outer, containment should be a non-permeable material with a lid that eliminates leakage (72).

Another source of contamination results from residual odor, sometimes referred to as inverse contamination or contamination of the working environment, which occurs when the target material leaves remnants or volatiles in the environment where it was placed. This can often occur when a substance is left in direct contact with a surface, such as a table or drawer, and when the substance is allowed a period of time to sit before the testing session begins (57). Secondary transfer of odor can occur when odor from one material is transferred onto a surface, and then from that surface onto a second container. This is likely to occur when a target material is removed from a location in the testing scenario and a second material is placed on top of existing residual odor. A similar effect can be seen from transfer by touch when the individual preparing the test touches one testing material, contaminating the set of gloves, and then touches another testing material with the same gloves. Papet (115) and the UK Centre for the Protection of National Infrastructure each specifically warn against these risks, and suggest placing target materials on often-replaced barriers such as wax paper and changing gloves frequently to help limit residual odors (72, 115). Such contact contamination is, however, even more complicated. For example, a probable solution to such contact contamination is having separate individuals emplace target and non-target odor samples, but this strategy may result in the availability of a discriminable difference between the odor of the two individuals being associated with the different samples. This may happen regardless of the wearing and regular changing of gloves as canine olfaction is sensitive enough to detect the effluent from individuals that contaminates samples by just the individuals being in close proximity to the samples.

The duration of residual odors depends on the testing material itself, the substrate being contaminated by the residual odor, the amount of contamination, and environmental conditions. For example, residues from narcotics or essential oil have been shown to be detected anywhere from 2 to 48 h after removal of the odor source (12, 114, 116). Dogs have been able to detect human remains residue in soil up to 667 days after removal (117), and have been successful in locating blood on cotton swatches after five laundry cycles (118). Since residual odor can be difficult to predict, it is best to keep records of past testing odor locations to help identify apparent false alerts that are actually correct but caused by residual odor (72).

Effects of Wrapping/Containment

It is nearly impossible to present a target substance free of any type of container or packaging. Particularly in an operational setting, the target of interest is likely to be securely wrapped, packaged, and/or obscured in some manner. Even in this situation, odorants from the target are likely to be present on the outer barrier for a number of reasons. This form of contamination can be problematic if the goal is to assess dogs' ability to detect odor that is concealed. The durability of odorants on the outside of a container is dependent on the amount and manner deposited and the tendency for the outer material to absorb the odorant in question. The rate of diffusion or permeation of odorants through the wrapping or packaging material is also dependent on the material type and thickness.

In a testing situation, such as olfactory threshold measurements, it is important to keep in mind that all packaging and wrapping around a target material will absorb some amount of the odor, even in "non-stick" materials such as Teflon. Using TNT vapors pulsed at various surfaces, Poziomek et al. (119) demonstrated that the TNT adsorbed more strongly to some surfaces tested than others, and, in fact, Teflon was the optimal substrate for adsorption, retention, and recovery of TNT (119). Again, the molecular structure of the odorant and type of wrapping, as well as temperature and other environmental factors, will affect ab/dsorption, and like with permeation, ab/dsorption can change the ratios of odorants in the odor profile with certain odorants being retained more strongly than others. For example, when odor profiles from living and deceased people were collected onto a sorbent material, it was shown that the resulting instrumentally measured odor profiles were dependent on the type of adsorbent material used in collection (120, 121).

Similar to wrapping, buried odor behaves and is transported to the surface for detection through complex processes. A body of literature has been devoted to describing buried odor, particularly in the case of landmine (122–128) and cadaver detection (117, 129, 130). As an overview, the evolution of buried odor involves dynamic processes of absorption, diffusion, dissolution in water, transformation by microbes, and uptake by vegetation that change with changing conditions. A detailed discussion is beyond the scope of this review, but in summary, for a dog to detect buried odor, free odor molecules must diffuse through soil to the surface. However, free odorant molecules may absorb to soil particles or dissolve into water, where they then may be carried away with ground water or taken up by the roots from nearby vegetation. This is the reason handlers often report of dogs not indicating a buried hide at source, but instead at a nearby water source or tree. The movement of free odorants is dependent on the type of odorant, the soil type, porosity, and moisture content, and the temperature, thermal radiation from sunlight, and air movement above the burial. In general, as soil becomes dry, more odor molecules absorb to the soil particles, lowering the odor availability. Moisture in the water enhances diffusion and increases odor availability (122, 128, 131). Because of the multifaceted nature of buried odor movement and availability, constructing reproducible testing with known variables is challenging. As such, any testing conducted with buried odor should be carried out with great care with as many defined variables as possible.

Set Time

Allowing each dog to experience the odor in the same way each time requires the ability to confidently deliver a known and constant odor profile and odor concentration over the duration of a test or set of olfactory experiments. In order for the first and last dogs being tested (and all in between) to have access to the same concentration of odor, the sample must be delivered following a proper equilibration time for the chosen container and material being tested, commonly referred to as "set" or "soak" time in an operational or field setting. Unfortunately, there is no single equilibration time that is appropriate for all scenarios, but understanding the factors that affect equilibration time can assist researchers in making an educated decision given a particular set of experimental parameters. Many of the factors discussed above will affect soak time. In general, the higher the vapor pressure of the odorant of interest, the faster the system will come to equilibrium. The actual time will also be dependent on and change with the amount of material being used, the size and type of container, whether or not the odorant(s) must permeate through any sort of concealment, ambient temperature and humidity, air flow in the environment, and the presence and quantity of multiple odorants in the container. There are additional nuances to this, of course. For one instance, if the odorants do not simply evaporate/sublimate from the testing material, but instead evolve from a reaction of some sort, such as a decomposition reaction, time to equilibrium will also be dependent on the rate of that reaction. For instance, the explosive hexamethylene-triperoxide-diamine (HMTD) itself has a very low vapor pressure yielding very little molecular HMTD available as an odorant; however, it degrades under normal ambient conditions producing a number of detectable odorants, meaning equilibrium is dependent on the rate of the decomposition reaction (132).

Once an odorant has reached equilibrium in its container, the odor concentration will stay constant, assuming that none of the variables above change. However, this is not always the case. In one example, the odor profile associated with certain types of aluminum powder, a component of some homemade explosives, is derived from the breakdown of the stearic acid coating yielding a mix of odorants that, to humans, smell similar to crayons. Field measurements of the headspace of the aluminum powder on a cool morning yielded an abundance of odorants related to stearic acid decomposition, but when tested again later in the day on a warm afternoon the same amount of material yielded only very low levels of odorants. Further research indicated that exposure to heat generated by the sun on the warm afternoon actually drove off the odorants faster than they were produced from the stearic acid reaction (133). Though this describes a very unique set of circumstances and materials, it illustrated why it is important to consider not only the time required for equilibration, but also the duration the odor remains available. Depending on the source of the odor and the amount of substance being used, it is possible to deplete the available odorants over the duration of a lengthy test. Some commercially available training aids, for instance, have a short reported service-life of only several hours. In order to conduct a test that is reproducible and stable over its duration, it is thus imperative to be aware both when the substances being tested have reached equilibrium and when the odor begins to be depleted. The soak or set times selected by various canine certifying bodies are generally non-specific with many requiring a set time of at least 30 min with no maximum set time given (53, 134).

CONCLUSIONS

A lack of standardization in canine olfactory detection assessments, both in scientific research and in evaluations of operational canines, has led to a wide variability in results. This lack of standardization partially stems from the wide range of aspects examined by olfactory detection research. Nonetheless, attempts should be made to increase consistency in methodologies, such as standards for necessary controls to include and reporting of data, to allow for ease of interpreting results, internal validity of data, and making meaningful comparisons across studies. In this review, we discuss the range of factors that should be considered when designing and conducting canine olfactory detection studies, many of which have direct applications to operational testing.

It is important for researchers to conduct both basic and applied research related to canine detection. However, it should be cautioned that not all research can be extrapolated to operational performance due to variables discussed in this review. Specific variables influence the quantification of detection dog performance such as experimental design, testing bias, odor contamination, training aid storage/handling, odor characteristics, experimental controls, and statistical analysis. Methods for testing canine scent detection vary influencing the outcome metrics of performance and the validity of results. Operators, management teams, policy makers, and law enforcement rely on scientific data to make decisions, design policies, and to advance canine technologies. Therefore, scientists conducting research should incorporate as many operational constraints as possible so that the data can be applied to operational performance. In addition, operational teams should adopt rigorous scientific standards in order to scientifically validate their dogs' capabilities. This will lead to better informed decisions about capability, vulnerability, and risk analysis.

AUTHOR CONTRIBUTIONS

LL, SK, LD, and AS wrote the first draft of the manuscript. CA, LW, and MS made significant revisions and contributions to the content. All authors read and approved the final draft.

FUNDING

Funding for preparation of this review was provided by the Auburn University College of Veterinary Medicine Canine Performance Sciences Program through the generous support of Walt and Ginger Woltosz, the Richard G. and Dorothy Metcalf Endowment, Alan Kalter and Dr. Chris Lezottee Health Excellence Fund, and James M. Hoskins Endowment.

ACKNOWLEDGMENTS

We thank all the canine training professionals of Canine Performance Sciences and other programs, as well as the many canine practitioners who have generously shared their experience and expertise with us over many years and which has been invaluable in our understanding of canine olfactory detection.

REFERENCES

1. Johnen D, Heuwieser W, Fischer-Tenhagen C. Canine scent detection-Fact or fiction? *Appl Anim Behav Sci.* (2013) 148:201–8. doi: 10.1016/j.applanim.2013.09.002
2. Johnen D, Heuwieser W, Fischer-Tenhagen C. An approach to identify bias in scent detection dog testing. *Appl Anim Behav Sci.* (2017) 189:1–12. doi: 10.1016/j.applanim.2017.01.001
3. Taylor KD, Mills DS. The development and assessment of temperament tests for adult companion dogs. *J Vet Behav Clin Appl Res.* (2006) 1:94–108. doi: 10.1016/j.jveb.2006.09.002
4. Beebe SC, Howell TJ, Bennett PC. Using scent detection dogs in conservation settings: a review of scientific literature regarding

their selection. *Front Vet Sci.* (2016) 3:96. doi: 10.3389/fvets.2016.00096

5. Goldblatt A, Gazit I, Terkel J. Olfaction and explosives detector dogs. In: Helton W, editors. *Canine Ergonomics: The Science of Working Dogs.* Boca Raton: CRC Press (2009). p. 135–75.
6. Johnston JM, Myers LJ, Waggoner LP, Williams M. Determination of canine olfactory thesholds using operant laboratory methods. *Subst Detect Syst.* (1994) 2092:238–43. doi: 10.1117/12.171244
7. Krestel D, Passe D, Smith JC, Jonsson L. Behavioral determination of olfactory thresholds to amyl acetate in dogs. *Neurosci Biobehav Rev.* (1984) 8:169–74. doi: 10.1016/0149-7634(84)90037-X
8. Walker DB, Walker JC, Cavnar PJ, Taylor JL, Pickel DH, Hall SB, et al. Naturalistic quantification of canine olfactory sensitivity. *Appl Anim Behav Sci.* (2006) 97:241–54. doi: 10.1016/j.applanim.2005.07.009
9. Concha AR, Guest CM, Harris R, Pike TW, Feugier A, Zulch H, et al. Canine olfactory thresholds to amyl acetate in a biomedical detection scenario. *Front Vet Sci.* (2019) 5:345. doi: 10.3389/fvets.2018.00345
10. Polgár Z, Kinnunen M, Újváry D, Miklósi Á, Gácsi M. A test of canine olfactory capacity: comparing various dog breeds and wolves in a natural detection task. *PLoS ONE.* (2016) 11:e0154087. doi: 10.1371/journal.pone.0154087
11. Jenkins EK, DeChant MT, Perry EB. When the nose doesn't know: canine olfactory function associated with health, management, and potential links to microbiota. *Front Vet Sci.* (2018) 5:56. doi: 10.3389/fvets.2018.00056
12. Jezierski T, Adamkiewicz E, Walczak M, Sobczynska M, Górecka-Bruzda A, Ensminger J, et al. Efficacy of drug detection by fully-trained police dogs varies by breed, training level, type of drug and search environment. *Forensic Sci Int.* (2014) 237:112–8. doi: 10.1016/j.forsciint.2014.01.013
13. Hall NJ, Glenn K, Smith DW, Wynne CDL. Performance of pugs, German shepherds, and greyhounds (*Canis lupus familiaris*) on an odor-discrimination task. *J Comp Psychol.* (2015) 129:237–46. doi: 10.1037/a0039271
14. Gadbois S, Reeve C. Canine olfaction: scent, sign, and situation. In: *Domestic Dog Cognition and Behavior: The Scientific Study of Canis familiaris.* New York, NY: Springer (2014). p. 3–29. doi: 10.1007/978-3-642-53994-7
15. Jamieson LTJ, Baxter GS, Murray PJ. Identifying suitable detection dogs. *Appl Anim Behav Sci.* (2017) 197:1–7. doi: 10.1016/j.applanim.2017.06.010
16. Jones KE, Dashfield K, Downend AB, Otto CM. Search-and-rescue dogs: an overview for veterinarians. *J Am Vet Med Assoc.* (2004) 225:854–60. doi: 10.2460/javma.2004.225.854
17. Brucks D, Soliani M, Range F, Marshall-Pescini S. Reward type and behavioural patterns predict dogs' success in a delay of gratification paradigm. *Sci Rep.* (2017) 7:42459. doi: 10.1038/srep42459
18. Schoon GAA. Scent identification lineups by dogs (*Canis familiaris*): experimental design and forensic application. *Appl Anim Behav Sci.* (1996) 49:257–67. doi: 10.1016/0168-1591(95)00656-7
19. DeChant M, Ford C, Hall N. Effect of handler knowledge of the detection task on canine search behavior and performance. *Front Vet Sci.* (2020) 7:250. doi: 10.3389/fvets.2020.00250
20. Simon AG, Degreeff LE, Peranich K, Furton KG. *Canine Generalization to Molecularly Similar Odors and Odor Mixtures*. (No. NRL/MR/6181-18-9797). Naval Research Lab, Washington, DC (2018).
21. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* New York, NY: Routledge (1988).
22. Arden R, Bensky MK, Adams MJ. A review of cognitive abilities in dogs, 1911 through 2016: more individual differences, Please!. *Curr Dir Psychol Sci.* (2016) 25:307–12. doi: 10.1177/0963721416667718
23. Porritt F, Mansson R, Berry A, Cook N, Sibbald N, Nicklin S. Validation of a short odour discrimination test for working dogs. *Appl Anim Behav Sci.* (2015) 165:133–42. doi: 10.1016/j.applanim.2014.11.021
24. Edwards TL, Browne CM, Schoon A, Cox C, Poling A. Animal olfactory detection of human diseases: guidelines and systematic review. *J Vet Behav Clin Appl Res.* (2017) 20:59–73. doi: 10.1016/j.jveb.2017.05.002
25. Bryce E, Zurberg T, Zurberg M, Shajari S, Roscoe D. Identifying environmental reservoirs of *Clostridium difficile* with a scent detection dog: preliminary evaluation. *J Hosp Infect.* (2017) 97:140–5. doi: 10.1016/j.jhin.2017.05.023
26. Walczak M, Jezierski T, Górecka-Bruzda A, Sobczynska M, Ensminger J. Impact of individual training parameters and manner of taking breath odor samples on the reliability of canines as cancer screeners. *J Vet Behav Clin Appl Res.* (2012) 7:283–94. doi: 10.1016/j.jveb.2012.01.001
27. Gadbois S, Reeve C. The semiotic canine: scent processing dogs as research assistants in biomedical and environmental research. *Dog Behav.* (2016) 2:26–32. doi: 10.4454/DB.V2I3.43
28. Edwards TL. Automated canine scent-detection apparatus: technical description and training outcomes. *Chem Senses.* (2019) 44:449–55. doi: 10.1093/chemse/bjz039
29. Hall NJ, Wynne CDL. Odor mixture training enhances dogs' olfactory detection of Home-Made Explosive precursors. *Heliyon.* (2018) 4:e00947. doi: 10.1016/j.heliyon.2018.e00947
30. Macmillan NA, Creelman CD. *Detection Theory: A User's Guide.* New York, NY: Psychology Press (2004).
31. Helton WS, editor. Overview of scent detection work: issues and opportunities. In: *Canine Ergonomics: The Science of Working Dogs.* Boca Raton, FL: CRC Press (2009). p. 83–96.
32. Centre for the Protection of National Infrastructure. *Canine Odour Discrimination Test.* Detection Dog Guidance Notes (2018). Available online at: https://www.cpni.gov.uk/canine-detection-guidance-notes
33. Lu XCM, Slotnick BM, Silberberg AM. Odor matching and odor memory in the rat. *Physiol Behav.* (1993) 53:795–804. doi: 10.1016/0031-9384(93)90191-H
34. Fadel FR, Driscoll P, Pilot M, Wright H, Zulch H, Mills D. Differences in trait impulsivity indicate diversification of dog breeds into working and show lines. *Sci Rep.* (2016) 6:1–10. doi: 10.1038/srep22162
35. Johnen D, Heuwieser W, Fischer-Tenhagen C. How to train a dog to detect cows in heat-Training and success. *Appl Anim Behav Sci.* (2015) 171:39–46. doi: 10.1016/j.applanim.2015.08.019
36. Lit L. Evaluating learning tasks commonly applied in detection dog training. In: Helton W, editor. *Canine Ergonomics: The Science of Working Dogs*. Boca Raton: CRC Press (2009). p. 99–114.
37. Jezierski T, Walczak M, Górecka A. Information-seeking behaviour of sniffer dogs during match-to-sample training in the scent lineup. *Polish Psychol Bull.* (2008) 39:71–80. doi: 10.2478/v10059-008-0010-y
38. Murarka M, Vesley-Gross ZI, Essler JL, Smith PG, Hooda J, Drapkin R, et al. Testing ovarian cancer cell lines to train dogs to detect ovarian cancer from blood plasma: a pilot study. *J Vet Behav.* (2019) 32:42–8. doi: 10.1016/j.jveb.2019.04.010
39. Curran AM, Prada PA, Furton KG. Canine human scent identifications with post-blast debris collected from improvised explosive devices. *Forensic Sci Int.* (2010) 199:103–8. doi: 10.1016/j.forsciint.2010.03.021
40. Lit L, Schweitzer JB, Oberbauer AM. Handler beliefs affect scent detection dog outcomes. *Anim Cogn.* (2011) 14:387–94. doi: 10.1007/s10071-010-0373-2
41. Engeman RM, Vice DS, York D, Gruver KS. Sustained evaluation of the effectiveness of detector dogs for locating brown tree snakes in cargo outbound from Guam. *Int Biodeterior Biodegrad.* (2002) 49:101–6. doi: 10.1016/S0964-8305(01)00109-3
42. Porritt F, Shapiro M, Waggoner P, Mitchell E, Thomson T, Nicklin S, et al. Performance decline by search dogs in repetitive tasks, and mitigation strategies. *Appl Anim Behav Sci.* (2015) 166:112–22. doi: 10.1016/j.applanim.2015.02.013
43. Jamieson L, Baxter G, Murray P. You are not my handler! Impact of changing handlers on dogs' behaviours and detection performance. *Animals.* (2018) 8:176. doi: 10.3390/ani8100176
44. Zubedat S, Aga-Mizrachi S, Cymerblit-Sabba A, Shwartz J, Leon JF, Rozen S, et al. Human–animal interface: the effects of handler's stress on the performance of canines in an explosive detection task. *Appl Anim Behav Sci.* (2014) 158:69–75. doi: 10.1016/j.applanim.2014.05.004
45. Reed SE, Bidlack AL, Hurt A, Getz WM. Detection distance and environmental factors in conservation detection dog surveys. *J Wildlife Manage.* (2011) 75:243–51. doi: 10.1002/jwmg.8
46. Lazarowski L, Rogers B, Waggoner LP, Katz JS. When the nose knows: ontogenetic changes in detection dogs' (*Canis familiaris*) responsiveness to social and olfactory cues. *Anim Behav.* (2019) 153:61–8. doi: 10.1016/j.anbehav.2019.05.002

47. Szetei V, Miklósi Á, Topál J, Csányi V. When dogs seem to lose their nose: an investigation on the use of visual and olfactory cues in communicative context between dog and owner. *Appl Anim Behav Sci.* (2003) 83:141–52. doi: 10.1016/S0168-1591(03)00114-X
48. Soproni K, Miklósi A, Topál J, Csányi V. Comprehension of human communicative signs in pet dogs (*Canis familiaris*). *J Comp Psychol.* (2001) 115:122–6. doi: 10.1037/0735-7036.115.2.122
49. Virányi Z, Topál JÓ, Gácsi MÁ, Miklósi Á, Csányi V. Dogs respond appropriately to cues of humans' attentional focus. *Behav Process.* (2004) 66:161–72. doi: 10.1016/j.beproc.2004.01.012
50. Albuquerque N, Guo K, Wilkinson A, Savalli C, Otta E, Mills D. Dogs recognize dog and human emotions. *Biol Lett*. (2016) 12:20150883. doi: 10.1098/rsbl.2015.0883
51. Tuyttens FAM, de Graaf S, Heerkens JLT, Jacobs L, Nalon E, Ott S, et al. Observer bias in animal behaviour research: can we believe what we score, if we score what we believe? *Anim Behav.* (2014) 90:273–80. doi: 10.1016/j.anbehav.2014.02.007
52. Centre for the Protection of National Infrastructure. *Why You Should Train and Test Detection Dogs 'Double-Blind.'* Detection Dog Guidance Notes (2018). Available online at: https://www.cpni.gov.uk/canine-detection-guidance-notes
53. Scientific Working Group on Dog and Orthogonal detector Guidelines (SWGDOG). *Approved Guidelines: Substances – SC8.* (2011). Available online at: http://swgdog.fiu.org; Retrieved April 20, 2020 from https://ifri.fiu.edu/research/detector-dogs/swgdog/
54. Dorman DC, Foster ML, Fernhoff KE, Hess PR. Canine scent detection of canine cancer: a feasibility study. *Vet Med.* (2017) 8:69–76. doi: 10.2147/VMRR.S148594
55. Elliker KR, Sommerville BA, Broom DM, Neal DE, Armstrong S, Williams HC. Key considerations for the experimental training and evaluation of cancer odour detection dogs: lessons learnt from a double-blind, controlled trial of prostate cancer detection. *BMC Urol.* (2014) 14:22. doi: 10.1186/1471-2490-14-22
56. Angle TC, Passler T, Waggoner PL, Fischer TD, Rogers B, Galik PK, et al. Real-time detection of a virus using detection dogs. *Front Vet Sci.* (2016) 2:79. doi: 10.3389/fvets.2015.00079
57. Minhinnick S. Statistical reliability confounders and improvement in advanced dog training: patterns, routines, targets, alerts, distractors, reinforcement, and other issues. In: Jezierski T, Ensminger J, Papet LE, editors. *Canine Olfaction Science and Law*. Boca Raton: CRC Press; Taylor and Francis Group (2016). p. 197–212.
58. Scientific Working Group on Dog and Orthogonal detector Guidelines (SWGDOG). *SC1-abcdefghijk-Terminology.* (2011). Available online at: http://swgdog.fiu.edu/approved-guidelines; Retrieved January 21, 2019, from http://swgdog.fiu.edu/approved-guidelines
59. Mazur JE. *Learning and Behavior*. Upper Saddle River, NJ: Pearson Prentice Hall (2006).
60. Furton KG, Myers LJ. The scientific foundation and efficacy of the use of canines as chemical detectors for explosives. *Talanta.* (2001) 54:487–500. doi: 10.1016/S0039-9140(00)00546-4
61. Krichbaum S, Rogers B, Cox E, Waggoner LP, Katz JS. Odor span task in dogs (*Canis familiaris*). *Anim Cogn.* (2020) 23:571–80. doi: 10.1007/s10071-020-01362-7
62. Katz J, Wright A, Bodily K. Issues in the comparative cognition of abstract-concept learning. *Comp Cogn Behav Rev.* (2007) 2:79–92. doi: 10.3819/ccbr.2008.20005
63. Bhagavan S, Smith BH. Olfactory conditioning in the honey bee, *Apis mellifera*: effects of odor intensity. *Physiol Behav.* (1997) 61:107–17. doi: 10.1016/S0031-9384(96)00357-5
64. Cleland TA, Narla VA, Boudadi K. Multiple learning parameters differentially regulate olfactory generalization. *Behav. Neurosci.* (2009) 123:26–35. doi: 10.1038/jid.2014.371
65. Ghirlanda S, Enquist M. A century of generalization. *Anim Behav.* (2003) 66:15–36. doi: 10.1006/anbe.2003.2174
66. Oxley J, Waggoner L. Detection of explosives by dogs. In: Marshell M, Oxley J, editors. *Aspects of Explosives Detection.* Amsterdam: Elsevier Science (2009). p. 27–40.
67. Fischer-Tenhagen C, Johnen D, Heuwieser W, Becker R, Schallschmidt K, Nehls I. Odor perception by dogs: evaluating two training approaches for odor learning of sniffer dogs. *Chem Senses.* (2017) 42:435–41. doi: 10.1093/chemse/bjx020
68. Lazarowski L, Dorman DC. Explosives detection by military working dogs: olfactory generalization from components to mixtures. *Appl Anim Behav Sci.* (2014) 151:84–93. doi: 10.1016/j.applanim.2013.11.010
69. Wright HF, Wilkinson A, Croxton RS, Graham DK, Harding RC, Hodkinson HL, et al. Animals can assign novel odours to a known category. *Sci Rep.* (2017) 7:3–8. doi: 10.1038/s41598-017-09454-0
70. Hall NJ, Wynne CDL. Canine olfactory learning and behavior. In: Jezierski T, Ensminger J, Papet LE, editors. *Canine Olfaction Science and Law.* CRC Press; Taylor and Francis Group (2016). p. 123–37. doi: 10.1201/b20027
71. Nicklin S. *Why Use Dogs?* United Kingdom Defence Evaluation and Research Agency (2019).
72. Centre for the Protection of National Infrastructure. *Why and How to Control Contamination*. Detection Dog Guidance Notes (2018). Available online at: https://www.cpni.gov.uk/canine-detection-guidance-notes
73. Centre for the Protection of National Infrastructure. *Using Blanks and Interferents to Ensure Effective Detection Dog Training*. Detection Dog Guidance Notes (2018). Available online at https://www.cpni.gov.uk/canine-detection-guidance-notes
74. Kaulfuß P, Mills DS. Neophilia in domestic dogs (*Canis familiaris*) and its implication for studies of dog cognition. *Anim Cogn.* (2008) 11:553–6. doi: 10.1007/s10071-007-0128-x
75. Gazit I, Terkel J. Domination of olfaction over vision in explosives detection by dogs. *Appl Anim Behav Sci.* (2003) 82:65–73. doi: 10.1016/S0168-1591(03)00051-0
76. Wells DL, Hepper PG. Directional tracking in the domestic dog, *Canis familiaris*. *Appl Anim Behav Sci.* (2003) 84:297–305. doi: 10.1016/j.applanim.2003.08.009
77. Neitz J, Geist T, Jacobs GH. Color vision in the dog. *Visual Neurosci*. (1989) 3:119–25.
78. Jacobs GH, Deegan JF, Crognale MA, Fenwick JA. Photopigments of dogs and foxes and their implications for canid vison. *Visual Neurosci.* (1993) 10:173–80.
79. Rosengren A. Experiments on colour discrimination in dogs. *Acta Zool Fennica*. (1969) 121:3–19.
80. Byosiere SE, Chouinard PA, Howell TJ, Bennett PC. The effects of physical luminance on colour discrimination in dogs: a cautionary tale. *Appl Anim Behav Sci*. (2019) 212:58–65.
81. Moser AY, Brown WY, Bizo LA, Andrew NR, Taylor MK. Biosecurity dogs detect live insects after training with odor-proxy training aids: scent extract and dead specimens. *Chem Senses.* (2020) 45:179–86. doi: 10.1093/chemse/bjaa001
82. Williams M, Johnston JM. Training and maintaining the performance of dogs (*Canis familiaris*) on an increasing number of odor discriminations in a controlled setting. *Appl Anim Behav Sci.* (2002) 78:55–65. doi: 10.1016/S0168-1591(02)00081-3
83. Gazit I, Goldblatt A, Terkel J. The role of context specificity in learning: the effects of training context on explosives detection in dogs. *Anim Cogn.* (2005) 8:143–50. doi: 10.1007/s10071-004-0236-9
84. Partyka ML, Bond RF, Farrar J, Falco A, Cassens B, Cruse A, et al. Quantifying the sensitivity of scent detection dogs to identify fecal contamination on raw produce. *J Food Prot.* (2014) 77:6–14. doi: 10.4315/0362-028X.JFP-13-249
85. Lazarowski L, Foster ML, Gruen ME, Sherman BL, Fish RE, Milgram NW, et al. Olfactory discrimination and generalization of ammonium nitrate and structurally related odorants in Labrador retrievers. *Anim Cogn.* (2015) 18:1255–65. doi: 10.1007/s10071-015-0894-9
86. Angle C, Waggoner LP, Ferrando A, Haney P, Passler T. Canine detection of the volatilome: a review of implications for pathogen and disease detection. *Front Vet Sci.* (2016) 3:1–7. doi: 10.3389/fvets.2016.00047
87. Buszewski B, Ligor T, Jezierski T, Wenda-Piesik A, Walczak M, Rudnicka J. Identification of volatile lung cancer markers by gas chromatography-mass spectrometry: comparison with discrimination by canines. *Anal Bioanal Chem.* (2012) 404:141–6. doi: 10.1007/s00216-012-6102-8

88. Furton KG, Hong Y, Cheng Hsu YL, Luo T, Rose S, Walton J. Identification of odor signature chemicals in cocaine using solid-phase microextraction-gas chromatography and detector-dog response to isolated compounds spiked on U.S. paper currency. *J Chromatogr Sci.* (2002) 40:147–55. doi: 10.1093/chromsci/40.3.147
89. Harper RJ, Almirall JR, Furton KG. Identification of dominant odor chemicals emanating from explosives for use in developing optimal training aid combinations and mimics for canine detection. *Talanta.* (2005) 67:313–27. doi: 10.1016/j.talanta.2005.05.019
90. Johnston JM, Williams M, Waggoner LP, Edge CC, Dugan RE, Hallowell SF. Canine detection odor signatures for mine-related explosives. *Detect Remed Technol Mines Minelike Targets III.* (1998) 3392:490. doi: 10.1117/12.324221
91. Lorenzo N, Wan T, Harper RJ, Hsu YL, Chow M, Rose S, et al. Laboratory and field experiments used to identify *Canis lupus* var. *Anal Bioanal Chem.* (2003) 376:1212–24. doi: 10.1007/s00216-003-2018-7
92. Macias MS, Guerra-Diaz P, Almirall JR, Furton KG. Detection of piperonal emitted from polymer controlled odor mimic permeation systems utilizing *Canis familiaris* and solid phase microextraction–ion mobility spectrometry. *Forensic Sci Int.* (2010) 195:132–8. doi: 10.1016/j.forsciint.2009.12.006
93. Simon AG, Mills DEK, Furton KG. Chemical and canine analysis as complimentary techniques for the identification of active odors of the invasive fungus, *Raffaelea lauricola*. *Talanta.* (2017) 168:320–8. doi: 10.1016/j.talanta.2017.03.015
94. Dindal AB, Buchanan MV, Jenkins RA, Bayne CK. Determination of cocaine and heroin vapor pressures using commercial and illicit samples. *Analyst.* (2000) 125:1393–6. doi: 10.1039/b004298f
95. Rendine M, Fiore C, Bertozzi G, De Carlo D, Filetti V, Fortarezza P, et al. Decomposing human blood: canine detection odor signature and volatile organic compounds. *J. Forensic Sci.* (2019) 64:587–92. doi: 10.1111/1556-4029.13901
96. Verheggen F, Perrault KA, Megido RC, Dubois LM, Francis F, Haubruge E, et al. The odor of death: an overview of current knowledge on characterization and applications. *Bioscience.* (2017) 67:600–13. doi: 10.1093/biosci/bix046
97. DeGreeff LE, Peranich K, Simon AG. *Detection of Ammonium Nitrate Variants by Canine: A Study of Generalization Between Like Substances.* (No. NRL/MR/6181–18-9791). Naval Research Laboratory, Washington, DC (2018).
98. Lotspeich E, Kitts K, Goodpaster J. Headspace concentrations of explosive vapors in containers designed for canine testing and training: theory, experiment, and canine trials. *Forensic Sci Int.* (2012) 220:130–4. doi: 10.1016/j.forsciint.2012.02.009
99. Alhaddad AA, Salman M, Alyahya F. Factors affecting the rate of evaporation of the oil spilled on land during the gulf war. *Environ Technol.* (1995) 16:559–68. doi: 10.1080/09593331608616297
100. Harris BFS, Robinson JS, Agronomy F. Factors affecting the evaporation of moisture from the soil. *J Agric Res.* (1916) VII:439–61.
101. Legg BJ, Long IF. Microclimate factors affecting evaporation and transpiration. In: Hadas A, Swartzendruber D, Rijtema PE, Fuchs M, Yaron Y, editors. *Physical Aspects of Soil Water and Salts in Ecosystems. Ecological Studies (Analysis and Synthesis)*. Berlin; Heidelberg: Springer (1973). p. 275–85.
102. Degreeff LE. *The Chemistry of Odor and Odor Detection*. Chicago, IL: HITS K9 Seminar (2019).
103. Giordano BC, Lubrano AL, Field CR, Collins GE. Dynamic headspace generation and quantitation of triacetone triperoxide vapor. *J Chromatogr A.* (2014) 1331:38–43. doi: 10.1016/j.chroma.2014.01.045
104. Furton KG, Caraballo NI, Cerreta MM, Holness HK. Advances in the use of odour as forensic evidence through optimizing and standardizing instruments and canines. *Philos Trans R Soc Biol Sci.* (2015) 370:1674. doi: 10.1098/rstb.2014.0262
105. Moore S, MacCrehan W, Schantz M. Evaluation of vapor profiles of explosives over time using ATASS (Automated Training Aid Simulation using SPME). *Forensic Sci Int.* (2011) 212:90–5. doi: 10.1016/j.forsciint.2011.05.019
106. Rice S, Koziel JA. The relationship between chemical concentration and odor activity value explains the inconsistency in making a comprehensive surrogate scent training tool representative of illicit drugs. *Forensic Sci Int.* (2015) 257:257–70. doi: 10.1016/j.forsciint.2015.08.027
107. Becker F, Markee J, King JE. Studies on olfactory acuity in dogs. (1) Discriminatory behavior in problem box situations *Br J Anim Behav.* (1957) 5:94–103.
108. Grate JW, Ewing RG, Atkinson DA. Vapor-generation methods for explosives-detection research. *TrAC Trends Anal Chem.* (2012) 41:1–14. doi: 10.1016/j.trac.2012.08.007
109. King JE, Becker RF, Markee JE. Studies on olfactory discrimination in dogs: (3) ability to detect human odour trace. *Anim Behav.* (1964) 12:311–5. doi: 10.1016/0003-3472(64)90017-X
110. Stockham R, Slavin D, Kift W. Specialized use of human scent in criminal investigations. *Forensic Sci Commun.* Washington, DC: FBI (2004) 6.
111. Vyplelová P, Vokálek V, Pinc L, Pacáková Z, Bartoš L, Santariová M, Capková Z. Individual human odor fallout as detected by trained canines. *Forensic Sci Int.* (2014) 234:13–5. doi: 10.1016/j.forsciint.2013.10.018
112. Moulton DG. The use of animals in olfactory research. In: Gay WI, editor. *Methods of Animal Experimentation Volume IV: Environemnt and the Special Senses*. Academic Press (1973). p. 143–223.
113. Hallowell SF, Fischer DS, Brasher JD, Malone RL, Gresham GL, Rae C. Effectiveness of quality-control aids in verifying K-9-team explosive detection performance. *Chem Biol Based Technol Contraband Detect.* (1997) 2937:227–34. doi: 10.1117/12.266782
114. DeGreeff LE. *Elucidation and Durability of Odor Profiles of K9 Nose Work Materials*. (Report No.6180/0306). Naval Research Lab, Washington, DC (2020).
115. Papet LE. Narcotics and explosive odors: volatile organic compounds as training aids for olfactory detection. In: Jezierski T, Ensminger J, Papet LE, editors. *Canine Olfaction Science and Law*. CRC Press; Taylor and Francis Group (2016). p. 265–78.
116. Furton KG, Hsu Y-L, Luo T-Y, Norelus A, Rose S. Field and laboratory comparison of the sensitivity and reliability of cocaine detection on currency using chemical sensors, humans, K-9s, and SPME/GC/MS/MS analysis. *Invest Forensic Sci Technol.* (1999) 3576: 41–6. doi: 10.1117/12.334543
117. Alexander MB, Hodges TK, Bytheway J, Aitkenhead-Peterson JA. Application of soil in forensic science: residual odor and HRD dogs. *Forensic Sci Int.* (2015) 249:304–13. doi: 10.1016/j.forsciint.2015.01.025
118. Rust LT, Nizio KD, Wand MP, Forbes SL. Investigating the detection limits of scent-detection dogs to residual blood odour on clothing. *Forensic Chem.* (2018) 9:62–75. doi: 10.1016/j.forc.2018.05.002
119. Poziomek EJ, Almeer SH. Surface adsorption and retention of TNT vapors. In: *Enforcement and Security Technologies,* Vol. 3575. International Society for Optics and Photonics (1998). p. 392–402.
120. Degreeff LE, Furton KG. Collection and identification of human remains volatiles by non-contact, dynamic airflow sampling and SPME-GC/MS using various sorbent materials. *Anal Bioanal Chem.* (2011) 401:1295–307. doi: 10.1007/s00216-011-5167-0
121. Hudson-Holness DT, Furton KG. Comparison between human scent compounds collected on cotton and cotton blend materials for SPME-GC/MS analysis. *J Forensic Res.* (2010) 1:1–6. doi: 10.4172/2157-7145.1000101
122. Phelan JM, Webb SW. Chemical sensing for buried landmines: fundamental processes influencing trace chemical detection. In: *Mine Detection Dogs: Training, Operations, and Odour Detection*. Geneva: Geneva International Centre for Humanitarian Demining (2003). p. 209–86.
123. Goth A, McLean IG, Trevelyan J. How do dogs detect landmines? A summary of research results. In: *Mine Detection Dogs: Training, Operations, and Odour Detection.* Geneva: Geneva International Centre for Humanitarian Demining (2003). p. 43–52.
124. Jenkins TF, Leggett DC, Ranney TA. *Vapor Signatures From Military Explosives: Part 1. VAPOR TRANSPORT FROM BURIED MILITARY-GRADE TNT*. Ft. Belvoir, VA: U.S. Army Corps of Engineers, Cold Regions Research and Engineering Laboratory (1999).
125. George V, Jenkins TF, Leggett DC, Cragin JH, Phelan J, Oxley J, et al. Progress on determing the vapor signature of a buried landmine. In: *SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets IV*. Orlando, FL (1999).

126. Torres A, Padilla I, Hwang S. Physical modeling of 2,4-DNT gaseous diffusion through unsaturated soil. In: *SPIE Vol.* 6553 (2007). doi: 10.1117/12.723452
127. Leggett DC, Cragin JH, Jenkins TH, Ranney T. *Release of Exploisves-Related Vapors From Land Mines*. Hanover, NH: U.S. Army Corp of Engineers (2001).
128. Sargisson RJ, Mclean IG, Brown J, Bach H. Environmental determinants of landmine detection by dogs : findings from a large-scale study in Afghanistan. *J ERW Mine Act.* (2012) 16:74–81. doi: 10.13140/2.1.1505.8882
129. Alexander MB, Hodges TK, Wescott DJ, Aitkenhead-Peterson JA. The effects of soil texture on teh ability of human remains detection dogs to detect buried human remains. *J Forensic Sci.* (2016) 61:649–55.
130. Osterkamp T. *Detector Dogs and Scent Movement*. Boca Raton, FL: CRC Press (2020).
131. Baez B, Correa SN, Hernandez-Rivera SP, de Jesus M, Castro ME, Mina N, et al. Transport of explosives I: TNT in soil and its equilibrium vapor. *Detect Remed Technol Mines Minelike Targets IX.* (2004) 5415:1389. doi: 10.1117/12.542905
132. Steinkamp FL, DeGreeff LE, Collins GE, Rose-Pehrsson SL. Factors affecting the intramolecular decomposition of hexamethylene triperoxide diamine and implications for detection. *J Chromatogr A.* (2016) 1451:83–90. doi: 10.1016/j.chroma.2016.05.013
133. Peranich K, DeGreeff LE, Cerreta M. *Using Pressed and Variable Ratio Ammonium Nitrate - Aluminum to Train Explosive Detection Canines: Limited Objective Experiment.* Indian Head, MD: Naval Surface Warfare Center; Indian Head EOD Technology Division (2016).
134. United States Police Canine Association. *Certification Rules and Regulations.* (2014). Available online at: www.uspcaK9.com; Retrieved April 20, 2020 from http://uspca2.com/wp-content/uploads/2014/05/USPCARulebook2014.pdf

Conflict of Interest: AS was employed by the company AGS Forensics, LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 22 July 2020
doi: 10.3389/fvets.2020.00508

Efficacy of Dog Training With and Without Remote Electronic Collars vs. a Focus on Positive Reinforcement

*Lucy China, Daniel S. Mills and Jonathan J. Cooper**

Animal Behaviour, Cognition and Welfare Research Group, School of Life Sciences, University of Lincoln, Lincoln, United Kingdom

We assessed the efficacy of dog training with and without remote electronic collars compared to training with positive reinforcement. A total of 63 dogs with known off-lead behavioral problems such as poor recall were allocated to one of three training groups (each $n = 21$), receiving up to 150 min of training over 5 days to improve recall and general obedience. The 3 groups were: E-collar—manufacturer-nominated trainers who used electronic stimuli as part of their training program; Control 1—the same trainers following practices they would apply when not using electronic stimuli; and Control 2—independent, professional trainers who focused primarily on positive reinforcement for their training. Data collection focused on dogs' response to two commands: "Come" (recall to trainer) and "Sit" (place hindquarters on ground). These were the two most common commands used during training, with improving recall being the target behavior for the subject dogs. Measures of training efficacy included number of commands given to elicit the response and response latency. Control 2 achieved significantly better responses to both "Sit" and "Come" commands after a single instruction in the allocated time. These dogs also had shorter response latencies than the E-collar group. There was no significant difference in the proportion of command disobeyed between the three groups, although significantly fewer commands were given to the dogs in Control 2. There was no difference in the number of verbal cues used in each group, but Control 2 used fewer hand and lead signals, and Control 1 made more use of these signals than E-collar group. These findings refute the suggestion that training with an E-collar is either more efficient or results in less disobedience, even in the hands of experienced trainers. In many ways, training with positive reinforcement was found to be more effective at addressing the target behavior as well as general obedience training. This method of training also poses fewer risks to dog welfare and quality of the human-dog relationship. Given these results we suggest that there is no evidence to indicate that E-collar training is necessary, even for its most widely cited indication.

Keywords: dog training, dog welfare, electronic collar, reinforcement, punishment

OPEN ACCESS

Edited by:
Nathaniel James Hall,
Texas Tech University, United States

Reviewed by:
Erica Nan Feuerbacher,
Virginia Tech, United States
Alexandra Protopopova,
University of British Columbia, Canada

***Correspondence:**
Jonathan J. Cooper
jcooper@lincoln.ac.uk

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences, a section of the journal Frontiers in Veterinary Science

Received: *31 March 2020*
Accepted: *03 July 2020*
Published: *22 July 2020*

Citation:
China L, Mills DS and Cooper JJ (2020) Efficacy of Dog Training With and Without Remote Electronic Collars vs. a Focus on Positive Reinforcement. Front. Vet. Sci. 7:508. doi: 10.3389/fvets.2020.00508

INTRODUCTION

Successful obedience training of dogs requires effective use and timing of cues (often referred to as "signals") alongside reinforcement and/or punishment by dog trainers. Where dog training involves aversive or noxious stimuli, this can lead to punishment if dogs do not behave as desired (1, 2). A growing understanding of the application of learning theory to dog welfare has led many training organizations, welfare charities and academics to advocate what they consider to be more humane methods, with a greater focus on the use and timing of rewards (3–9).

Electronic training aids take a number of forms, but they commonly involve a collar-born device (E-collar) which can deliver a static electric stimulus to the dog's neck as well as a number of other stimuli, such as auditory or haptic/vibration signals (10). Collar-born devices include: remote, hand-operated devices; bark- or noise-activated control collars; and containment systems (or invisible fencing) (11). Generally, collars are designed to allow the auditory/haptic signals to be paired with the delivery of the electric stimulus as a form of "warning" cue. If the dog ignores this, the electric stimulus may be applied until the desired behavior is performed. In this way dogs may learn through a combination of negative reinforcement and classical conditioning to avoid the electric stimulus by performing the desired response, however, if the delivery of electric stimuli is poorly timed or inescapable, then undesirable associations may be formed (11–13). Opponents of E-collars have argued that because these devices use aversive stimuli to deter undesirable behavior, they pose an increased risk of undesirable training outcomes (such as negative changes in affective state or unanticipated associations) compared to reward-focused training, especially in the hands of poorly trained or inexperienced owners (14–18). In contrast, those who advocate the use of remote E-collars have argued that the devices, especially in the hands of experienced trainers, can be used as to modify behavior through negative reinforcement, with limited exposure to noxious stimuli, so are a valuable training aid. Collar manufacturers suggest that an advantage of these systems is that they give handlers control over a dog even at distance (19), and effectively suppress highly motivated behaviors, such as predatory behavior; a cause of livestock chasing or unintentional killing of wildlife (20–23). It has also been claimed that where E-collars are successful in treating behavioral problems, dogs may avoid unnecessarily euthanasia, an outcome that would be distressing to the owners (24).

The use of E-collars in dog training appears to be declining in the UK, from an estimate of 6% of all pet dogs in 2012 (25) to around 1% in 2019 (26). This decrease may reflect current government policy on the devices in Wales [devices banned under the (27)] and Scotland [not condoned in dog training and use may lead to punishment (28)], with restrictive legislation proposed for England (29) as well as high-profile campaigns against their use [e.g., by the (18)]. Nevertheless, these figures while appearing relatively low still suggest about 100,000 dogs in the UK are subject to E-collar use, and these devices remain legal in many other countries.

Research studies are cited selectively by both advocates and opponents of E-collars to support their claims, often with insufficient appreciation of the quality of experimental design or with a biased evaluation of evidence, such as the multiple possible interpretations of isolated behavioral indicators of welfare (11). However, the necessity of these devices [which has been used to justify their continued use e.g. (30)] depends on their efficacy compared to other training approaches (11, 31). Efficacy can be assessed objectively using specific target behavioral measures, and the use of professionally designed regimes delivered by experienced trainers can reduce the risk of sampling bias. In the current study we directly assessed the efficacy of the use of electronic collars to improve recall (the target behavior) and general obedience in dogs compared to training without E-collars. Dogs used in this study were referred to experienced, professional trainers as their owners had been experiencing significant obedience problems, including poor recall, but also chasing livestock and/or aggressive behavior to other dogs. The current study focussed on remote, hand-operated devices, as these were the most commonly used form in the UK at time of study (25, 32); being primarily used as a means of discouraging chasing behavior and improving recall. We used training records collected during DEFRA funded research (33) on behalf of the UK government. In contrast to the previously published work (31), where efficacy was assessed by owner feedback, this study recorded the speed and reliability of response after each command, in order to derive a more rigorous, systematic and objective measure of efficacy.

MATERIALS AND METHODS

Data were extracted from dog training videos, which were originally recorded as part of a DEFRA funded study (33) collected over a period of 6 months in 2010/11. Details of the recruitment of dogs, the training regimes and video data collection have previously been published (31, 33), so the methods presented here provide an overview with additional details of differences in the approach taken in the current study.

Ethical approval was provided by the University of Lincoln Research Ethics Committee. Owners and trainers that participated in the study gave their informed consent for the use of their dogs and video recordings in the study. Home Office Inspectorate were consulted, and indicated that the work did not constitute a procedure and consequently a Home Office License would not be required based on the following conditions: E-collar use was legal in England and Scotland at the time of the study; dogs were being referred for behaviors commonly associated with E-collar use in the UK; the training was being conducted by experienced professional trainers using normal training programmes with the informed consent of owners.

Training Groups

All dogs used in this study had been referred for behavioral concerns including poor recall and livestock worrying and owners had been recommended to seek professional training to resolve those problems. The 63 dogs involved in the study were all older than 9 months of age and had no prior experience with electronic collars. Dogs in E-collar and Control Group 1 were trained in Autumn/Winter 2010 and were randomly allocated to their training group. Dogs in Control Group 2 were trained in Spring 2011, meaning subjects could be recruited to match the dogs trained with E-collars on the basis of referred behavioral problem and owner's assessment of severity. The 3 training groups were as follows:

- E-collar Group (EC: $n = 21$): dogs were trained using active electronic collars to improve recall and general obedience by experienced, manufacturer-nominated trainers (ECMA) (chosen to represent best-practice use of the E-collar). Trainers followed approved practice as recommended by ECMA, including assessing the dog's sensitivity to electric stimuli prior to training, and pairing vibration cue with the electric signal with the aim of modifying behavior through negative reinforcement. Dogs in this group also experienced positive reinforcement, such as rewarding dogs with food and negative reinforcement such as lead pressure.
- Control Group 1 (C1: $n = 21$): dogs were trained by the same trainers who worked with the E-collar group, using a mix of food rewarded positive reinforcement and negative reinforcement such as lead pressure to improve recall and general obedience but without use of electronic stimuli;
- Control Group 2 (C2: $n = 21$): dogs were trained to improve recall and general obedience by experienced professional trainers who were members of Association of Pet Dog Trainers (APDT UK); an organization which does not support the use of E-collars in dog training (chosen to represent best-practice use of positive reinforcement or "reward-based training").

The dog population used in this study was broadly similar to the populations described by Blackwell et al. (25) in their survey of use of electronic training aids, and there were no significant differences between the dogs allocated to the three treatments in type of dog or reason for referral (31). Gundogs (25%), cross-breeds (25%), pastoral (17%) and terriers (13%) were the most commonly represented breed types with similar numbers in each treatment group, whereas there were no dogs from toy or utility breed groups. 34 (54%) dogs were female, with 21 (33%) of these neutered and 13 (21%) entire female dogs. Of the 29 male dogs, there was also a slightly higher number of neutered dogs (19, 30% of total population) than entire male dogs (10 or 16%); however there were no significant gender biases between the treatment groups. Chasing was the most common reason for referral in the study population (51 out of 63 dogs or 81% of population), representing 18 dogs in the E-Collar Group, 17 dogs in Control Group 1 and 16 dogs in Control Group 2. Sheep or lambs were the most commonly cited chase target, where owners reported chasing as a problem behavior, although owners also listed other livestock such as horses and poultry, wildlife such as rabbits or squirrels, as well as cars and joggers as targets for chasing. The remaining dogs had either been referred for poor recall (9 dogs of which 1 was in E-collar Group and 4 each in Control Group) or aggressive interactions with other dogs whilst off lead (3 dogs, 2 of which were in E-collar Group, 1 in Control Group 2 and none in Control Group 1). The majority of owners described their dogs as exhibiting the referred behavior "Always" (31 dogs or 49% of population), or "Frequently" (24 dogs or 38% of population indicating the high severity as perceived by owners. When these two ratings were pooled, there was no difference in owner assessment of severity between the three groups (31).

All dogs in the study wore an E-collar during training sessions in order for data analyzers to be blind to training group during video observation. Dogs in the Control groups wore a de-activated or "dummy" collar, whilst the e-collars worn by dogs in E-collar group were active and useable by the trainers. Training mainly occurred in field locations, with penned sheep, penned chickens and other (on lead) dogs, as potential distractors during training. Dogs were primarily kept on 10 m long leads throughout training session; however, trainers had the option to drop the lead or remove the lead from the dog when considered appropriate. During training dogs were normally within 1 m of the trainer (around 70% of time in all three groups) with <5% of time spent more than 5 m distant from trainer (in all three groups). Trainers in all groups had access to food rewards and could use them as the trainer deemed appropriate during training. Previous work (31) had indicated that whilst dogs in Control Group 2 received fewer signals per 15 min training session than dogs in E-collar Group or Control Group 1 (32 signals compared with 59 and 56, respectively), they were much more likely to receive food reward following a successful response, than dogs in Control Group 1 or E-collar group. Preliminary observations prior to this study to determine which commands were most common in the three groups confirmed these previous reports with food estimated to be used about 5 times more frequently as a reward during training by Control Group 2, than E-collar or Control Group 1 (34). This rate of reward would be consistent with the emphasis on reward based training in Control Group 2, compared to a mix of training approaches in the other treatment groups. Two training sessions were recorded daily (one in the morning and one in the afternoon) for each dog, for up to 5 consecutive days, producing an average of 28.5 ± 4.5 (mean ± SD) minutes of video record per dog per day, and up to 150 min over the 5 training days.

Data Collection

Data for the current study were taken from the two training sessions on the first, third and fifth day of training for each dog. Measures focused on indicators of efficacy and reliability of obeying command, including latency to complete response and number of commands required to complete desired response. Data collection focused on the two commands that were most commonly used in all 3 training groups, and could be easily distinguished from video data. These were a "Come" command normally used for recall of dogs when at a distance from the trainer and a "Sit" command normally used to require the dog to place its hind-quarters on ground and remain stationary for brief periods of time (See **Table 1**). "Come" and "Sit" commands were chosen for several reasons. Both commands could be clearly identified and obtained from the videos, across all groups, and could not be confused with other commands. During preliminary analysis of video records, these commands were also found to

TABLE 1 | Descriptions of the mode of delivery (or signals) for the Come and Sit commands.

Command given	Command type	Description
Come	Verbal	The dog is encouraged to return to the trainer/owner from distance upon the verbal "come" command; noises of encouragement given after this include clicking, whistling, kissing-sounds, etc.; related verbal expressions such as "let's go," "come on" etc and use of the dog's name
	Hand signal	The dog is encouraged to return to the trainer/owner from distance upon the visual hand signal of a beckoning motion from the arm and hand extended away from the body and the arm of hand is repeatedly drawn toward the body; may also be gestured by the patting of the trainer/owner's leg. May be accompanied by other more physical actions noted at the time
	Lead signal	The dog is encouraged to return to the trainer/owner from distance following a tug on the lead being toward the trainer/owner or the lead is flicked to bring the dog toward the trainer/owner. May be accompanied by other more physical actions noted at the time
Sit	Verbal	The dog is asked to place its rear end on the ground upon issuing of the "sit" command verbally
	Hand signal	The dog is asked to place its rear end on the ground by the hand of the trainer/owner being brought up toward the chest/center of the body or the trainer pointing their finger down over the dog's head. May be accompanied by other more physical actions noted at the time
	Lead signal	The dog is asked to place its rear end on the ground upon the lead being pulled vertically above the dog's head or down toward the ground. May be accompanied by other more physical actions noted at the time

be the most commonly used in all three training groups. Three forms of signal or mode of delivery of training signals were noted in preliminary observation; verbal, hand and lead, and these are also defined in context in **Table 1**. We also recorded: if dogs began the recall response after a single "Come" command (Come); if multiple commands (Come+) were used to initiate the recall response; or if the dog did not initiate the response (disobey; see **Table 2**). Similarly, we recorded: if a sit response was completed after a single signal (Sit); if multiple signals were needed (Sit+); or if the dog did not perform a Sit response to the "Sit" commands (disobey). In preliminary observations, to determine timeframes for these definitions of outcome, most dogs responded within 2 s of initial command, and where dogs were given additional commands, this was normally limited to 2 or occasionally 3 commands within the 10 s of the initial command. Where dogs had not completed the response within 10 s of initial command, trainers normally ceased this sequence of commands and after a brief rest normally longer than 10 s would resume with a new command. This approach was similar across the three training groups, so the definition of successful responses and disobey could be applied to all groups. To control for the different number of commands given, absolute values were converted into % of commands to compare reliability of response between the three groups. Where dogs responded to the "Come" command the latency was recorded as the time from delivery of first command to the dog initiating the recall response, whereas latency to sit was recorded as the time to place hind-quarters on ground following delivery of the first "Sit" command signal.

Data Extraction and Statistical Analysis

Training videos were viewed in a random order and blinded, such that the viewer could not associate dogs with their respective group, using Solomon Coder software (version: beta 17.03.22). Following collection, raw data was extracted from the Solomon Coder files into a Microsoft Excel spreadsheet, separating each dog into their allocated training group. Data for the number of commands (**Table 3**) were analyzed per training session. A small number of sessions focused on just recall or just sit, so morning and afternoon sessions were aggregated, for analysis of the percentage of dogs responding to the first signal, multiple signals, or disobeying and for the calculation of latencies. This provided a single daily measure for each dog. Previous work Cooper et al. (31) had indicated no significant differences in dogs' behavior between morning and afternoon sessions, and exploratory comparison of morning and afternoon data in this study was consistent with this.

Statistical analysis of the data was conducted using Minitab 17.0, using General Linear Models (GLMs). Training groups and days (1, 3, and 5) were treated as fixed factors, whilst individual dog IDs were random factors nested within the training groups. As the focus was on efficacy outcomes we focused on main effects and did not include interactions within our models, so as not to unnecessarily inflate the degrees of freedom in the models. Unless stated otherwise data is presented as mean ± standard error, since our focus was on differences between groups and not group variability.

RESULTS

Number of Commands, Signals, and Responses

On average 20.3 ± 0.6 commands were given per training session, of which 15.7 ± 0.6 (77%) were obeyed on first command, 4.1 ± 0.2 (20%) obeyed after multiple commands and only 0.6 ± 0.1 (3%) disobeyed. On average the number of signals per training session was 26.8 ± 0.8. The majority of signals were verbal with 17.8 ± 0.8 verbal signals per session (66% of all signals). There were 5.2 ± 0.3 hand signals per training session (19% of all signals) and 3.8 ± 0.4 lead signals (14%). There was no difference in the number of verbal signals given to dogs in the 3 training groups (**Table 3**), but Control Group 1 consistently received more hand and lead signals than dogs trained with E-collars,

TABLE 2 | Description of dogs' responses to "Come" and "Sit" commands.

Command given	Command response	Description
Come	Obeys after first command	The dog correctly responds to the "come" command by taking steps, at any speed, toward the trainer/owner following the first instance of the command being given
	Obeys after repeated commands	The dog correctly responds to the "come" command by taking steps, at any speed, toward the trainer/owner following multiple instances of the command being given
	Disobey	The dog fails to appropriately respond to the "come" command, either by failing to move toward the trainer/owner or no correct response within 10 s of the first command thus acting as a cut-off point for the command
Sit	Obeys after first command	The dog correctly responds to the "sit" command by placing the hind quarters on the ground following the first instance of the command being given
	Obeys after repeated commands	The dog correctly responds to the "sit" command by placing the hind quarters on the ground following the first instance of the command being given
	Disobey	The dog fails to appropriately respond to the "sit" command, either by failing to place the hind quarters on the ground or no correct response occur within 10 s of the first command thus acting as a cut-off point for the command

TABLE 3 | Mean number of commands given per training session (±SE) for dogs trained with E-collars and the two control groups, including number of verbal, hand, and lead signals, number of times a single "Come" and "Sit" command were given and numbers of times multiple signals were given for each command (Come+ and Sit+) and the number of times dogs obeyed on first comment, obeyed after multiple commands (Obey+) or did not obey.

Command given	Mean ± Standard Error of commands given			F-Ratio from GLM
	E-Collar	Control 1	Control 2	
Verbal	16.5 ± 1.4	20.5 ± 1.6	16.6 ± 1.1	$F_{(2,293)} = 3.05$, $P = 0.051$
Hand	$5.4 \pm 0.4_a$	$8.9 \pm 0.7_b$	$1.6 \pm 0.2_c$	$F_{(2,293)} = 57.7$, $P < 0.001$
Lead	$4.2 \pm 0.5_a$	$7.5 \pm 1.0_b$	$0.1 \pm 0.0_c$	$F_{2,293)} = 39.6$, $P < 0.001$
Sit	$12.5 \pm 0.8_a$	$16.2 \pm 1.0_b$	$3.4 \pm 0.5_c$	$F_{(2,293)} = 69.2$, $P < 0.001$
Sit+	$3.4 \pm 0.4_a$	$5.5 \pm 0.6_b$	$0.6 \pm 0.1_c$	$F_{(2,293)} = 35.4$, $P < 0.001$
Come	$7.4 \pm 0.6_a$	$10.2 \pm 0.8_b$	$11.8 \pm 0.8_b$	$F_{(2,293)} = 8.92$, $P < 0.001$
Come+	$2.9 \pm 0.5_a$	$4.9 \pm 0.7_b$	$2.5 \pm 0.3_a$	$F_{(2,293)} = 6.84$, $P = 0.001$
Obey	$15.4 \pm 1.1_a$	$19.2 \pm 1.1_b$	$12.8 \pm 0.9_a$	$F_{(2,293)} = 8.78$, $P < 0.001$
Obey+	$4.1 \pm 0.3_a$	$6.3 \pm 0.5_b$	$2.0 \pm 0.2_a$	$F_{(2,293)} = 37.5$, $P < 0.001$
Disobey	$0.4 \pm 0.1_a$	$1.0 \pm 1.1_b$	0.4 ± 0.1a	$F_{(2,293)} = 9.50$, $P < 0.001$

Different subscripts (a, b, and c) indicate where training groups differed based on Tukey pair-wise comparisons.

whilst Control Group 2 consistently had fewer hand and lead signals than the other groups. As a consequence, Control Group 1 received most signals during training, whilst Control Group 2 received fewer signals during the training period than the other groups [$F_{(2, 293)} = 30.2$, $P < 0.001$].

Control Group 2 performed fewer "Sit" responses during training than the E-collar group and Control Group 1, following single commands, following use of multiple commands (**Table 3**) and overall [$F_{(2, 293)} = 74.5$, $P < 0.001$]. Control Group 1 performed the most "Sit" and most "Come" responses following multiple commands, whilst the E-collar Group performed least "Come" responses following a single command and in total [$F_{(2, 293)} = 5.51$, $P = 0.005$].

Control Group 1 exhibited more disobeys than either the E-collar training group or Control Group 2 (**Table 3**), but also completed more responses after single and multiple commands as they received most commands of the three training groups. When the percentage of responses was analyzed to account for the different number of commands between the training groups, there was no difference in percentage of disobeys between the three training groups (**Table 4**). Control Group 2, however, had a higher percentage of performing both "Come" and "Sit" responses on first command and lower percentage following multiple commands than either Control Group 1 or the E-collar Group.

Training day had no effect on number of commands or response rate, except for the use of signals (**Figures 1–3**). Use of lead signals declined from day 1 to day 5 [**Figure 3**: $F_{(2,293)} = 17.5$, $p < 0.001$] and use of hands signals was most common on day 3 [**Figure 2**; $F_{(2,293)} = 4.04$, $p = 0.018$]. There was however no change in number of verbal signals used over the training days, and overall total number of signals used did not differ across training days [$F_{(2,293)} = 0.16$, $P = 0.85$].

Latency to Respond

Overall, the mean latency to respond to the "Come" command was 1.24 ± 0.05 s, whereas dogs took a mean of 1.64 ± 0.06 s to complete the "Sit" commands. There were significant differences in latency to respond to both the "Come" [$F_{(2,114)} = 5.89$; $p = 0.04$] and the "Sit" command [$F_{(2,101)} = 12.3$; $P < 0.001$]

TABLE 4 | Mean percentage of "Come" and "Sit" commands (± SE) obeyed after a single signal, obeyed after multiple signals (Obey+) or not obeyed for dogs trained with E-collars and the two control groups.

Percentage	Mean ± Standard Error of commands given			*F*-Ratio from GLM
	E-Collar	Control 1	Control 2	
% Obey Come	$71.0 \pm 3.2_a$	$72.4 \pm 2.7_a$	$82.5 \pm 2.3_b$	$F_{(2,114)} = 4.46, P = 0.015$
% Obey+ Come	$26.3 \pm 2.8_a$	$24.4 \pm 2.4_a$	$15.4 \pm 2.2_b$	$F_{(2,114)} = 4.33, P = 0.017$
% Disobey Come	2.7 ± 0.1	3.2 ± 0.01	2.1 ± 0.01	$F_{(2,114)} = 0.66, P = 0.52$
% Obey Sit	$76.8 \pm 2.8_a$	$72.7 \pm 2.7_a$	$83.5 \pm 3.2_b$	$F_{(2,114)} = 3.49, P = 0.036$
% Obey+ Sit	$18.9 \pm 2.0_a$	$21.9 \pm 2.1_a$	$10.6 \pm 2.1_b$	$F_{(2,114)} = 6.69, P = 0.002$
% Disobey Sit	4.4 ± 0.2	5.4 ± 0.2	3.7 ± 0.1	$F_{(2,114)} = 0.36, P = 0.70$

Different subscripts (a and b) indicate where training groups differed based on Tukey pair-wise comparisons.

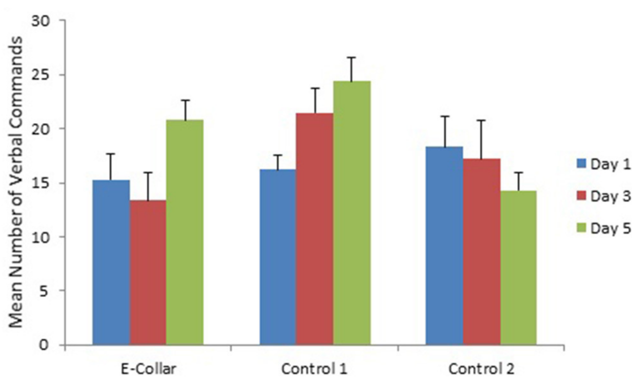


FIGURE 1 | The mean (with SE) number of verbal commands given to dogs in the E-collar training group and the two Control groups over the 3 training days. See **Tables 3**, **4** for analysis of differences between groups; no significant differences between training days.

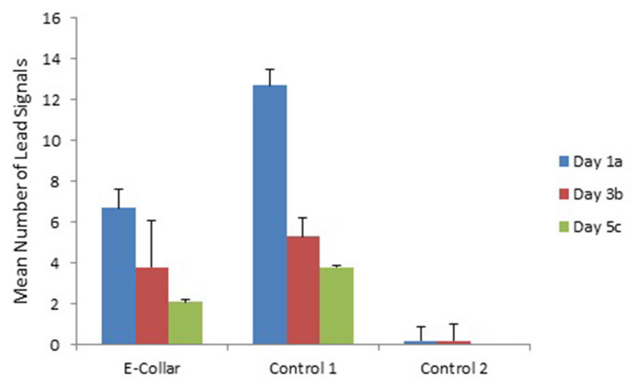


FIGURE 3 | The mean (with SE) number of lead commands given to dogs in the E-collar training group and the two Control groups over the 3 training days. Subscripts (a, b, and c) indicate where training days differed based on Tukey pair-wise comparisons. See **Tables 3**, **4** for analysis of differences between groups.

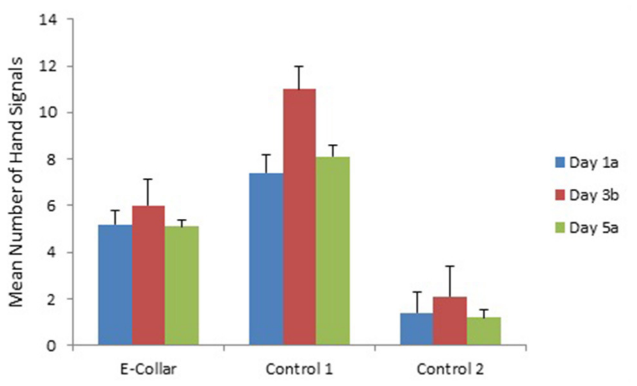


FIGURE 2 | The mean (with SE) number of hand commands given to dogs in each training group over the 3 training days. Subscripts (a and b) indicate where training days differed based on Tukey pair-wise comparisons. See **Tables 3**, **4** for analysis of differences between groups.

between the training groups (**Table 5**). For the "Come" command there was a shorter latency to respond by Control Group 2 compared with the E-collar Group. The difference in latency to respond to the "Sit" command was largely similar to that of the "Come" command, however Control Group 2 responded sooner than both the E-collar Group and Control Group 1.

Although the E-collar Group and Control Group 1 appeared to show a decline in latency to respond to the "Come" command over the study period (**Figure 4**) there was no significant change in latency to come between the 3 training days [$F_{(2,114)} = 1.82$; $P = 0.17$]. In contrast there was a change in latency to sit [$F_{(2,101)} = 5.61$; $P = 0.005$] with longer latencies to sit on day 3 and day 5 compared to day 1 (**Table 5**), which was related to increased latency in the E-collar training group and Control Group 1, as training progressed (**Figure 5**).

DISCUSSION

Each of the three training groups had successful training outcomes to both "Come" and "Sit" commands. The proportion of responses that were performed following first command was high in all three groups, and the proportion of disobeys was low throughout the study and did not differ between training groups. These findings are consistent with owner satisfaction

TABLE 5 | Mean latency to complete response in seconds from initial command (± standard deviation) for those dogs that completed come and sit responses from the E-collar and 2 Control training groups and on different days.

Command given	Latency to respond			GLM
	E-Collar	**Control 1**	**Control 2**	
Come	$1.35 \pm 0.11_a$	$1.24 \pm 0.09_{ab}$	$1.13 \pm 0.05_b$	$F_{(2,114)} = 5.89, P = 0.04$
Sit	$1.67 \pm 0.11_a$	$1.81 \pm 0.12_a$	$1.36 \pm 0.11_b$	$F_{(2,101)} = 12.3, P < 0.001$
	Day 1	**Day 3**	**Day 5**	
Come	1.26 ± 0.10	1.30 ± 0.08	1.14 ± 0.08	$F_{(2,114)} = 1.82, P = 0.17$
Sit	$1.44 \pm 0.10_a$	$1.72 \pm 0.11_b$	$1.76 \pm 0.15_b$	$F_{(2,101)} = 5.61, P = 0.005$

Different subscripts (a and b) indicate where training groups differed based on Tukey pair-wise comparisons.

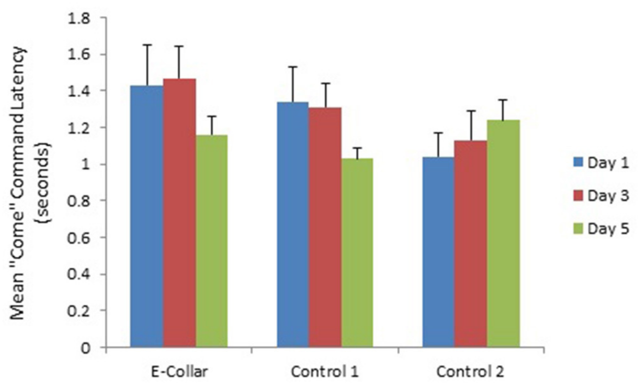


FIGURE 4 | The mean (with SE) latency to respond to "Come" command by dogs in the E-collar training group and the two Control groups over the 3 training days.

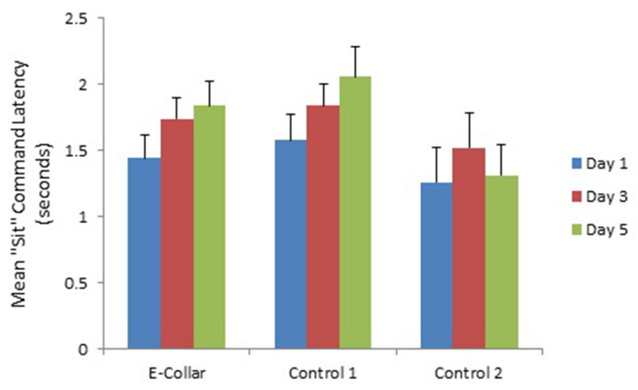


FIGURE 5 | The mean (with SE) latency to respond to "Sit" command by dogs in the E-collar training group and the two Control groups over the 3 training days.

with training outcomes as reported previously (31) and should be expected as all trainers were professionals, with extensive experience of training dogs to improve recall and general obedience. The reward-based Control Group 2, however, had a higher proportion of obeys after first command to both "Come" and "Sit" commands and required fewer multiple commands to initiate a recall or complete a sit response. This suggests that the reward-based training was the most effective approach not only for recall which was the target behavior in training, but also for other commands, even though the reward based trainers did not spend as much of their time training on sit command as the other two training groups.

Latencies to respond also indicate successful training outcomes in all three groups with dogs beginning to return to the trainer on average 1.24 s after delivery of a "Come" command and dogs completing the sit response on average 1.64 s after a "Sit" command. The slightly longer latency to sit potentially reflected this measure being based on completion of response, whereas latency to "Come" response was determined from the initiation of recall with dogs beginning to return to trainer. Although differences between groups were small, dogs in Control Group 2, showed a shorter latency to begin to return than the E-collar Group, which is consistent with the higher proportion of responses seen following a single command in this group. There were also differences between the groups in time to complete the sit response, with Control Group 2 being faster to complete this response than both the E-collar group and Control Group 1. This was also consistent with a higher proportion of dogs completing this response after a single command. It is noteworthy that there was little difference in latency to sit between the three groups on the first day of training, as dogs in all three groups had a reliable response to the "Sit" command before training, but longer latencies in the E-collar and Control 1 group become apparent as training progressed. These findings are consistent with the reported public perception that E-collars have lower success rates than reward-based training for recall and chase problems (25), and concerns regarding efficacy of training programs involving potentially aversive stimuli raised by Hiby et al. (3), Rooney and Cowan (4), Fernandes et al. (5), Ziv (6), and Masson et al. (7, 35).

Two factors apart from the use of electric stimuli during training should be explored before drawing conclusions with regard to the efficacy of the three training methods. The first relates to the weather conditions as E-collar Group and Control Group 1 were trained in mid-winter, whereas Control Group 2 were trained 4 months later in early Spring. This was in part due to availability of industry nominated trainers, but also allowed time to select dogs from the larger population available for training without E-collars to best match those referred to E-collar Group. Although there has been no published work on seasonal variation in training outcomes in dogs, there are likely to be variation in environmental conditions, that may impact on these outcomes. Indeed the winter training period in particular featured some extreme weather conditions with lying snow and low daytime temperatures as well as milder periods. For this reason, as part of the exploratory analysis of data during the original project (AW1402a), weekly variation in data were investigated in each group and no differences were found with respect to command use, dog behavior or training outcomes, suggesting weekly changes in environmental had minimal effects, and that trainers maintained consistent approaches to training over the weeks of data collection, despite the challenges of field conditions.

The second relates to differences in the general approaches to training between the three groups and in particular between Control Group 2 trainers and those in E-Collar and Control Group 1. Firstly, Control Group 2 appeared to primarily target recall training, with less time spent on other commands including sit, whereas the E-collar Group and Control Group 1 chose to work on both recall and general obedience including sitting (**Table 3**), perhaps indicating a greater focus on controlling the dog as well as achieving the target goal behavior. Furthermore, whilst the use of verbal signals was similar between the three groups, hand and in particular lead signals were less frequently used by Control Group 2 than either Control Group 1 or the E-collar Group; with Control Group 1 making more use of hand and lead signals during training than the E-collar Group. The use of multiple signals in training can have variable effects, with, for example, the use of additional contingencies such as lead pressure during a recall command, potentially affecting the rate of learning of the desired response. Improvement in learning would depend to some extent on the multiple signals being delivered consistently, and even then, dogs may form more reliable associations with some stimuli than others due to learning and perceptual biases or the nature of delivery. For example, it has been reported that visual signals during dog training may overshadow verbal ones when used at the same time (36). The explanation for differences in learning outcome may therefore lie in the degree to which dogs were exposed to rewarding and potentially aversive stimuli in the three groups and the range of signals used to guide the dogs' behavior.

Broadly speaking, dogs in Control Group 2 were asked to complete a recall task in response to verbal signals and normally received food reward(s) on return to trainer. Hand signals were rarely used and even though dogs were often on lead, lead pressure was very rarely recorded. As a consequence a single signal was used to cue the desired behavior and a single contingency (food) associated with successful completion of response. Similarly, training of a sit used a verbal "sit" command, with dogs receiving food reward once response had been completed. In summary, this group appeared to use the simplest and clearest contingencies for associative learning.

Dogs in the E-collar group were trained in accordance with industry best practice, with dogs' sensitivity to E-collar settings assessed early in training, and training focussed on associating the pre-warning cue, a collar born vibration, with exposure to the electric stimuli. In this way, the intensity of the electric stimulus could have been moderated to match the dog's tolerance and dogs could learn to modify their behavior to avoid exposure to the electric stimulus; a form of negative reinforcement. This sophisticated use of e-collars contrasts with that of some trainers reported in Cooper et al. (31), who used e-collars at their maximum settings and applied the electric stimulus after the dogs engaged in undesirable behavior, such as sheep chasing, without the use of the pre-warning cues. As buttons to deliver pre-warning cues were on same handset as the button for electric stimulus, it was not possible to reliably determine when electric stimuli were applied, so we should be cautious about inferring when stimuli were used during training schedules. For example, although one might predict that there would be more use of electric stimuli during early training as sensitivity is determined and an association formed between stimulus and desired response, or that electric stimuli would be more likely to be applied if the dog did not respond to initial command this cannot be verified from our data. For example, in previous published work (31), where vocalizations and abrupt changes in posture were recorded when dogs were remote from trainers, there was no evidence of change in frequency over 5 days of training. This freedom to adjust application of stimuli as part of the training program, as well as inclusion of other approaches to training the target or other behaviors, was consistent with the ethical approval of our project as well as our aim to assess best practice as advocated by the industry. Therefore, so long as dogs were not exposed to inescapable punishment, and trainers followed industry standards, we could not artificially impose standardized training programs, nor could we preclude trainers from using other signals and/or contingencies during training such as hand and lead signals. As a consequence, although we did not have the control over variables of experimental investigations of e-collar training [e.g., (37–39)], we did meet our aim of evaluating professional training of companion dogs with typically referred behaviors in the field.

Dogs in Control Group 1 were trained by the same trainers as the E-collar group and were expected to follow the same training approaches but without use of E-collar stimuli. Dogs in this group wore a de-activated dummy collar (as did dogs in Control Group 2) to control for the wearing of an unfamiliar device as well as part of the process of blinding observers to treatment in video analysis. As a consequence these dogs experienced collar fitting at start of each training session, but were not exposed to electric or vibration stimuli during training. These trainers therefore also used a mix of verbal, hand and lead signals, as the E-collar Group, but relatively few food rewards during training. It was also clear that the dogs received more lead and hand signals than the dogs in the E-collar group. Hand signals, involved not only hand gestures, but were also accompanied in some instances by physical contact with the dogs to gain their attention, stopping of ongoing behavior or pushing the dog into the desired position, whilst lead signals could be accompanied by what appeared to be sharp pulls on the lead. This more physical and potentially aversive use of contact or lead pressure was not observed in any of the videos relating to Control Group 2 but were clearly identified in both the E-collar Group and Control Group 1. These qualitative observations support the suggestion that the trainers involved in both the E-collar training and Control Group 1 were again more focussed on forcing compliance rather than shaping the desired response (40).

In summary, an important strategy within the reward-focused training of Control Group 2 was the positive reinforcement of successive approximations of the desired behavior, with mainly verbal signals, in order to build a strong contingency between command word and response (40). In contrast the E-collar group and Control Group 1 used a variety of signals and contingencies, including some potentially aversive handling and lead pressure during training. With good timing, these could result in negative reinforcement, although poor timing or imposition of the noxious stimuli in response to failure to perform the desired behavior would constitute a form of punishment. It has been frequently argued that the use of aversives in dog training results in poorer learning outcomes and poses greater welfare risks compared with largely reward based training (3–6). Our results demonstrate through direct evidence from real life situations, that the reward-focused training was, indeed, more efficient than methods which included potentially aversive stimuli such as electric stimuli or excessive lead pressure. Whilst our results may reflect general differences in training style of the trainer groups involved in the study rather than use of E-collar *per se*, we would argue that because the trainers who used E-collars were put forward by industry representatives as exemplars of best practice; their data (at least in relation to E-collar use) should be taken to represent a best case scenario for professional E-collar training. It is likely that less experienced trainers and owners would be less skilled and thus less effective in their use of the device [See (25, 35)].

Overall, the professional use of a reward-focused training regime, as demonstrated by Control Group 2, was superior to E-collar and Control Group 1 in every measure of efficacy where there was a significant difference. In addition, dogs in Control Group 1 showed no better learning outcomes than those in the E-collar group, indicating industry nominated trainers were as effective at modifying undesirable behavior, when they did not use e-collars as one of their training methods. Given the better target behavior response parameters associated with a reward-focused training programme, and the finding that the use of an E-collar did not create a greater deterrent for disobedience; we conclude that an E-collar is unnecessary for effective recall training. Given the additional potential risks to the animal's well-being associated with use of an E-collar (7, 25, 31, 38, 39), we conclude that dog training with these devices causes unnecessary suffering, due to the increased risk of a dog's well-being is compromised through their use, without good evidence of improved outcomes.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by University of Lincoln Research Ethics Committee. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

LC undertook video observation and behavioral coding of the training videos, as well as initial statistical analysis, and led the writing of the main article. JC and DM were LC's supervisors for her Master's thesis, providing support throughout. All authors made similar contributions to the final manuscript.

FUNDING

We would like to thank DEFRA for funding the original study (AW1402A). We would also like to thank University of Lincoln, who provided an alumni support bursary for LC's Master's by Research thesis.

ACKNOWLEDGMENTS

The authors wish to thank DEFRA for funding of original project (AW1402A) and ECMA and APDT for supporting the nomination of the best trainers available and support in recruitment of dogs through referrals. We would also thank the dog owners for participation in this study and volunteering their dogs. We would also like to acknowledge Hannah Wright and Jessica Hardiman, for working with trainers in the field and recording of the video records and Marie Delpech, Emma Cosby, Rachael Nicklin, and Molly Taylor who previously worked with LC on the development of methods for assessing efficacy of training methods for this study.

REFERENCES

1. Burch MR. *The Evolution of Modern-Day Dog Training and Obedience*. (2002). Available online at: http://www.naiaonline.org/articles/article/the-evolution-of-modern-day-dog-training#sthash.kMpTEvbM.vcAxIBoe.dpbs (accessed November 27, 2019).
2. Gray SA. *A Short History of Dog Training*. US: Animal Wellness Magazine (2018). Available online at: https://animalwellnessmagazine.com/history-dog-training/ (accessed November 29, 2019).
3. Hiby EF, Rooney NJ, Bradshaw JWS. Dog training methods: their use, effectiveness and interaction with behaviour and welfare. *Anim Welf*. (2004) 13:63–9.

4. Rooney NJ, Cowan S. Training methods and owner–dog interactions: links with dog behaviour and learning ability. *Appl Anim Behav Sci.* (2011) 132:169–77. doi: 10.1016/j.applanim.2011.03.007
5. Fernandes JG, Olsson IAS, de Castro ACV. Do aversive-based training methods actually compromise dog welfare? A literature review. *Appl Anim Behav Sci.* (2017) 196:1–12. doi: 10.1016/j.applanim.2017.07.001
6. Ziv G. The effects of using aversive training methods in dogs—a review. *J Vet Behav Clin Appl Res*. (2017) 19:50–60. doi: 10.1016/j.jveb.2017.02.004
7. Masson S, de la Vega S, Gazzano A, Mariti C, Pereira GDG, Halsberghe C, et al. Electronic training devices: discussion on the pros and cons of their use in dogs as a basis for the position statement of the European Society of Veterinary Clinical Ethology. *J Vet Behav Clin Appl Res*. (2018) 25:71–75. doi: 10.1016/j.jveb.2018.02.006
8. APDT-UK. *Association of Pet Dog Trainers*. APDT. Available online at: https://apdt.co.uk/ (accessed March 30, 2020).
9. Stallard C. *A Ban on Using Electric Shock Collars Is Great – but We Need to Ban Sales of Aversive Devices to Protect Pets*. Blue Cross. Available online at: https://www.bluecross.org.uk/we-need-to-ban-aversive-devices-to-protect-pets (accessed November 25, 2019).
10. Lines JA, Van Driel K, Cooper JJ. The characteristics of electronic training collars for dogs. *Vet Rec*. (2013) 172:288. doi: 10.1136/vr.101144
11. Companion Animal Welfare Council, CAWC. *The Use of Electric Pulse Training Aids (EPTAs) in Companion Animals*. (2012). Available online at: http://eprints.lincoln.ac.uk/14640/1/CAWC%20ecollar%20report.pdf (accessed December 3, 2019).
12. Polsky RH. Electronic shock collars: are they worth the risks? *J Am Anim Hosp Assoc*. (1994) 30:463–8.
13. Polsky RH. Can aggression in dogs be elicited through the use of electronic pet containment systems? *J Appl Anim Welf Sci*. (2000) 3:345–57. doi: 10.1207/S15327604JAWS0304_6
14. Blackwell EJ, Casey RA. The use of shock collars and their impact on the welfare of dogs: A review of the current literature (2006). Available online at: https://positively.com/files/The-Use-of-Shock-Collars-and-Their-Impact-on-the-Welfare-of-Dogs.pdf (accessed November 10, 2019).
15. Ward M, MacKellar I. *Shock Collars – The Shocking Truth*. APBC (2010). Available online at: https://www.apbc.org.uk/pet-owner-article/shock-collars-the-shocking-truth/ (accessed June 18, 2019).
16. Dogs Trust. *We Call on Government to Ban Electric Shock Collars*. Dogs Trust (2018). Available online at: https://www.dogstrust.org.uk/news-events/news/electric-shock-collars-news (accessed November 25, 2019).
17. O'Brien C. *RSPCA Highlight Importance of Electric Shock Collar Ban – 100 Months on*. RSPCA (2018). Available online at: https://news.rspca.org.uk/2018/08/17/rspca-highlight-importance-of-electric-shock-collar-ban-100-months-on/ (accessed November 24, 2019).
18. Kennel Club. *UK Wide Electric Shock Collar Ban Will Save Hundreds of Thousands of Dogs From Aversive Training Devices*. Kennel Club (2018). Available online at: https://www.thekennelclub.org.uk/press-releases/2018/august/uk-wide-electric-shock-collar-ban-will-save-hundreds-of-thousands-of-dogs-from-aversive-training-devices/ (accessed March 29, 2020).
19. Electronic Collars Manufacturers Association, ECMA. *EMCA Products: Electronic Training Products Fall Into 3 Main Types*. EMCA (1996). Available online at: http://ecma.eu.com/product/ (accessed November 27, 2019).
20. Christiansen FO, Bakken M, Braastad BO. Behavioural differences between three breed groups of hunting dogs confronted with domestic sheep. *Appl Anim Behav Sci.* (2001) 72:115–29. doi: 10.1016/S0168-1591(00)00202-1
21. Christiansen FO, Bakken M, Braastad BO. Social facilitation of predatory, sheep-chasing behaviour in Norwegian Elkhounds, grey. *Appl Anim Behav Sci.* (2001) 72:105–14. doi: 10.1016/S0168-1591(00)00208-2
22. Salgirli Y, Schalke E, Boehm I, Hackbarth H. Comparison of learning effects and stress between 3 different training methods (electronic training collar, pinch collar and quitting signal) in Belgian Malinois Police Dogs. *Rev Med Vet*. (2012) 163:530–5.
23. Dale AR, Statham S, Podlesnik CA, Eliffe D. The acquisition and maintenance of dogs' aversion responses to kiwi (*Apteryx* spp.) training stimuli across time and locations. *Appl Anim Behav Sci.* (2013) 146:107–11. doi: 10.1016/j.applanim.2013.04.006
24. Robertson. *Consultation on Potential Controls or Prohibition of Electronic Training Aids in Scotland: Analysis of Consultation Responses*. The Scottish Government (2016). Available online at: http://www.gov.scot/Resource/0050/00503099.pdf (accessed June 16, 2019).
25. Blackwell EJ, Bolster C, Richards G, Loftus BA, Casey RA. The use of electronic collars for training domestic dogs: estimated prevalence, reasons and risk factors for use, and owner perceived success as compared to other training methods. *BMC Vet Res*. (2012) 8:93. doi: 10.1186/1746-6148-8-93
26. PDSA. *PDSA Animal Wellbeing (PAW) Report 2019*. PDSA (2019). Available online at: https://www.pdsa.org.uk/media/7420/2019-paw-report_downloadable.pdf (accessed November 27, 2019).
27. *Animal Welfare (Electronic Collars) (Wales) Regulations 2010 (No. 943 W. 97)*. Welsh Assembly (2010). Available online at: https://www.legislation.gov.uk/wsi/2010/943/contents/made (accessed June 10, 2019).
28. Scottish Government. Scottish Government Policy on Electronic Training Collars. Scottish Government (2018). Available online at: http://www.gov.scot/Topics/farmingrural/Agriculture/animal-welfare/AnimalWelfare/companion/electronictrainingaids (accessed June 10, 2019).
29. DEFRA. A ban on electronic training collars for cats and dogs in England. Department for Environment, Food & Rural Affairs (2018). Available online at: https://consult.defra.gov.uk/animal-health-and-welfare/ban-on-electronic-training-collars-cats-and-dogs/supporting__documents/ecollarsconsultdocument.pdf (accessed June 10, 2019).
30. Ripley K. *Everything You Need to Know About Training With E-Collars*. Wide Open Pets (2019). Available online at: https://www.wideopenpets.com/everything-need-know-training-E-collars/ (accessed March 29, 2020).
31. Cooper JJ, Cracknell N, Hardiman J, Wright H, Mills D. The welfare consequences and efficacy of training pet dogs with remote electronic training collars in comparison to reward based training. *PLoS ONE*. (2014) 9:e102722. doi: 10.1371/journal.pone.0102722
32. DEFRA. *Studies to Assess the Effect of Pet Training Aids, Specifically Remote Static Pulse Systems, on the Welfare of Domestic Dogs - AW1402*. Final report prepared by Prof. Jonathan Cooper, Dr. Hannah Wright, Prof. Daniel Mills (University of Lincoln); Dr. Rachel Casey, Dr. Emily Blackwell (University of Bristol); Katja van Driel (Food and Environment Research Agency); Dr. Jeff Lines (Silsoe Livestock System) (2013). Available online at: http://sciencesearch.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&Completed=0&ProjectID=15332 (accessed June 16, 2019).
33. DEFRA. *Studies to Assess the Effect of Pet Training Aids, Specifically Remote Static Pulse Systems, on the Welfare of Domestic Dogs; Field Study of Dogs in Training - AW1402A*. Final report prepared by Prof. Jonathan Cooper, Dr. Nina Cracknell, Jessica Hardiman and Prof. Daniel Mills (University of Lincoln) (2013). Available online at: http://sciencesearch.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&Completed=0&ProjectID=17568#discription (accessed June 16, 2019).
34. China L. *Comparison of E-collar and Reward Based Training on Acute Behavioural Responses and Efficacy of Training in Pet Dogs*. BSc. University of Lincoln (2018).
35. Masson S, Nigron I, Gaultier E. Questionnaire survey on the use of different e-collar types in France in everyday life with a view to providing recommendations for possible future regulations. *J Vet Behav Clin Appl Res*. (2018) 26:48–60. doi: 10.1016/j.jveb.2018.05.004
36. Skyrme R, Mills DS. An investigation into potential overshadowing of verbal commands during training. *J Vet Behav Clin Appl Res*. (2010) 1:42. doi: 10.1016/j.jveb.2009.09.039
37. Beerda B, Schilder MBH, van Hoff Jan ARAM, de Vries HW, Mola JA. Behavioural, saliva cortisol and heart rate responses to different types of stimuli in dogs. *Appl Anim Behav Sci*. (1998) 58:365–81. doi: 10.1016/S0168-1591(97)00145-7

38. Schilder MB, van der Borg JA. Training dogs with help of the shock collar: short and long term behavioural effects. *Appl Anim Behav Sci*. (2004) 85:319–34. doi: 10.1016/j.applanim.2003.10.004
39. Schalke E, Stichnoth J, Ott S, Jones-Baade R. Clinical signs caused by the use of electric training collars on dogs in everyday life situations. *Appl Anim Behav Sci*. (2007) 105:369–80. doi: 10.1016/j.applanim.2006.11.002
40. Mills DS. Training and learning protocols. In: Horwitz DF, Mills DS, editors. *BSAVA Manual of Canine and Feline Behavioural Medicine*. Cheltenham Glos: BSAVA (2009). p. 49–64.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

METHODS
published: 13 August 2020
doi: 10.3389/fvets.2020.00470

The Penn Vet Working Dog Center Fit to Work Program: A Formalized Method for Assessing and Developing Foundational Canine Physical Fitness

Brian D. Farr[1,2]*, Meghan T. Ramos[2]* and Cynthia M. Otto[2]

[1] Army Medical Department Student Detachment, 187th Medical Battalion, Medical Professional Training Brigade, Joint Base San Antonio - Fort Sam Houston, San Antonio, TX, United States, [2] Penn Vet Working Dog Center, Clinical Sciences and Advanced Medicine, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, United States

OPEN ACCESS

Edited by:
Erik Hofmeister,
Auburn University, United States

Reviewed by:
Lowell Paul Waggoner,
Auburn University, United States
John J. Ensminger,
Private Practitioner, Rhinebeck, NY, United States

***Correspondence:**
Brian D. Farr
brian.farr.dvm@gmail.com
Meghan T. Ramos
megramos@upenn.edu

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences, a section of the journal Frontiers in Veterinary Science

Received: 08 May 2020
Accepted: 25 June 2020
Published: 13 August 2020

Citation:
Farr BD, Ramos MT and Otto CM (2020) The Penn Vet Working Dog Center Fit to Work Program: A Formalized Method for Assessing and Developing Foundational Canine Physical Fitness. Front. Vet. Sci. 7:470. doi: 10.3389/fvets.2020.00470

Fit to Work is a formalized working dog foundational physical fitness assessment and development program. The Penn Vet Working Dog Center developed this program to address the needs of working dog handlers, trainers, and programs for simple, effective, and efficient methods to develop and assess working dog physical fitness. Fit to Work focuses on the foundational fitness modalities of strength, stability, mobility, and proprioception. The Penn Vet Working Dog Center piloted and refined this program over 3 months in a closed population of 31 working dogs in training. Fit to Work consists of posture development and maintenance, warm-up and cool-down routines, training exercises, and assessment methods. To simplify implementation for dogs and personnel, the foundational training program incorporates a discrete number of exercises, standardized progression steps, defined criteria for progression, and a reduced emphasis on learned behaviors. Fit to Work also enables safe and progressive assessment of foundational fitness through a tiered and inexpensive process. Future research will focus on validation of training and assessment methods, development of assessment standards, and correlation of physical fitness with operational performance.

Keywords: canine physical fitness, fitness assessment, exercise program, musculoskeletal system, working dogs, performance

INTRODUCTION

Canine Physical Fitness

Canine physical fitness is a recognized requirement for safe and effective performance of working dogs. Current definitions of canine physical fitness are vague and are frequently tied to a specific type of activity (e.g., running, agility). Extrapolating from human literature, physical fitness is defined by The American College for Sports Medicine as "a set of attributes that people have or achieve" (1). The American College for Sports Medicine further defines physical fitness into measurable health and skill-related (athletic ability) attributes (1, 2). The health-related components of physical fitness are: cardiorespiratory endurance, muscular endurance, muscular strength, body composition (e.g., amount of fat vs. muscle), and mobility (1, 2).

The skill-related components of physical fitness are agility, balance, coordination, speed, power, and reaction time (1). Defining canine physical fitness in these terms of measurable components enables further investigation into the details of each component and their impact on the working dog's physical requirements to perform a variety of athletic tasks. While there are other aspects of fitness important for working dogs (e.g., mental), this article will use the term "fitness" to solely refer to physical fitness.

Foundational Fitness

Basic canine fitness requires a foundational fitness program coupled with a scientifically based and repeatable assessment process. Foundational fitness addresses mobility, stability (e.g., core, forelimb, hindlimb), strength, and proprioception required to perform basic physical tasks such as running, jumping (up and down), navigating unstable surfaces, and quickly or abruptly changing direction. The definition for each foundational fitness component can be found in **Table 1**. These components were selected to primarily target the muscles of the core, hind limbs, and supporting soft tissue structures (e.g., fascia, ligaments, etc.) that are not primarily engaged when a dog performs routine physical activities. Dogs naturally bear approximately two thirds of their weight on their forelimbs; thus, representing the brunt of musculoskeletal development and engagement during normal physical activity (3, 4). The uneven distribution of weight along with common hip and spinal anomalies may impact muscle development and maintenance of the hindlimbs and core, predisposing these areas to injury or degenerative changes during moderate and vigorous activities encountered in the majority of working dog careers (5–7). Development of a dog's musculoskeletal system in the optimal biomechanical alignment through whole body proprioception, muscle stability and strength of the core, forelimbs, and hindlimbs, and mobility may decrease a dog's susceptibility to, the severity of, or the recovery time for an injury.

This canine foundational fitness program is designed to be incorporated into an established working dog training program (e.g., search and rescue, law enforcement, military, etc.). The dog's career training develops the baseline cardiovascular endurance and skill related components of career-specific fitness. However, additional career-specific fitness training should be pursued to optimize career performance. A canine foundational fitness program for working dogs is designed to incorporate exercises that require minimal additional training of the dog and handler, include clear incremental exercise progressions, utilize easily accessible and low-cost equipment, and should also be applicable to a wide variety of working dog careers.

Foundational Fitness Assessment

One of the biggest challenges in evaluating the current level of canine fitness and the impact of any intervention, such as a fitness program, is the establishment of a formalized, repeatable method for assessment. To date, there are no peer-reviewed, published studies of systematic foundational programs with assessments specifically for working dogs. The methods for creating a foundational fitness assessment should produce scientifically valid, reliable, and reproducible canine fitness data (8–10). Clear and objective outcome variables, such as duration of exercise, number of repetitions, or distance traveled are critical. The assessment should be designed to account for intra-dog and inter-dog variability and learning effect of the dog, but lead to results that allow the dog to be compared to the population (9). The individual performing the assessment should not require extensive experience, thereby maximizing the intra-rater or inter-rater reliability. Minimizing human interaction with the dog during the assessment is important to prevent skewing the data. Finally, skills being assessed should directly translate to the realistic and operational athletic requirements of working dogs (8, 11).

The Fit to Work Program

The Fit to Work (FTW) foundational exercises and associated assessment method presented here are designed to create a formalized, highly reproducible, and inexpensive method to create the first stage of a balanced fitness program for working dogs. The goals of the program are to meet the

TABLE 1 | Definitions of foundational canine fitness components.

Foundational fitness definitions	
Term	**Definition**
Foundational canine fitness	The mobility, muscle stability (core, forelimb, hindlimb), muscle strength, and proprioception required to perform basic physical activities such as running, jumping (up and down), navigating unstable surfaces, and quickly or abruptly changing direction that a dog needs to perform its job-related tasks.
Mobility	The ability to move the body without limitation from joints, muscles, tendons, or other connective tissues.
Strength	The ability of a muscle group to exert maximal force.
Stability	The ability for the muscles and tissues supporting a joint to resist unwanted or abnormal movement.
Forelimb stability	The ability to maintain the shoulder, elbow, and carpus in biomechanically optimal positions.
Hindlimb stability	The ability to maintain the hip, stifle, and tarsus in biomechanically optimal positions.
Core stability	The ability to maintain the spine (cervical, thoracic, lumbosacral) in biomechanically optimal positions.
Proprioception	The perception or awareness of the position and movement of the body (12).

foundational fitness requirements of working dogs that are not currently addressed in the traditional training disciplines. The program was designed with the realistic expectations and time constraints of the working dog training and utilization worlds. To address the gap in knowledge and lack of additional training time, the FTW program contains a discrete number of exercises, standardized progression steps, and defined criteria for progression. The FTW program significantly reduces the requirement for learned behaviors to shorten the learning curve and enable a novice working dog handler to develop their dog's physical fitness. We will present each of the components of the program, their description, the purpose of the exercise, the known or anticipated role in performance and injury prevention, a recommended progression of difficulty, a brief summary of contraindications, and suggested training approaches.

METHODS

Pilot Implementation

All dogs included in the implementation below are owned by the Penn Vet Working Dog Center, School of Veterinary Medicine, University of Pennsylvania. The Fit to Work (FTW) program is included in the Penn Vet Working Dog Center (PVWDC) Puppy Foundation Program protocol 804547 approved by The University of Pennsylvania Institutional Care and Use Committee.

The FTW program was piloted on a closed population of 31 dogs in training for careers in search and rescue, law enforcement, single-purpose detection, and medical detection. Each dog was fostered on nights and weekends by an individual or family who did not perform any FTW activities. The dogs ranged in age from 2 months to 6 years of age and were of the Labrador Retriever ($n = 15$), German Shepherd Dog ($n = 8$), Belgian Malinois ($n = 3$), Dutch Shepherd ($n = 3$), Small Munsterlander ($n = 1$), and Doberman Pinscher ($n = 1$) breeds. The dogs were assigned to 5 full-time trainers who each had responsibility for 3–8 dogs. The training team was assisted by ~20 part-time undergraduate interns and ~30 adult volunteers.

Foundational Fitness Training Structure

The foundational fitness training consists of a series of daily and three times weekly exercises. Daily exercises develop posture via the Posture Down and Posture Sit and enhance mobility via the Warm-Up and Cool-down exercises prior to and immediately after any moderate or vigorous activities. Three times a week, the dog participates in the foundational fitness exercises which focus on strength, stability, balance, and proprioception. These exercises are divided into two circuits consisting of three exercises per circuit. Circuit One consists of the Posture Down or Chipmunk, the Squat, and the Back-up. Circuit Two consists of the Plank, the Pivot, and the Back-up. Each circuit is performed twice either consecutively or in an alternating fashion. Each exercise within the circuit is performed for 30–60 s. The entire training session takes between 15 and 20 min excluding equipment setup and break down time.

TABLE 2 | Summary of the foundational fitness exercises and their primary and secondary targets and the preferred method of rewarding during the exercise.

Foundational fitness exercise summary												
	Exercise name	**Primary and secondary targets**								**Reward method**		
		Mobility	**Core stability**	**Core strength**	**Hindlimb stability**	**Hindlimb extension**	**Forelimb stability**	**Forelimb extension**	**Proprioception**	**T**	**IF**	**CF**
Daily exercise	Figure-8	P	S		S		S		S	X	X	
	Paws-up	P	S		S				S	X	X	
	Four-position cookie stretch	P							S	X		X
	Posture sit	S	P	S	S				S	X		X
Weekly circuit	Posture down	S	P	S	S				S			X
	Chipmunk		P	P	S		S		P			X
	Plank		P	S	S		S	S	S			X
	Pivot		P	S*	P		S		S	X	X	
	Squat		S	S*	S	P			S	X	X	
	Back-up		S		S		S	P*		X	X	

*Primary and secondary targets key: P, primary target; S, secondary target; *at the higher progression levels. Reward method key: T, Toy; IF, intermittent food reward; CF, continuous food reward.*

FIGURE 1 | (A) Shows the Paws-up. **(B–D)** Show the Figure-8. **(E–H)** Show the Four-Position Cookie Stretch [**(E)** Nose to hip, **(F)** Nose to shoulder, **(G)** Nose to rear foot, and **(H)** Nose to chest].

A summary of the foundational fitness exercises, their fitness modality targets, and recommended reward methods can be found in **Table 2**.

Foundational Fitness Exercises

Warm-Up and Cool-Down

Description

To complete the Warm-up, the dog walks for 30 s, trots for 30 s, performs a Paws-up for 15 s on the handler's arm or an object (**Figure 1A**), and performs 3 Figure-8s (**Figures 1B–D**) between the handler's legs or around two objects or people set 45–90 cm (18–36 in) apart. The Warm-up should take ∼90 s to complete. To complete the Cool-down, the dog walks for 30 s while the handler observes for any physical or behavioral abnormalities, performs a Paws-up for 15 s, and performs a Four-Position Cookie Stretch (**Figures 1E–H**) on each side. Finally, the handler checks the dog's paws, pads, and nails for signs of injury. The Cool-down should take ∼120 s to complete.

Purpose

The Warm-up should be completed before all activities to prepare the dog's body for the upcoming movements. The walk and trot portion of the Warm-up progressively increase body temperature and tissue mobility (12). The Paws-up targets the hip and abdominal tissues (predominantly the iliopsoas, psoas major, and rectus abdominis muscles) for extension, and the Figure-8 prepares the neck and trunk tissues (predominantly the lateral muscles of the neck, extrinsic muscles of the forelimb, internal and external abdominal oblique muscles, and epaxial spinal muscles) for lateral movement.

The Cool-down should be performed after all activity to maintain and increase mobility while tissues are warm and to identify any injuries sustained during training. The walk portion of the Cool-down allows the dog's heart rate and breathing to begin to decrease, and gives the handler the opportunity to identify any lameness. The Paws-up targets the mobility of hip and abdominal tissues, and the Four-Position Cookie Stretch targets the mobility of the neck and trunk. Finally, checking the dog's paws, pads, and nails after training allows rapid identification of issues in these injury-prone areas.

Role in performance and injury prevention

The walk and trot portions of the Warm-up increase blood flow and oxygen delivery to muscles and connective tissues and increase body temperature to prepare for higher-intensity activity (13, 14). The Paws-up and Figure-8 may optimize movement and decrease injury by allowing the dog access to an increased tissue range of motion (15–17). Both movements may also increase neuromuscular activation and thus athletic performance (18, 19). The entire Cool-down gives the dog's body time to recover from activity before returning to rest and the handler an opportunity to identify injuries that were not evident when the dog was engaged in activity. While post-exercise stretching appears to have limited effects on muscle soreness (20), the increased tissue mobility from the Paws-up and Four-Position Cookie Stretch may decrease future tissue injury risk (21, 22) and increase future performance (23). Checking the dog's paws, pads, and nails allows early identification and rapid treatment of performance-limiting injuries.

Contraindications

Without guidance from a veterinarian, active exercise and thus the Warm-up and Cool-down are not recommended for dogs with suspected musculoskeletal abnormalities or cardiorespiratory disease. With supervised rehabilitation, the use of and adaptations to the Warm-up and Cool-down can be customized to the dog's injury or condition.

Progression

The range and thus the intensity of the Paws-up may be increased by adjusting the height of the handler's arm or the object used.

The intensity of the Figure-8 may be increased by decreasing the width between the handler's legs (or between the objects that the dog is navigating) or by increasing the speed of execution. The range and thus the intensity of the Four-Position Cookie Stretch may be increased by extending the dog's nose closer to each targeted position or by increasing the duration at each position.

Training

While many ways exist to perform or train the Warm-up and Cool-down exercises, we have found the following to work well for our population of working dogs. To train the Paws-up on an arm, have the dog sit, kneel next to them, scoop your forearm under their forelimbs, and use a toy or food reward to lure them into a stand as you slowly rise. Use the toy or food reward to maintain them in the Paws-up position for the desired duration. Common technique errors include rewarding too high above the dog's head which overextends the dog's neck and allowing the dog to rest a significant portion of their weight on your forearm.

To train the Figure-8, start with your feet approximately twice your shoulder width distance apart. Use a "Touch" command or lure the dog between your legs, then with alternating hands direct the dog in one direction (e.g., clockwise around your right leg), then back through your legs and around the other leg in the opposite direction (e.g., counterclockwise around your left leg). Common technique errors include inadequate rate of reinforcement, giving improperly timed commands and imprecise luring, all of which prevent the dog from sharply turning around your legs. Good form also requires good handler form, standing up straight is the goal.

To train the Four-Position Cookie Stretch, lure the dog in a standing position so that its spine is aligned perpendicular to your spine. Gently place your opposite hand under their thorax or abdomen to maintain their position against you. Lure their nose away from your body, in a plane parallel to their spine to their hip, mark, and reward. Then lure their nose to their hip and then gradually move it as close to their shoulder as possible, while maintaining it in the plane parallel to their spine, mark, and reward. Lure their nose to their hip and then move the lure distally until their nose is as close to their rear foot as possible, mark, and reward. Finally, lure their nose between their forelimbs and as close to their chest as possible, mark, and reward. A common technique error is not maintaining the dog's position against your body which allows them to decrease the intensity of the exercise.

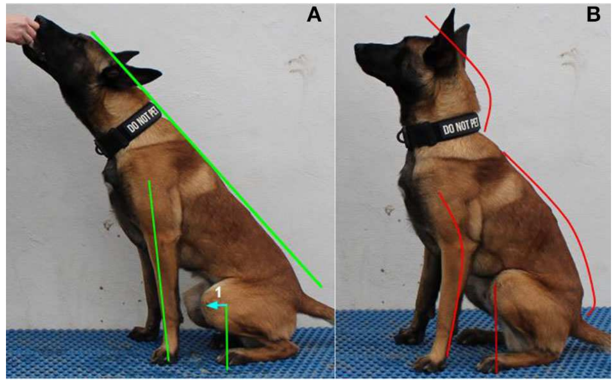


FIGURE 2 | (A) Shows the Posture Sit in the correct position. Note the straight line from the head to the base of the tail. The forelimb is extended in a straight line. The stifle is dorsal to or just cranial to the digits (line 1 shows the range of the stifle over the digits). **(B)** Shows the Posture Sit in the incorrect position. Note the rounded forelimb and spine (cervical, thoracic, and lumbosacral). The stifle is positioned caudal to the digits.

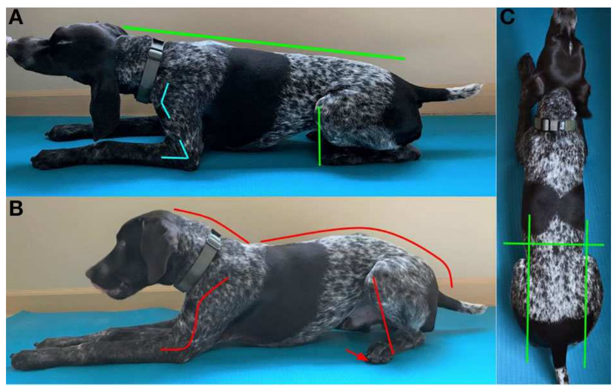


FIGURE 3 | (A) Shows the Posture Down in the correct position. Note the straight line from the head to the base of the tail. The forelimb is flexed at the elbow and shoulder. The stifle is dorsal to or just cranial to the digits. **(B)** Shows the Posture Down in the incorrect position. Note the extended elbow and rounded spine (cervical, thoracic, and lumbosacral). The hindlimbs are abducted and the stifle is caudal to the digits. The red arrow indicates the abduction of the hind foot. **(C)** Shows the correct position of the Posture Down from above (the dorsal view) the dog. All four ipsilateral limbs are aligned in a sagittal plane, and the hindlimb digits are obscured by the stifles.

Posture Sit and Posture Down

Description

To perform the Posture Sit, the dog should sit with its coxofemoral joint, stifle, tarsus, and hindlimb digits in the same straight sagittal plane (**Figure 2**). The forelimbs should be half a stride length in front of the hindlimb digits. The forelimb paws should be directly under the shoulder and should remain on the ground throughout the exercise. The shoulder (glenohumeral joint), elbow, carpus, and forelimb digits should be aligned in the same straight sagittal plane as the hindlimbs. To achieve the Posture Sit position, the dog extends its spine and rolls its pelvis forward to form a straight line from the nose to the base of the tail (**Figure 2A**). During the movement, the dog's stifles should move dorsal to or just cranial to the hindlimb digits.

To perform the Posture Down, the dog should lay down so that the ipsilateral right and left limbs are within the same sagittal plane, often referred to as a "sphinx" position (**Figure 3**). To achieve the Posture Down position, the dog flexes its shoulder and elbow 10–20° depending on size of the dog, rolls its pelvis forward, and extends its spine to form a straight line from the nose to the base of the tail (**Figure 3A**). During the movement, the dog's stifles should move dorsal to or just cranial to the hindlimb digits.

Purpose

Performing the Posture Sit and Posture Down primarily targets stabilizer muscles of the spine (sagittal extension) and abdominal engagement (sagittal flexion). The primary muscle groups of the Posture Sit and Posture Down are the cervical spine extensor muscle (splenius), trunk extensor muscles—epaxials (transversospinalis, longissimus, iliocostalis), latissimus dorsi, hip stabilizers (gluteal, psoas major, iliacus, adductor, and piriformis muscles), and abdominal muscle (rectus abdominis). Proprioception of spine, shoulder, elbow, lumbosacral, coxofemoral joint, and stifle are secondarily developed during both the Posture Sit and Posture Down.

Role in performance

Performing the Posture Sit and Posture Down strengthens the dog's core stability musculature and supports optimal biomechanical alignment by creating a proprioceptive memory of correct posture (5, 24, 25). Promoting correct posture in a sit and down position establishes the foundation necessary to perform activities safely (25). All dogs benefit from correct posture whether in athletic performance, tactical operations, or as a household pet.

Role in injury prevention

Engaging in Posture Sits and Posture Downs may prevent repetitive stress on the spine resulting in kyphosis and susceptibility to traumatic or alignment-related injuries (22). Additionally, observation of a dog's ability to achieve the proper Posture Sit and Posture Down can act as a screen for subtle injuries. A dog that is reluctant to extend its spine in the Posture Sit or Posture Down may be experiencing lower back pain. A dog that is reluctant to sit or down in a straight sagittal plane and abducts a limb may be experiencing hip or stifle pain. Recognition of these subtle changes can lead to earlier diagnosis of an injury and prevention of further injury.

Progression

Teaching correct posture development in both the sit and down positions when the dog is a puppy (8 weeks) and maintained throughout the dog's life is highly encouraged. For healthy dogs, progression in the Posture Sit and Posture Down is achieved by increasing the duration of holding the correct position and then destabilizing the surface on which the dog is performing the exercise. For beginner dogs, the duration may be as short as 1–2 s. The dog is expected to maintain a 30 s hold in the correct position before progressing to an unstable surface. An unstable surface should be a flat platform with the destabilization component located underneath the platform. Performing this exercise on a flat surface allows the trainer to evaluate if the dog is not remaining in a proper position and/or is consistently favoring one side over another. If the exercise is performed on two unstable surfaces such as balance discs alone, the dog can make subtle changes in body posture that may result in asymmetric muscle development and lack of engagement of the smaller secondary musculature.

Contraindications

The Posture Sit and Posture Down are safe for healthy dogs of all ages. Caution in progression of the Posture Sit and Posture Down should be taken with suspected spine, hip, or stifle abnormalities. While critical to the rehabilitation process of many musculoskeletal injuries, the Posture Sit and Posture Down should only be performed for these patients under the guidance of the dog's veterinarian.

Training

To train the Posture Sit and Posture Down, have the dog perform a correct (square) sit or (sphinx) down position as described above. Placing the dog on a narrow platform or between two objects will encourage the dog to align all limbs in the sagittal plane. Once in the proper square sit or sphinx down, a treat or toy lure at a dog's resting nose level height should be utilized to encourage the dog to move forward slightly [2–7 cm (1–3 in) depending on the dog's size]. The dog should be lured until the spine is straight and the stifles are dorsal to or just cranial to the hindlimb digits. If the treat or toy lure placement is too high, the dog will attempt to stand. If the treat or toy lure placement is too low in the Posture Down, the dog will lift its hocks or attempt to crawl. If the treat or toy lure placement is too low in the Posture Sit, the dog will attempt to lay down or will round its back. Time the delivery of the reward to the dog based on the dog's progression level and deliver the reward while the dog is holding the correct posture position.

Pivot

Description

To perform the Pivot, the dog places its forepaws on an elevated stable object (e.g., standard concrete block) and steps laterally (sidesteps) with its hindpaws around the object both clockwise and counterclockwise (**Figures 4A–H**). The Pivot is performed continuously in one direction for a specific duration or number of rotations and then repeated in the opposite direction after a rest interval.

Purpose

The Pivot primarily develops the dog's hip, stifle, and tarsus stability. Core stability (predominantly sagittal and lateral flexion), hindlimb proprioception, and hip, stifle, and tarsus extension are secondarily developed. Hip stability is provided by the hip abductors (superficial, middle, and deep gluteal muscles) and adductors (adductor, gracilis, semimembranosus, sartorius, and pectineus muscles). Stifle and tarsus stability may also benefit from the improved control over the lower extremity provided by the stable hip during movement (26).

Role in performance

The Pivot strengthens the dog's hip stability musculature and supports optimal biomechanical alignment of the hip, stifle, and tarsus when the dog is moving and stationary (26). A stable hip may increase power generation when jumping or sprinting (27), provide effective lateral force and stable footing when turning (28), and align the hips for subsequent movements when landing from a jump (29). A dog that is required to perform movements

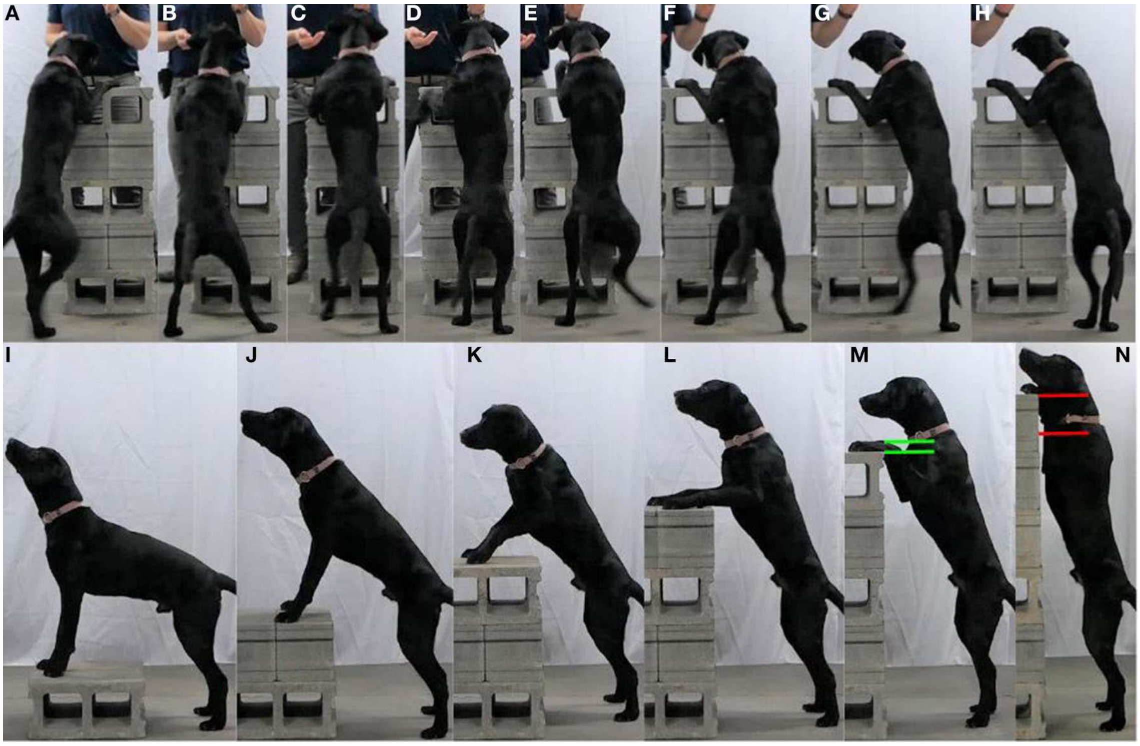


FIGURE 4 | (A–H) Illustrate the Pivot. **(I–N)** Show progression of the Pivot [**(I)** Level 1, **(J)** Level 2, **(K)** Level 3, **(L)** Level 4, **(M)** Level 5 (maximum for this dog), and **(N)** Level 6 (too high for this dog)].

with the forelimbs elevated, weight shifted to the hindlimbs, and the hip extended (e.g., searching elevated surfaces or vehicles) may also benefit from the Pivot.

Role in injury prevention

Stable hips developed by the Pivot may prevent traumatic (30) or alignment-related injuries and assist in limiting the progression of orthopedic disease (31). A dog that missteps or begins to slip into a splayed hindlimb position may incur an iliopsoas muscle, hip adductor group, or hip joint injury. If this dog has more stable hips from training the Pivot, it may have a better chance of recovering its footing and preventing injury (30). Increased hip stability helps maintain optimal motion of the hip and stifle joints and may reduce injuries caused by poor joint alignment (32, 33). Although there are no clinical trials to evaluate the impact, dogs with early dysplastic hips may benefit from increased muscular support and reduced joint instability, and dogs with stable cruciate ligament disease may benefit from the improved lower extremity control.

Progression

For mature dogs, the Pivot is usually trained on a 20 cm (8 in) high and 40 cm (16 in) square object (e.g., two standard concrete blocks) (**Figure 4I**), although puppies under 3–4 months of age or small breed dogs may benefit from an object that is only 20 cm (8 in) square (e.g., two standard bricks). Once the dog is able to complete three revolutions in 30 s in each direction, the object height is increased by 20 cm (8 in) (**Figures 4J–M**). Elevation of the object height progressively shifts a greater percentage of the dog's weight onto their hindlimbs. This process is continued until the next height progression would result in the dog's forepaws being elevated higher than the shoulder joint (**Figure 4N**). Further progression is primarily provided by external weight in the form of a weight vest with weight increments scaled to the weight of the dog (e.g., for a 20 kg (44 lb) dog start with 2.3 kg (5 lb) of external weight). Alternate methods of progression include destabilizing the hindpaw surface or increasing hindlimb lateral resistance (water or exercise band).

Contraindications

Without guidance from a veterinarian, the Pivot is not recommended for dogs with suspected spine, hip, stifle, tarsus, or hindpaw abnormalities. With supervised rehabilitation, the Pivot can be adapted to the dog's injury or condition.

Training

While there are many ways to perform or train the Pivot, the following works well for our population of working dogs. First, have the dog place their forepaws on the elevated object, stand next to the dog's flank, provide the verbal cue to "Step," provide gentle body pressure by moving into the dog, and then mark and reward any lateral movement of the hind foot while the forelimbs remain on the object. Hold the reward in the right hand when performing the Pivot clockwise and in the left hand when

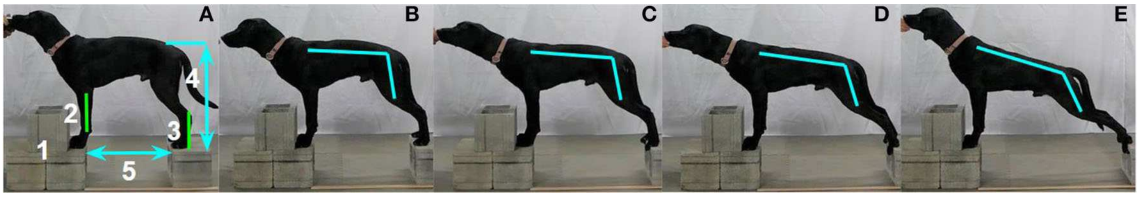


FIGURE 5 | (A) Shows the Plank setup (1 shows the recommended three-block front object), starting position [2 illustrates a vertical distal forelimb and 3 illustrates vertical metatarsals (rear pastern)] and measurements (4 is the hip height and 5 is the distance between the objects). **(B–E)** Show progression of the Plank [**(B)** Level 2, **(C)** Level 4, **(D)** Level 6, and **(E)** Level 8].

performing the Pivot counterclockwise. Work toward having the dog moving at least a quarter rotation ahead of you and rewarding after each quarter rotation. A well-trained dog should perform the Pivot in a position 180° from the handler and be able to complete a 30 s interval prior to reward. Common technique errors include relying solely on body pressure after an initial learning period, using the hand closest to the dog to reward, rewarding too far above or away from the dog's mouth or too infrequently, or allowing the dog to move too quickly and skip with or cross their hind feet. A soft high-value reward (e.g., small piece of cheese) assists our population of working dogs to rapidly swallow the reward and resume movement.

Plank

Description

To perform the Plank, the dog stands with its forepaws on one stable object (e.g., standard concrete block) and its hindpaws on a second stable object of equal height. Place (or cue the dog to place) their forepaws on the rear edge of the front object and their hindpaws on the front edge of the rear object. Adjust the distance between the objects until both the dog's distal forelimbs (elbow to carpus) and metatarsals (rear pastern) are vertical (**Figure 5A**, Lines 2 and 3). Measure and record the distance between the objects (**Figure 5A**, Line 5). Next, measure the dog's height from the top of the object to the dorsum of the dog directly above the coxofemoral joint (**Figure 5A**, Line 4) and use this to determine the dog's height-adjusted movement increment (**Table 3**). Then, move the two objects apart to the desired level, and the dog holds the position for the specified duration or until they step down. A proper Plank is performed when the distal forelimbs are maintained in a vertical position and the hindpaws are in contact with the rear object (**Figures 5B–E**).

Purpose

The Plank primarily develops a dog's core stability (predominantly resistance to thoracic and lumbar spine extension). Isometric elbow and carpus extension are secondarily developed. Core stability in the Plank is predominantly provided by the rectus abdominis, external abdominal oblique, iliopsoas, and psoas minor muscles. In humans, the spinal extension muscles have also been shown to play a role (34, 35).

Role in performance

The Plank strengthens the muscles that provide stability to a dog's spine. This increased stability may enhance the dog's ability to generate whole-body power (36–38), run (39–42), and perform single-leg movements (43). While improved core stability has differing effects on agility in humans (44, 45), the differences in anatomy may enhance the role of core stability for canine agility. A dog performing movements with the forepaws and hindpaws on separate surfaces, the forepaws on elevated surfaces, or any paw on unstable surfaces may benefit from training the Plank (34, 46).

Role in injury prevention

Training movements that develop a dog's spinal stability may increase the likelihood of maintaining optimal biomechanical alignment when gravity or the motion of the dog's body cause thoracic or lumbar spine hyperextension. A dog that repeatedly hurdles obstacles (e.g., agility or law enforcement), spends prolonged time in a forepaw-elevated position (e.g., searching vehicles or elevated surfaces), traverses unstable surfaces that predispose it to a fall, (e.g., disaster search and rescue) encounters powerful force to the spine (e.g., criminal apprehension), or is at risk of spine injury or intervertebral disc disease may benefit from training the Plank.

Progression

The Plank is usually trained in intervals of up to 30 s at a particular level. The dog is given 45 s in which to accumulate 30 s of proper Plank. The 30 s duration time is paused when the dog moves out of the proper position (by shifting backwards so the distal forelimbs are no longer vertical or moving a hindpaw off the rear object) and the timer is resumed when the proper position is achieved again. The interval is stopped when the dog accumulates 30 s, the 45 s time elapses, or the dog steps off either object.

Once a dog is able to accumulate 30 s of proper Plank within the 45 s window, the objects can be moved to the next level (**Figures 5B–E**). This process is continued until the dog completes 30 s at Plank—Level 9. Further progression is primarily provided by external weight in the form of a weight vest with weight increments scaled to the weight of the dog [e.g., for a 20 kg (44 lb) dog start with 1.1 kg (2.5 lb) of external weight]. Alternate methods of progression include destabilizing either the forepaw surface, the hindpaw surface, or both.

Contraindications

Without guidance from a veterinarian, the Plank is not recommended for dogs with suspected spine or hip

TABLE 3 | Plank hip height measurements, corresponding Plank level increment, total distance for each hip height and Plank level, and approximate hip angle achieved.

Plank hip height, increment, and level All measurements are in centimeters										
Hip Height	**Increment**	**Level 1**	**Level 2**	**Level 3**	**Level 4**	**Level 5**	**Level 6**	**Level 7**	**Level 8**	**Level 9**
Approx. hip angle		93°	96°	99°	103°	107°	112°	122°	135°	150°
<20	1	1	2	3	4	5	6	7	8	9
20–26	1.5	1.5	3	4.5	6	7.5	9	10.5	12	13.5
27–33	2	2	4	6	8	10	12	14	16	18
34–40	2.5	2.5	5	7.5	10	12.5	15	17.5	20	22.5
41–47	3	3	6	9	12	15	18	21	24	27
48–54	3.5	3.5	7	10.5	14	17.5	21	24.5	28	31.5
55–61	4	4	8	12	16	20	24	28	32	36
62–69	4.5	4.5	9	13.5	18	22.5	27	31.5	36	40.5
>70	5	5	10	15	20	25	30	35	40	45

abnormalities. With supervised rehabilitation, the Plank can be adapted to the dog's injury or condition.

Training

While there are many ways to perform or train the Plank, the following works well for our population of working dogs. Position the stable objects (e.g., standard concrete blocks) at the starting position (Level 0), and have the dog step onto the rear object and walk across to the front object. For some dogs, placing an additional object between the front and back objects will facilitate the dog walking across the gap. An extra object may be used on top of the front object to assist the dog with maintaining its forepaws on the inside edge (**Figure 5A**). Gently slide the rear object to the desired level (an assistant may be useful for this step). Use a "Touch" command or a food reward to obtain and maintain the proper Plank position. Common technique errors include failure to adjust the dog's feet to the inside edges of the objects and improper positioning of the reward so the dog's neck is improperly aligned or the forelimbs are not vertical. A continuous high-value reward (e.g., frozen peanut butter in a cup) assists our population of working dogs in maintaining the proper position.

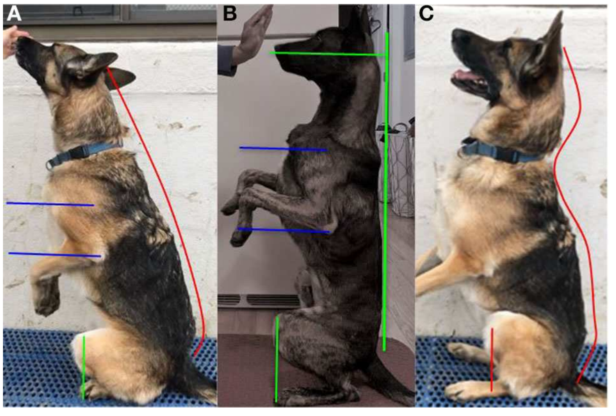


FIGURE 6 | (A) Shows the beginning progression of the Chipmunk—note the stifle dorsal to the digits. This position is held for <2 s. **(B)** Shows the correct Chipmunk position. The spine is straight, the stifles are dorsal to the digits, the carpi are suspended between the shoulders and elbows, and the head is in a neutral position. **(C)** Shows an incorrect Chipmunk. The spine is rounded in all three segments, the stifle is caudal to the digits, the dog's entire weight is resting on the ischium of the pelvis that is touching the ground.

Chipmunk

Description

The Chipmunk is also known as the "beg," "sit pretty," "sit up," or "sit erect." To perform the final correct position, the dog must have a square sit as described in the Posture Sit exercise. Once in a sit, the dog's forelimbs are suspended off the ground with the carpi suspended between elbow and shoulder height, the dog maintains a straight spine, and holds their head in a neutral position facing forward (**Figure 6B**).

Purpose

The Chipmunk primarily develops the dog's core stability (resistance to spinal extension, sagittal flexion, and transverse flexion). Secondarily, isometric shoulder and elbow contraction and minor concentric contraction of quadriceps muscles are involved. The primary stability muscles engaged during the Chipmunk are the rectus abdominis, external abdominal oblique, iliopsoas, transversospinalis, longissimus, and iliocostalis muscles.

Role in performance

The Chipmunk strengthens the muscles that provide stability to a dog's spine (4). Along with muscle development, the Chipmunk enhances the dog's whole body balance and proprioception (4). This increased stability, balance, and proprioception may enhance the dog's ability to produce power, improve endurance, and enhance agility (46, 47).

Role in injury prevention

Training exercises that enhance a dog's balance, proprioception, and core strength may protect them during uncontrolled movements which result in body misalignment, spinal hyperextension or compression, such as jumping, apprehension training, or ladder climbing (41, 43).

Progression

A Chipmunk should only be attempted once the dog has achieved consistent proper Posture Sits and Posture Downs. Achievement of the correct Posture Sit and Posture Down ensures that the dog has sufficient body awareness and muscle development of the epaxial and abdominal muscles. As described in the training section below, the dog should gradually be taught to engage its core in short training sessions before expecting the final Chipmunk position. Once a dog can achieve the proper form of a Chipmunk for a duration of 30 s (**Figure 6B**), it can progress to more advanced levels. To increase the difficulty of a Chipmunk, move the lure left and right, encouraging the dog to shift its weight. The further away from the dog that the lure is located, the more difficult the exercise becomes. Once the dog can maintain its balance for 30 s while weight shifting, they can progress. The final progression is to perform the Chipmunk on an uneven or unstable surface. Progression of the Chipmunk requires patience, hypervigilance of posture, and commitment to long-term training.

Contraindications

The Chipmunk is not recommended for dogs with suspected spine (cervical, thoracic, or lumbosacral), hip, or stifle abnormalities, unless under direct supervision of a veterinarian.

Training

Training sessions for the Chipmunk should be short, and the dog should not perform more than 3–6 repetitions during any single session. To begin to train the Chipmunk, place the dog in a Posture Sit and place a lure just above the dog's head to encourage lifting of the forelimbs. Reward as soon as the forelimbs leave the ground [<15 cm (6 in)] (**Figure 6A**). Once the dog lifts its forelimbs to 15 cm (6 in), the next progression is to have the dog lift the forelimbs to a height between the elbow and shoulder. Next, encourage the dog to progress to holding themselves with a straight spine, square sit, neutral head position, and stifles dorsal to the hindlimb digits. If the dog's stifles are not dorsal to the hindlimb digits the dog is not engaging its core musculature appropriately. Do not increase the duration of the Chipmunk hold until the dog can consistently hold itself in the proper posture (**Figure 6B**). If proper posture is achieved, then progress to the more advanced Chipmunk described in the progression section. A dog may require weeks to months to attain the proper posture in the Chipmunk. Signs that a dog is fatiguing are muscle fasciculations, reluctance to lift forelimbs, attempts to place forelimbs on an object/handler, or the stifles consistently moving caudally away from the hindlimb digits.

Squat

Description

To perform the Squat, the dog places its forepaws on an elevated object, sits onto a restricted area platform, and then returns to the starting position (**Figures 7A–E**). The Squat is performed for a specific duration or number of repetitions.

Purpose

The Squat primarily develops the dog's hip, stifle, and tarsus extension strength (48). Core stability (predominantly sagittal flexion) and hindlimb proprioception are secondarily developed. Hip extension is primarily provided by the gluteal (superficial, middle, and deep), semitendinosus, and semimembranosus muscles. The gracilis, piriformis, and quadratus femoris muscles also contribute to hip extension. Stifle extension is primarily provided by the quadriceps femoris, sartorius, tensor fasciae latae, and biceps femoris muscles. Tarsal extension is primarily provided by the gracilis, biceps femoris, and semitendinosus muscles.

Role in performance

Forceful hip, stifle, and tarsus extension developed from performing the Squat may increase the dog's ability to jump, sprint, and change direction. Dogs needing to jump up, onto an elevated surface, over an object, or across a gap may benefit from the Squat. Dogs rising from a down or sitting position or those rapidly accelerating while sprinting can benefit from the Squat. Dogs doing single-leg movements, climbing stairs, or rapidly changing direction may also perform better as a result of incorporating the Squat into their fitness plan (49).

Role in injury prevention

Strong hips, stifles, and tarsi developed by the Squat may be less prone to injury (50). Increased hip musculature may reduce the risk of hip dislocation and provide stability to a dysplastic hip (31). Developing stifle extension musculature may reduce cranial movement of the tibia relative to the femur and provide increased support to the cranial cruciate ligament. Dogs may also experience fewer or lower severity tarsal sprains or common calcanean tendon injuries (51).

Progression

For mature dogs, the Squat is usually trained on a 20 cm (8 in) high stable object (e.g., standard concrete block) (**Figure 7F**), although puppies under 3–4 months of age or small breed dogs may benefit from an object that is only 10 cm (4 in) high (e.g., standard brick). Once the dog is able to complete 7 repetitions revolutions in 30 s, the object height is increased by 20 cm (8 in) or by 10 cm (4 in) for young or small breed dogs (**Figures 7G–I**). Elevation of the object height progressively shifts a greater percentage of the dog's weight onto their hindlimbs. This process is continued until the next height progression would result in the dog's forepaws being elevated higher than the shoulder joint at the bottom of the movement (**Figure 7J**). Further progression is primarily provided by external weight in the form of a weight vest with weight increments scaled to the weight of the dog [e.g., for a 20 kg (44 lb) dog start with 2.3 kg (5 lb) of external weight]. Alternate methods of progression include destabilizing the hindpaw surface.

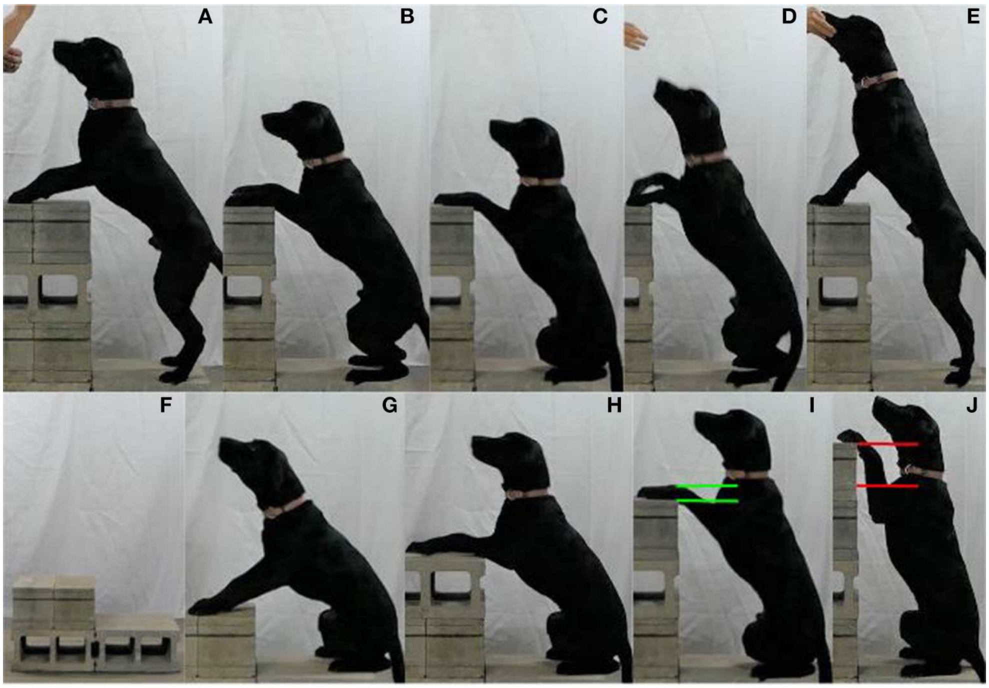


FIGURE 7 | (A–E) Illustrate the Squat. **(F)** Shows the restricted area platform. **(G–J)** Show progression of the Squat [**(G)** Level 1, **(H)** Level 2, **(I)** Level 3 (maximum for this dog), and **(J)** Level 4 (too high for this dog)].

Contraindications

Without guidance from a veterinarian, the Squat is not recommended for dogs with suspected spine, hip, stifle, or tarsus abnormalities. With supervised rehabilitation, the Squat may be adapted to the dog's injury or condition.

Training

A restricted area platform typically 40 cm (16 in) square (e.g., two standard concrete blocks) is used to maintain the forepaws on the elevated object and the hindpaws at a consistent distance from the object. While there are many ways to perform or train the Squat, the following works well for our population of working dogs. Have the dog step onto the platform and then onto the elevated object. Provide the verbal cue to "Sit" and use a food reward or toy to guide the dog's nose backward. Mark and reward any flexion of the hips, stifles, or tarsi. Progress toward a full sit with both forepaws on the elevated object and the majority of the hindlimb below the hock on the platform. Use a "Touch" command or lure the dog to return to the starting position. A soft high-value reward (e.g., small piece of cheese) assists our population of working dogs to rapidly return to the starting position, swallow the reward, and begin the next repetition.

Back-up

Description

To perform the Back-up exercise, beginning at a walk, the dog moves in a backwards or reverse motion by picking up and pushing off the ground with each foot to propel themselves backwards. The dog's spine should remain parallel with the floor during the entire movement (**Figure 8A**).

Purpose

Back-up is utilized for proprioception of the hindlimbs and primary targeting of the biceps brachii and quadriceps muscles (4, 24). The dog's normal gait is a pulling mechanism that primarily engages the triceps and hamstrings. The swing phase of the gait, which primarily involves the biceps and quadriceps muscles, is a passive motion (23, 45, 47). Backwards walking transitions the active and passive phases of the walking and trotting gait to build the biceps brachii, brachialis, and hamstring (semimembranosus and semitendinosus) muscles (24).

Role in performance

Performing the Back-up exercise focuses on the extensors of the hindlimb while complementing the flexion focused Squat exercise (24, 49, 52). The enhanced proprioceptive and activation of neuromuscular pathways that are not naturally targeted in a dog may increase athletic performance. Placement of backwards walking between the flexion focused exercises of the fitness circuit allows for active recovery time for the flexor muscles. The whole-body training approach may lead to balanced power, stamina, and enhance proprioception (24, 49, 52).

Role in injury prevention

Training of secondary muscle groups, activation of secondary gait neuromuscular pathways, and enhancing hindlimb proprioception may protect a dog from injury (3, 22, 28, 48, 49).

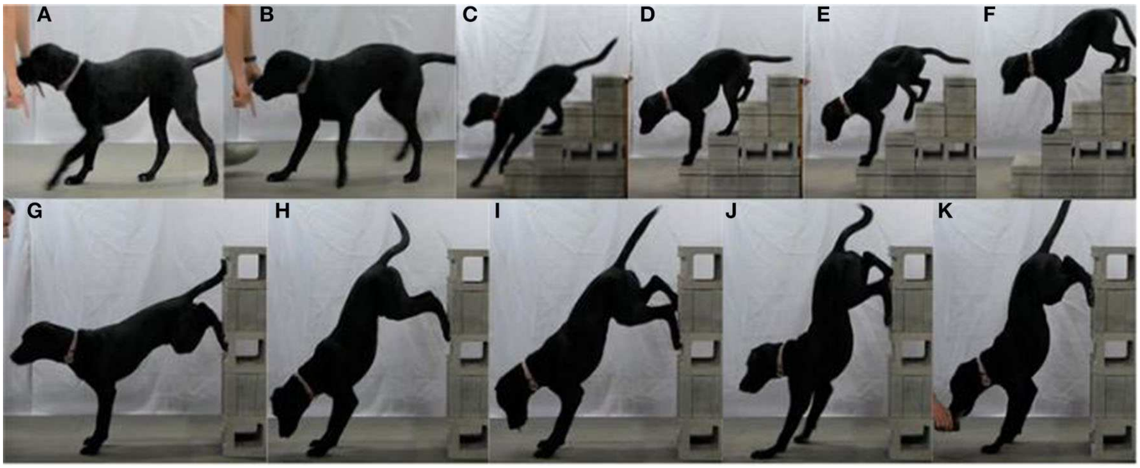


FIGURE 8 | (A,B) Show the first progression of Back-up on the ground. **(C–F)** Shows a more advanced progression of the Back-up, backing up stairs. **(G–K)** Shows another advanced progression of the Back-up, backing up a vertical wall.

Utilizing the Back-up during the fitness circuit maintains warmer tissue temperatures which improves tissue elasticity decreasing the risk of tissue damage during the remainder of the exercises (19, 20).

Contraindications

The Back-up is not recommended for dogs with suspected musculoskeletal abnormalities. The Back-up should be used with caution in dogs with diagnosed lumbosacral pain, hip dysplasia or osteoarthritis, or cruciate ligament disease.

Progression

Progression for the Back-up can be broken down into several stages (**Figures 8A–K**). The first goal is to have the dog walk backwards a distance of 3 m (9.8 ft) 3 times in 30 s. The gait must be smooth and intentional, with each paw leaving the ground independently. Dragging of the paws or a kyphotic spine is incorrect form and the dog should not progress to the next level of difficulty. After walking backwards, the dog progresses to trotting backwards with the same criteria as the walk. After trotting backwards, the Back-up progresses to backing up stairs. Each step should be 15–30 cm (6–10 in) in height. The dog must be able to back up 18 steps (may be split into 3 trips up and down 6 steps) in 30 s. The criteria for progression to the next level is that all four limbs must independently clear each step. Simultaneous "hopping" of the hindlimbs over the stairs is improper execution of the exercise. Following backing up the stairs, the Back-up progresses to backing up a vertical wall into a forelimb only "handstand".

Training

There are several ways to train the Back-up exercise. One method is to place the dog between two objects such as a wall and chair. The handler then steps in front of the dog with a lure in front of the dog's nose. The handler takes a step toward the dog and the dog will respond by stepping backwards. During initial training of the Back-up, the dog should be rewarded after each step backwards. Once the dog understands the movement, the reward frequency can be decreased. Moving toward the dog too quickly or placing the lure too low causes the dog to round its spine. Placing the lure too high causes the dog to sit. As a general principle for training the Back-up, a dog should take 2–3 steps backwards for every step forward by the handler.

Foundational Fitness Assessment

Overview

The Foundational Fitness Assessment (FFA) objectively measures a dog's fitness across the foundational fitness components. We developed this assessment to aid in measuring a dog's current fitness, identifying any change in fitness after completing a training program, adjusting training programs, and comparing the fitness of dogs in similar age ranges, breeds, or careers. Measuring a dog's current fitness was the primary impetus for the development of this assessment. We wanted a way to objectively assess the fitness of the dogs in the PVWDC training program in order to establish a baseline before they began the foundational training. We also desired a formalized assessment to determine if the training program we developed was making a difference in a dog's fitness. This kind of assessment will now allow us to compare different training program styles, methods, durations, frequency, and equipment. We hypothesized that dogs would become more fit during the course of a training program, and a formalized assessment would allow us to adjust that training program to maximize their fitness development. Finally, we wanted to be able to objectively compare the fitness of dogs. This will eventually allow us to develop age, breed, and career-specific standards and scoring.

We prioritized durability, accuracy, and simplicity when we designed the FFA. In this setting, a durable assessment is one that is easily implemented in a non-professional setting, involves inexpensive and readily available equipment, and is quick to conduct. Accuracy amongst evaluators (inter-rater and intra-rater reliability) and between assessments (inter-assessment and intra-assessment) are also important. Finally, we wanted a simple test that either does not require new behaviors or utilizes behaviors that could be rapidly learned. Our end state is an assessment that could work for most dogs, by most evaluators, and in most environments.

As a result of our experience during the pilot implementation, the FFA is divided into two levels. Level One consists of two tests that evaluate core stability and whole-body power generation (predominantly hindlimb extension strength) without the requirement for learned behaviors. Level Two utilizes two behaviors that take ∼4 weeks to learn but that provide an independent assessment of hindlimb extension strength and hindlimb stability. The two levels may be used separately or together, and future work will focus on validating their independent and consecutive use.

Foundational Fitness Assessment—Level One

Overview

The Foundational Fitness Assessment-Level One (FFA-L1) consists of the Sprint Test (ST) and the Progressive Plank Test (PPLT). This assessment requires minimal and inexpensive equipment and takes ∼15–20 min per dog to complete. While the order of the FFA-L1 requires further exploration, we propose that the PPLT should be completed after the ST so that core stability fatigue does not affect the maximal sprint. Some dogs may require the PPLT before the ST for behavioral reasons. Regardless of the order, the dog should be given at least 5 min to recover between tests.

Sprint Test

The ST assesses a dog's ability to generate whole-body power during the acceleration phase (initial 25 m) of sprinting. Sprinting is predominantly a hindlimb extension movement (53), so the ST is primarily an assessment of hip, stifle, and tarsal extension strength. To perform the ST, an area of flat, level, and smooth ground (preferably grass, dirt, or turf) at least 50 m (164 ft) long and 10 m (33 ft) wide must be identified. As the dog will perform a maximal effort attempt, every effort to ensure the dog's safety must be taken.

The dog must start from a down position (chest and tarsi on the ground) with its entire body behind the starting line. Light restraint may be used to minimize the obedience requirements. Accurately measure the 25 m (82 ft) course and place a narrow but conspicuous marker (e.g., cone) to define the finish line. Any method may be used to motivate the dog to sprint maximally, but the motivation should be located at least 10 m beyond the finish line to encourage maximal effort for the duration of the test. A toy reward may be held by the handler and a "Come" or "Here" command given. A toy may be thrown beyond the finish line, but this method is not preferred as some dogs will decrease their effort to track the toy in flight. For a dog trained to a bite sleeve, a decoy running away from the finish line may be used.

Due to the need for precise measurement and the inherent margin of error with manual timing the ST must be recorded using a video camera capable of at least 30 frames per second (accurate to 0.033 s) and preferably 60 frames per second (accurate to 0.017 s) (54). The camera should be positioned in line with the finish line and out of the way of the dog's path. The camera is aimed at the starting line to capture the start and then rotated to capture the finish. The video is then analyzed to determine the interval (in hundredths of a second) between the first motion of the dog and the first portion of the body to cross the finish line or object. The dog is given three attempts with at least 2 min rest between attempts. The ST score is expressed as the seconds (to the hundredths place) of the fastest attempt (e.g., 3.08 s).

Progressive Plank Test

The PPLT assesses a dog's trunk muscle endurance in a safe and objective manner. The plank (or prone bridge) is reliably used to assess human core muscle endurance (10, 44, 47, 55–59). To perform the PPLT, the dog's Plank measurements are first obtained (see above). Next, the dog is given 45 s in which to accumulate 30 s of proper Plank position at Level 2. If the dog successfully completes this stage, they are given 30 s of rest. The dog is then given 45 s in which to accumulate 30 s of proper Plank position at Level 4. If the dog successfully completes this stage, they are given 30 s of rest, and the process is repeated for Level 6. If the dog successfully completes this stage, they are given 30 s of rest before attempting a final maximum duration Plank at Level 8. The dog is allowed 15 s of improper position during this final stage. The PPLT is terminated when the dog steps off either block, fails to accumulate 30 s of proper Plank position within 45 s during the Level 2, 4, or 6 stages, or accumulates 15 s of improper Plank position during the final Level 8 stage. The PPLT score is expressed as the final level and seconds (rounded down to the nearest second) completed at that level. Examples are 4–0:24 (24 s at Level 4) or 8–2:15 (2 min and 15 s at Level 8).

Foundational Fitness Assessment—Level Two

Overview

The Foundational Fitness Assessment-Level Two (FFA-L2) consists of the Progressive Pivot Test (PPT) and the Progressive Squat Test (PST). This assessment requires ∼4–6 weeks of prior training, minimal and inexpensive equipment and takes ∼15–20 min per dog. While the order of the FFA-L2 requires further exploration, we propose that the PPT should be conducted before the PST so that fatigue from the Squat does not affect hip stability. The dog should be given at least 5 min to recover after completion of the final PPT interval before beginning the PST.

Progressive Pivot Test

The PPT assesses a dog's hip stability in a safe, objective, and specific manner. To perform the PPT, the dog completes up to 3 complete Pivot rotations in under 30 s in both directions at successively higher levels until the dog's maximum level is reached. The dog has 45 s of total time in which to complete 30 s of active Pivot. The 30 s active time is paused if the dog steps off the object or if there is a handler or reward issue. The interval ends when the dog successfully completes the 3 rotations, the 30 s of active time elapses, or the 45 s of total time elapses. After completion of a level in one direction, the dog is given 30 s of rest before attempting that level in the opposite direction.

If the dog successfully completes the level, the object height is increased to the next level. If the dog successfully completes a level in one direction but does not complete the level in the

TABLE 4 | Proposed Foundational Fitness Assessment scoring system.

Proposed FFA scoring system			
Score	**Description**	**Percentile range**	**Percentage of results (%)**
0	At risk	0th–10th	10
1	Minimal	11th–36th	25
2	Effective	37th–63rd	26
3	Excellent	64th–89th	25
4	Outstanding	90th–100th	10

other direction, the object height is not increased. This process is continued until either the dog fails to complete a level or the object height is raised to the dog's maximum level. Then, the dog completes a final 2 min maximum effort in each direction. The PPT score is expressed as the final level and number of rotations (rounded down to the nearest quarter rotation) in the clockwise (expressed first) and counterclockwise directions. Examples are 3–2.75/1.5 (two and three-quarter rotations clockwise and one and one-half rotations counterclockwise at Level 3) and M-10.25/12 (ten and one-quarter rotations clockwise and twelve rotations counterclockwise at this dog's maximum level).

Progressive Squat Test

The PST assesses a dog's hip, stifle, and tarsus extension strength and endurance in a safe, objective, and specific manner. To perform the PST, the dog completes up to 7 Squats in under 30 s at successively higher levels until the dog's maximum level is reached. The dog has 45 s of total time in which to complete 30 s of active Squats. The 30 s active time is paused if the dog steps off the object or platform or if there is a handler or reward issue. The interval ends when the dog successfully completes the 7 Squats, the 30 s of active time elapses, or the 45 s of total time elapses.

If the dog successfully completes the level, the object height is increased to the next level. After successful completion of a level, the dog is given 30 s of rest before attempting the next level. This process is continued until either the dog fails to complete a level or the object height is raised to the dog's maximum level. Then, the dog completes a final 2 min maximum effort. The PST score is expressed as the final level and number of Squats completed at that level. Examples include 2–5 (5 Squats at Level 2) and M-14 (14 Squats at this dog's maximum level).

Scoring and Standard Development

Our goal is to assess the foundational fitness of a sufficient number of dogs in order to develop both a scoring system and age, breed, and career-specific standards. We aim to develop a bell curve of results for each assessment. Then, we propose a scoring system where results clustered near the mean receive an average score while results above the mean receive higher scores and those below the mean receive lower scores. See **Table 4** for more detail.

We also propose further work to explore the relationship between foundational fitness results, injury, and objective career-specific performance measures. We are interested in identifying the effect a dog's fitness has on the likelihood it will experience an injury, the type of injury (acute or degenerative), the severity of injury, and the duration of recovery. We hypothesize that more fit dogs will be injured less frequently, experience fewer degenerative injuries, be injured less severely, and recover faster than their less fit counterparts. We also believe a dog's fitness is integral to its ability to perform its job or sport. While performance in some dog activities is easy to measure (e.g., distance in dock diving or time in agility), other dog activities are harder to measure or have more poorly defined performance metrics (e.g., explosive detection or urban search and rescue). We hypothesize that more fit dogs will perform better at their activities whether those activities are predominantly physical or less physically-focused.

RESULTS

Our aim was to develop a formalized method to develop and assess foundational fitness in working dogs. We implemented the FTW program in the PVWDC population over an ∼3-month period consisting of training personnel, familiarizing dogs, initial assessment, and regular training. Our pilot results are summarized below.

Safety

The FTW program was safe in this group of dogs, under the conditions tested. We assessed 31 dogs on two occasions without injury or negative effect on training. The same group of dogs conducted ∼600 foundational training sessions during the familiarization and regular training periods. We identified several minor abrasions from contact with the concrete blocks but no other injuries. The formalized Cool-down routine with its paw, pad, and nail check conducted after every training session facilitated identification and treatment of injuries before the dog left for the day. Without this system, injuries sustained during FTW or other training could have been overlooked or identified at home by a non-professional foster without medical resources.

The PVWDC veterinarians also experienced the common phenomenon in human athletic performance where the athletic trainer identifies an athlete's musculoskeletal issues before a medical professional does. On numerous occasions we had a non-trainer individual conduct FTW and then bring us a data-driven issue. Comments such as "I was just doing Pivot with Ivey, and both times she did 2–2.5 rotations clockwise but only 1–1.5 rotations counterclockwise. Can you come take a look?" resulted in a targeted musculoskeletal examination, identification of the issue (unilateral quadriceps femoris muscle soreness in this case), treatment, and rapid return to full performance.

Finally, we believe FTW to be safe for young dogs. The primary concerns with fitness training in these dogs are the effects of excessive force on epiphyseal plates and repetitive motion on the skeleton and joints (60–63). The resistance focus, short duration, and low-impact nature of the foundational fitness exercises and assessments is not likely to cause musculoskeletal damage, although monitoring is needed. In addition, regular joint loading in a structured and progressive manner in young dogs may decrease the risk of musculoskeletal damage (64–68).

Ease of Implementation

FTW appears to be an accessible method to develop and assess fitness in working dogs. Dogs of diverse breeds, temperaments, and ranging in age from 2 months to 3 years learned to perform each of the movements. Dogs were also appropriately challenged by the program. After initial familiarization, dogs who tested well on the Plank were then advanced to higher levels while dogs who scored lower on the Plank were provided lower progressions appropriate for them.

Puppies started learning proprioception and height-adjusted hip stability during their first week at the PVWDC, and dogs that had trained fitness for years in our program adjusted as well. We found that dogs that started FTW earlier (before 6 months of age), were exposed to the exercises more frequently, and practiced the exercises outside of the foundational training sessions rapidly learned the exercises and progressed. While dogs over 6 months of age and dogs that were only exposed to the exercises during the foundational training sessions took longer to learn the exercises, the limited number of exercises, repetition, and formalized criteria and progressions assisted their learning.

FTW is also an accessible program for the people involved in assessing and developing working dog fitness. The feedback from trainers, interns, and volunteers on the foundational fitness training program centered on simplicity and ease of progression. Our personnel found the reduced number of movements and formalized structure simple to implement. The clarified criterion and standardized levels made determining when and how to progress the difficulty of a movement easy.

The trainers, interns, and volunteers universally adopted the revised FTW program. The feedback we received identified an increase in perceived relevance to a working dog's future career as a major factor in this adoption. Also, the significantly reduced training requirements increased the confidence of trainers with less marker (e.g., clicker) training experience. We consistently observed trainers taking 2–3 weeks to become comfortable enough to teach an intern who then took the same length of time before teaching a volunteer. It took ∼4–6 weeks for the FTW knowledge to pass through four generations (FTW expert to trainer to intern to volunteer) and make a novice competent to perform the training.

DISCUSSION

We developed a formalized method to train and assess foundational fitness modalities for working dogs. The PVWDC FTW program incorporates posture development and frequent reinforcement of this behavior, warm-up and cool-down routines to prepare for and recover from training, methods to efficiently train foundational fitness, and a two-level format for assessing foundational fitness. We implemented this program in a working dog training facility and demonstrated the program's safety and ease of implementation.

Some limitations to this initial implementation are the young age of some of the dogs in our population, the prior exposure of our dogs and personnel to fitness training, the unique structure of the PVWDC, and the short duration of the implementation period. The young age of some of the dogs in our program allowed us to determine optimal methods for early familiarization with fitness exercises but limited our ability to see how more mature dogs would fare. In addition, PVWDC has a culture of fitness, and our personnel were accustomed to performing fitness training with our dogs. These factors likely shortened the learning curve for both our dogs and our personnel. While not completely unique among working dog programs, our pattern of bringing dogs into our program at 8 weeks of age and training with them on a daily basis until 12–24 months of age provided more and earlier contact time than some programs. Finally, the formalized and quantitative methods described in this paper represent an implementation period of ∼4 months. The number of refinements accomplished in that time indicate more are likely as we continue to train working dog fitness.

The aim of this paper was to initially describe these pilot techniques to assess and train foundational fitness for working dogs. Thus, further prospective studies are needed to validate the assessment components and training exercises. Also, this initial implementation allowed us to define initial progression levels for some exercises (Pivot, Plank, and Squat). The progression for other exercises (Posture Down, Chipmunk, and Back-up) is more generally described. Further work remains to be done to define the optimal method, sequence, and rate to progress each exercise.

Formalized canine fitness programs are rare. Even fewer are suitable for the functional fitness requirements and temperaments of working dogs or the time and logistical constraints of a kennel or training facility. In contrast, the human fitness realm has numerous methods to assess and train fitness for tactical athletes in comparable organizations (8, 11, 69–73). In addition, throughout this article we have cited research into the effectiveness of various human fitness exercises and methods; working dogs need similar evidence-based methods to assess and train fitness. We believe the PVWDC FTW program is one step in that direction.

We anticipate this work will serve to add momentum to the growing field of canine performance medicine. One gap deserving future research is the quantification of muscle activity for various fitness exercises. While these investigations have been started with rehabilitation movements (49), much work remains to be done to understand which exercises most effectively activate the desired musculature. Another area for future research is developing alternate exercises and progressions to develop foundational fitness. While we believe exercises like the Squat and Pivot are cornerstones of fitness, a dog's continued progression depends on incorporating alternate methods to develop the same fitness modalities.

In addition to the exercises themselves, we believe research needs to be done to provide evidence for the optimum programming and periodization methods. For training sessions, the frequency, timing throughout the day, and timing relative to work or other training need to be explored. The proper order of exercises within the session along with the optimum combination of sets, repetitions, and intensities needs to be determined. The canine fitness realm also needs safe and effective methods for developing strength and power similar to the barbell and kettlebell for human fitness.

Finally, formalized canine fitness must grow beyond these foundational fitness roots. Various working and sporting dog careers and disciplines have specific fitness requirements that should be layered on top of foundational fitness. Dogs need speed, power, endurance, and agility to perform in these careers, and their handlers and trainers need evidence-based training methods and assessments to help them improve those modalities.

We believe multidisciplinary collaboration is the key to unlocking progress toward filling these gaps. Working dog handlers, experienced trainers, and canine performance-oriented veterinarians should partner with a diverse array of similarly-oriented scientists to solve these challenges. The PVWDC is eager to collaborate with like-minded individuals, kennels, programs, and organizations.

ETHICS STATEMENT

The animal study was reviewed and approved by the University of Pennsylvania Institutional Animal Care and Use Committee.

AUTHOR CONTRIBUTIONS

BF and MR developed the method, conducted the pilot implementation, and participated in manuscript preparation. CO developed the method and participated in manuscript preparation. All authors contributed to the article and approved the submitted version.

FUNDING

Research reported in this publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number TL1TR001880.

ACKNOWLEDGMENTS

This work has been built on the experience of many experts. We appreciate their influence on our thinking and approach to canine physical fitness. We would like to acknowledge the trainers, interns, and volunteers at the PVWDC for their work in implementing the FTW program.

REFERENCES

1. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep.* (1985) 100:126–31.
2. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee I-M, et al. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc.* (2011) 43:1334–59. doi: 10.1249/MSS.0b013e318213fefb
3. Chong TE, Davies HMS. Investigating canine elbow joint stabilisation through mechanical constraints of the deep fascia and other soft tissues. *J Anat.* (2018) 232:407–21. doi: 10.1111/joa.12763
4. Zink MC, Van Dyke JB. *Canine Sports Medicine and Rehabilitation*. Arnes, AI: John Wiley & Sons, Incorporated (2011). Available online at: http://ebookcentral.proquest.com/lib/upenn-ebooks/detail.action?docID=1132530 (accessed May 8, 2020).
5. Cain B, Jones JC, Holásková I, Freeman L, Pierce B. Feasibility for measuring transverse area ratios and asymmetry of lumbosacral region paraspinal muscles in working dogs using computed tomography. *Front Vet Sci.* (2016) 3:34. doi: 10.3389/fvets.2016.00034
6. Adrian CP, Haussler KK, Kawcak CE, Reiser RF, Riegger-Krugh C, Palmer RH, et al. Gait and electromyographic alterations due to early onset of injury and eventual rupture of the cranial cruciate ligament in dogs: a pilot study. *Vet Surg.* (2019) 48:388–400. doi: 10.1111/vsu.13178
7. Otto CM, Cobb ML, Wilsson E. Editorial: working dogs: form and function. *Front Vet Sci.* (2019) 6:351. doi: 10.3389/fvets.2019.00351
8. Bear R, Sanders M, Pompili J, Stucky L, Walters A, Simmons J, et al. Development of the tactical human optimization, rapid rehabilitation, and reconditioning program military operator readiness assessment for the special forces operator. *Strength Cond J.* (2016) 38:55–60. doi: 10.1519/SSC.0000000000000258
9. Smith AJ, Lilley E. The role of the three Rs in improving the planning and reproducibility of animal experiments. *Animals.* (2019) 9:975. doi: 10.3390/ani9110975
10. O'Keeffe BT, Donnelly AE, MacDonncha C. Test-retest reliability of student-administered health-related fitness tests in school settings. *Pediatr Exerc Sci.* (2019) 32:48–57. doi: 10.1123/pes.2019-0166
11. Army Combat Fitness Test. *Army Combat Fitness Test.* Available online at: https://www.army.mil/acft (accessed May 7, 2020).
12. Definition of proprioception. *Dictionary.com. www.dictionary.com.* Available online at: https://www.dictionary.com/browse/proprioception (accessed May 7, 2020).
13. Smith CA. The warm-up procedure: to stretch or not to stretch. a brief review. *J Orthop Sports Phys Ther.* (1994) 19:12–7. doi: 10.2519/jospt.1994.19.1.12
14. Davies CT, Young K. Effect of temperature on the contractile properties and muscle power of triceps surae in humans. *J Appl Physiol.* (1983) 55:191–5. doi: 10.1152/jappl.1983.55.1.191
15. Safran MR, Seaber AV, Garrett WE. Warm-up and muscular injury prevention an update. *Sports Med.* (1989) 8:239–49. doi: 10.2165/00007256-198908040-00004
16. Munnings D. Comparison of the time dogs take to run over a set line of jumps both before and after an osteopathic treatment. *Eur School Osteopath*. 26.
17. Zink CM. *Conditioning the Canine Athlete*. 1st ed. Ellicott, CO: Canine Sports Productions (1997).
18. Ayala F, Calderón-López A, Delgado-Gosálbez JC, Parra-Sánchez S, Pomares-Noguera C, Hernández-Sánchez S, et al. Acute effects of three neuromuscular warm-up strategies on several physical performance measures in football players. *PLoS ONE.* (2017) 12:e0169660. doi: 10.1371/journal.pone.0169660
19. Škof B, Strojnik V. The effect of two warm-up protocols on some biomechanical parameters of the neuromuscular system of middledistance runners. *J Strength Cond Res.* (2007) 21:394–9. doi: 10.1519/00124278-200705000-00018
20. Andersen JC. Stretching before and after exercise: effect on muscle soreness and injury risk. *J Athl Train.* (2005) 40:218–20.
21. Hrysomallis C. Injury incidence, risk factors and prevention in Australian rules football. *Sports Med.* (2013) 43:339–54. doi: 10.1007/s40279-013-0034-0
22. de la Motte SJ, Lisman P, Gribbin TC, Murphy K, Deuster PA. Systematic review of the association between physical fitness and musculoskeletal injury risk: part 3—flexibility, power, speed, balance, and agility. *J Strength Cond Res.* (2019) 33:1723–35. doi: 10.1519/JSC.0000000000002382

23. Witvrouw E, Mahieu N, Danneels L, McNair P. Stretching and injury prevention. *Sports Med.* (2004) 34:443–9. doi: 10.2165/00007256-200434070-00003
24. Buford JA, Zernicke RF, Smith JL. Adaptive control for backward quadrupedal walking. I. posture and hindlimb kinematics. *J Neurophysiol.* (1990) 64:745–55. doi: 10.1152/jn.1990.64.3.745
25. Chiou S-Y, Gottardi SEA, Hodges PW, Strutton PH. Corticospinal excitability of trunk muscles during different postural tasks. *PLoS ONE.* (2016) 11:e0147650. doi: 10.1371/journal.pone.0147650
26. Boling M, Padua D. Relationship between hip strength and trunk, hip, and knee kinematics during a jump-landing task in individuals with patellofemoral pain. *Int J Sports Phys Ther.* (2013) 8:661–9.
27. Blazevich AJ. Optimizing hip musculature for greater sprint running speed. *Strength Cond J.* (2000) 22:22. doi: 10.1519/00126548-200004000-00007
28. Neptune RR, Wright IC, van den Bogert AJ. Muscle coordination and function during cutting movements. *Med Sci Sports Exerc.* (1999) 31:294–302. doi: 10.1097/00005768-199902000-00014
29. Powers CM. The influence of altered lower-extremity kinematics on patellofemoral joint dysfunction: a theoretical perspective. *J Orthop Sports Phys Ther.* (2003) 33:639–46. doi: 10.2519/jospt.2003.33.11.639
30. Powers CM, Ghoddosi N, Straub RK, Khayambashi K. Hip strength as a predictor of ankle sprains in male soccer players: a prospective study. *J Athl Train.* (2017) 52:1048–55. doi: 10.4085/1062-6050-52.11.18
31. Dycus DL, Levine D, Marcellin-Little DJ. Physical rehabilitation for the management of canine hip dysplasia. *Vet Clin North Am Small Anim Pract.* (2017) 47:823–50. doi: 10.1016/j.cvsm.2017.02.006
32. Ginos BE, Kozuchowski J, Vaughn AS, Krause DA, Hollman JH, Ginos BE, et al. Relationships between knee valgus, hip-muscle strength, and hip-muscle recruitment during a single-limb step-down. *J Sport Rehabil.* 18:104–17. doi: 10.1123/jsr.18.1.104
33. Read PJ, Oliver JL, De Ste Croix MBA, Myer GD, Lloyd RS. Neuromuscular risk factors for knee and ankle ligament injuries in male youth soccer players. *Sports Med.* (2016) 46:1059–66. doi: 10.1007/s40279-016-0479-z
34. Oliveira R, Vaz JR, Silva L, Pezarat-Correia P, Correia JP, Vaz JR, et al. Trunk muscle activation, fatigue and low back pain in tennis players. *J Sci Med Sport.* (2016) 19:311–6. doi: 10.1016/j.jsams.2015.04.002
35. Calatayud J, Casaña J, Martín F, Jakobsen MD, Colado JC, Gargallo P, et al. Trunk muscle activity during different variations of the supine plank exercise. *Musculoskelet Sci Pract.* (2017) 28:54–8. doi: 10.1016/j.msksp.2017.01.011
36. Fujita S, Kusano S, Sugiura Y, Sakuraba K, Kubota A, Sakuma K, et al. A 100-m sprint time is associated with deep trunk muscle thickness in collegiate male sprinters. *Front Sports Act Living.* (2019) 1:32. doi: 10.3389/fspor.2019.00032
37. Kibler WB, Press J, Sciascia A. The role of core stability in athletic function. *Sports Med.* (2006) 36:189–98. doi: 10.2165/00007256-200636030-00001
38. Craven BR, Chilibeck PD, Spink KS, Grona SL, Butcher SJ, Craven BR, et al. The effect of trunk stability training on vertical takeoff velocity. *J Orthop Sports Phys Ther.* 37:223–31. doi: 10.2519/jospt.2007.2331
39. Hung K-C, Chung H-W, Yu CC-W, Lai H-C, Sun F-H. Effects of 8-week core training on core endurance and running economy. *PLoS ONE.* (2019) 14:e0213158. doi: 10.1371/journal.pone.0213158
40. Tong TK, Wu S, Nie J, Baker JS, Lin H. The occurrence of core muscle fatigue during high-intensity running exercise and its limitation to performance: the role of respiratory work. *J Sports Sci Med.* (2014) 13:244–51.
41. Rivera CE. Core and lumbopelvic stabilization in runners. *Phys Med Rehabil Clin North Am.* (2016) 27:319–37. doi: 10.1016/j.pmr.2015.09.003
42. Mokha M, Sato K. Does core strength training influence running kinetics, lower-extremity stability, and 5000-m performance in runners? *J Strength Cond Res.* 23:133–40. doi: 10.1519/JSC.0b013e31818eb0c5
43. Burnham J, Mc Y, Ke R, R M, Br W, Dl J, et al. Relationship of hip and trunk muscle function with single leg step-down performance: implications for return to play screening and rehabilitation. *Phys Ther Sport.* (2016) 22:66–73. doi: 10.1016/j.ptsp.2016.05.007
44. Imai A, Kaneoka K. The relationship between trunk endurance plank tests and athletic performance tests in adolescent soccer players. *Int J Sports Phys Ther.* (2016) 11:718–24.
45. Greene FS, Perryman E, Cleary CJ, Cook SB. Core stability and athletic performance in male and female lacrosse players. *Int J Exerc Sci.* (2019) 12:1138–48.
46. Huxel Bliven KC, Anderson BE. Core stability training for injury prevention. *Sports Health.* (2013) 5:514–22. doi: 10.1177/1941738113481200
47. Tong TK, Wu S, Nie J. Sport-specific endurance plank test for evaluation of global core muscle function. *Phys Ther Sport.* (2014) 15:58–63. doi: 10.1016/j.ptsp.2013.03.003
48. Ellis RG, Rankin JW, Hutchinson JR. Limb kinematics, kinetics and muscle dynamics during the sit-to-stand transition in greyhounds. *Front Bioeng Biotechnol.* (2018) 6:162. doi: 10.3389/fbioe.2018.00162
49. McLean H, Millis D, Levine D. Surface electromyography of the vastus lateralis, biceps femoris, and gluteus medius in dogs during stance, walking, trotting, and selected therapeutic exercises. *Front Vet Sci.* (2019) 6:211. doi: 10.3389/fvets.2019.00211
50. Tate J, Suckut T, Wages J, Lyles H, Perrin B. The associations between hip strength and hip kinematics during a single leg hop in recreational athletes post acl reconstruction compared to healthy controls. *Int J Sports Phys Ther.* (2017) 12:341–51.
51. De Ridder R, Witvrouw E, Dolphens M, Roosen P, Van Ginckel A. Hip strength as an intrinsic risk factor for lateral ankle sprains in youth soccer players: a 3-season prospective study. *Am J Sports Med.* (2017) 45:410–6. doi: 10.1177/0363546516672650
52. Goslow GE, Seehermanf HJ, Taylor CR, Mccutchln MN, Heglund NC. Electrical activity and relative length changes of dog limb muscles as a function of speed and gait. *J Exp Biol.* (1981) 94:15–42.
53. Williams SB, Wilson AM, Rhodes L, Andrews J, Payne RC. Functional anatomy and muscle moment arms of the pelvic limb of an elite sprinting athlete: the racing greyhound (Canis familiaris). *J Anat.* (2008) 213:361–72. doi: 10.1111/j.1469-7580.2008.00961.x
54. Faux DA, Godolphin J. Manual timing in physics experiments: error and uncertainty. *Am J Phys.* (2019) 87:110–5. doi: 10.1119/1.5085437
55. Durall CJ, Greene PF, Kernozek TW. A comparison of two isometric tests of trunk flexor endurance. *J Strength Cond Res.* (2012) 26:1939–44. doi: 10.1519/JSC.0b013e318237ea1c
56. Boyer C, Tremblay M, Saunders T, McFarlane A, Borghese M, Lloyd M, et al. Feasibility, validity, and reliability of the plank isometric hold as a field-based assessment of torso muscular endurance for children 8–12 years of age. *Pediatr Exerc Sci.* (2013) 25:407–22. doi: 10.1123/pes.25.3.407
57. Bohannon RW, Steffl M, Glenney SS, Green M, Cashwell L, Prajerova K, et al. The prone bridge test: performance, validity, and reliability among older and younger adults. *J Bodyw Mov Ther.* (2018) 22:385–9. doi: 10.1016/j.jbmt.2017.07.005
58. Gonzalez SL, Diaz AM, Plummer HA, Michener LA. Musculoskeletal screening to identify female collegiate rowers at risk for low back pain. *J Athl Train.* (2018) 53:1173–80. doi: 10.4085/1062-6050-50-17
59. Strand SL, Hjelm J, Shoepe TC, Fajardo MA. Norms for an isometric muscle endurance test. *J Hum Kinet.* (2014) 40:93–102. doi: 10.2478/hukin-2014-0011
60. O'brien TR, Morgan JP, Suter PF. Epiphyseal plate injury in the dog: a radiographic study of growth disturbance in the forelimb. *J Small Anim Pract.* (1971) 12:19–36. doi: 10.1111/j.1748-5827.1971.tb05631.x
61. Arokoski J, Kiviranta I, Jurvelin J, Tammi M, Helminen HJ. Long-distance running causes site-dependent decrease of cartilage glycosaminoglycan content in the knee joints of beagle dogs. *Arthritis Rheum.* (1993) 36:1451–9. doi: 10.1002/art.1780361018
62. Lammi MJ, Häkkinen TP, Parkkinen JJ, Hyttinen MM, Jortikka M, Helminen HJ, et al. Adaptation of canine femoral head articular cartilage to long distance running exercise in young beagles. *Ann Rheum Dis.* (1993) 52:369–7. doi: 10.1136/ard.52.5.369
63. Puustjärvi K, Karjalainen P, Nieminen J, Helminen HJ, Soimakallio S, Kivimäki T, et al. Effects of Long-term running on spinal mineral content in dogs. *Calcif Tissue Int.* (1991) 49:S81–2. doi: 10.1007/BF02555099
64. Dahab KS, McCambridge TM. Strength training in children and adolescents: raising the bar for young athletes? *Sports Health.* (2009) 1:223–6. doi: 10.1177/1941738109334215
65. Guy JA, Micheli LJ. Strength training for children and adolescents. *JAAOS.* (2001) 9:29–36. doi: 10.5435/00124635-200101000-00004

66. Faigenbaum AD, Myer GD. Resistance training among young athletes: safety, efficacy and injury prevention effects. *Br J Sports Med.* (2010) 44:56–63. doi: 10.1136/bjsm.2009.068098
67. Benjamin HJ, Glow KM. Strength training for children and adolescents: what can physicians recommend? *Phys Sportsmed.* 31:19–26. doi: 10.3810/psm.2003.09.482
68. Faigenbaum AD, Lloyd RS, MacDonald J, Myer GD. Citius, altius, fortius: beneficial effects of resistance training for young athletes: narrative review. *Br J Sports Med.* (2016) 50:3–7. doi: 10.1136/bjsports-2015-094621
69. *Work Capacity Test Information | US Forest Service*. Available online at: https://www.fs.usda.gov/managing-land/fire/safety/wct (accessed May 7, 2020).
70. Anderson GS, Plecas D, Segger T. Police officer physical ability testing – re-validating a selection criterion. *Policing Int J Police Strateg Manage.* (2001) 24:8–31. doi: 10.1108/13639510110382232
71. Army Physical Readiness Training (PRT) Information. *ArmyPRT.com.* Available online at: http://www.armyprt.com (accessed May 7, 2020).
72. TSAC Professionals. Available online at: https://www.nsca.com/professional-development/professional-areas/tsac-professionals/ (accessed May 7, 2020).
73. HITT Programs. Available online at: https://www.fitness.marines.mil/hitt_programs/ (accessed May 7, 2020).

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer LW declared a past co-authorship with one of the authors CO to the handling editor.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 14 August 2020
doi: 10.3389/fvets.2020.00528

Computed Tomographic Evaluation of the Sacroiliac Joints of Young Working Labrador Retrievers of Various Work Status Groups: Detected Lesions Vary Among the Different Groups and Finite Element Analyses of the Static Pelvis Yields Repeatable Measures of Sacroiliac Ligament Joint Strain

Michael Carnevale[1], Jeryl Jones[1], Gang Li[2], Julia Sharp[3], Katherine Olson[1] and William Bridges[4]*

[1] Department of Animal and Veterinary Sciences, Clemson University, Clemson, SC, United States, [2] Department of Mechanical Engineering, Clemson University, Clemson, SC, United States, [3] Department of Statistics, Colorado State University, Fort Collins, CO, United States, [4] School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC, United States

OPEN ACCESS

Edited by:
Wendy Irene Baltzer, Massey University, New Zealand

Reviewed by:
Joseph Wakshlag, Cornell University, United States
Jennifer M. Gambino, IDEXX Laboratories, Inc., United States

***Correspondence:**
Jeryl Jones
jerylj@clemson.edu

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences, a section of the journal Frontiers in Veterinary Science

Received: *14 February 2020*
Accepted: *08 July 2020*
Published: *14 August 2020*

Citation:
Carnevale M, Jones J, Li G, Sharp J, Olson K and Bridges W (2020) Computed Tomographic Evaluation of the Sacroiliac Joints of Young Working Labrador Retrievers of Various Work Status Groups: Detected Lesions Vary Among the Different Groups and Finite Element Analyses of the Static Pelvis Yields Repeatable Measures of Sacroiliac Ligament Joint Strain. Front. Vet. Sci. 7:528. doi: 10.3389/fvets.2020.00528

Musculoskeletal injuries can lead to a working dog being withdrawn from service prior to retirement. During training exercises, young working dogs are often required to perform repetitive tasks, including adoption of an upright posture (or "hupp" task). Non-invasive, quantitative methods would be helpful for supporting research on effects of these repetitive tasks on sacroiliac joints (SIJ). Furthering our understanding of lesions in and biomechanical stresses on the SIJ could provide insight into possible training modifications for minimizing risks of SIJ injury. Aims of this retrospective, secondary analysis, exploratory study were to test hypotheses that (1) mean numbers of SIJ computed tomographic (CT) lesions/dog would differ among work status groups in young working Labrador Retrievers; (2) a methodology for using CT data and finite element analysis (FEA) to quantify SIJ ligament strain in the static canine pelvis would be feasible; and (3) this FEA methodology would yield repeatable measures of SIJ ligament strain. Clinical and CT data for 22 Labrador retriever working dogs, aged 11–48 months, were retrospectively reviewed. Dogs were categorized into three work status groups (Breeder, Detection, Other). A veterinary radiologist who was unaware of dog group status recorded numbers of CT lesions for each SIJ, based on previously published criteria. Mean numbers of SIJ CT lesions/dog were compared among dog work status groups. An *a priori* FEA model was created from the CT images of one of the dogs using image analysis software packages. Using tissue properties previously published for the human pelvis, various directional loads ($n = 8$) and forces (48 ligament strain values) were placed on the canine model in five trials. Repeatability was tested using regression analysis. There was a significantly greater mean number of subchondral sclerosis lesions in left SIJ of Breeder vs. Detection dogs, a significantly greater mean number of subchondral cysts in right SIJ for Detection vs. Breeder dogs, and a significantly greater mean number of subchondral cysts in right SIJ of Other vs. Breeder dogs ($p < 0.05$). Finite element modeling and analysis using CT data was feasible and yielded repeatable results in 47/48 (98%) of tests at each combination of strain, ligament, and side.

Keywords: computed tomography, sacroiliac joint, finite element analysis, finite element modeling, working dogs, sacroiliac joint disease

INTRODUCTION

Working dogs are important contributors for police, security, search and rescue, and military missions throughout the world (1–4). Detection training and working tasks often require working dogs to repetitively assume an upright stance ("hupp" position), with all of their weight placed on their hind legs and therefore with increased forces being applied to their sacroiliac joints (SIJ) (1). It has been proposed that, because SIJ are innervated with pain receptors, injury, or degenerative disease may be one of the causes of lower back pain in working dogs (1, 5). Rigorous training in repetitive agility-type motions in young dogs has also been proposed to increase the likelihood of developing chronic joint injuries later in life due to the under-developed physes (6). In people, this premise has been supported, given that most spinal injuries in young athletes have been found to occur after a sudden increase in the intensity and frequency of training (7–9). This clinical problem has been termed "overtraining" (10). Bone scintigraphy of the sacroiliac joint (SIJ) in young human athletes reporting lower back pain showed increased radiopharmaceutical uptake in one or both joints signifying increased bone turnover in athletes reporting lower back pain without any known specific trauma or reported radiographic abnormalities (7, 9). Cumulative musculoskeletal injuries are the most common injuries among working dogs and an important cause of early retirement (1, 3, 4). Published studies describing sacroiliac joint (SIJ) lesions in working dogs and methods for assessing possible effects of working tasks on canine SIJ are currently lacking.

The complex anatomy of canine SIJ has been previously detailed (11, 12). The SIJ consist of both synovial and fibrocartilaginous components. Joints are surrounded by the sheet-like dorsal and ventral sacroiliac ligaments. These ligaments play a role in stabilization of the SIJ and pelvis. The SIJ is also supported by interosseous ligaments connecting the articular surfaces of the sacral and ilial wings and a sacrotuberous ligament that connects the caudodorsal margins of the S3 vertebra to the dorsal margins of the ischiatic tuberosities of the ischium. The sacrotuberous ligament also plays a role in limiting pelvic range of motion. It has been proposed that hormonal changes in intact female dogs could predispose them to developing calcifications in the SIJ (5, 13). In puppies, the paired right and left hemispheres of the pelvis are joined together by a pelvic symphysis (11). In puppies, this is a fibrous ligament. As the dog ages, that ligament starts to ossify merging the two hemispheres into one. There is limited evidence evaluating the effects of strenuous exercise in dogs with open physes, however studies in rats report conflicting evidence that, while exercise may limit longitudinal bone growth and cause epiphyseal trabecular thinning in young animals, it may also lead to increased bone mineral density, cortical thickness and improved load bearing (14, 15).

Non-invasive techniques for quantifying SIJ lesions and theoretically quantifying stress and strain in sacroiliac joint ligaments would be helpful for supporting future research on effects of working and/or training tasks on the SIJ of young working dogs. Furthering our understanding of the presence of lesions in-, and biomechanical stresses on the SIJ could provide necessary insight into detection, effective screening, surveillance, and training modifications for mitigating conditions that increase the risk of SIJ injury and modification of training protocols for optimal performance and career longevity. Computed tomography (CT) has been validated as a method for characterizing SIJ lesions based on comparisons with gross pathology in a study of cadaver canine specimens (16). However, descriptions of CT SIJ lesions in working dogs have not been published. Mechanical stress and strain forces applied to sacroiliac joints have been estimated for dogs of varying weights evaluating only bony components (17). Authors of that report concluded that sacroiliac joint biomechanics were likely affected primarily by dog weight and the shapes of the bony components of the sacroiliac joints. However, soft tissue components were not evaluated. Finite element analysis (FEA) is the process of running numerous mathematical calculations on a computer-generated model that is broken down into simple geometry (mesh) (18). This model can then be used in a simulation to predict outcomes based off user-input. Previous human studies have demonstrated that FEA can model stress and strain in the SIJ and other pelvic structures (18–24). Computed tomography (CT) or magnetic resonance imaging (MRI) techniques were used in these studies to construct three-dimensional (3D) models based on patient-specific anatomy and conformation. Forces are applied to any specific spot on the extrapolated 3D model to calculate stresses and strains in different areas of that model. However, these previous studies did not describe repeatability of FEA measurements. Finite element analysis methods for modeling the canine lumbar spine (25) pelvis (26), and femur (27, 28) have been published. At the time of the current study, no peer-reviewed publications were found describing FEA methodologies for modeling canine SIJ ligament strain based on patient-specific CT data. Factors contributing to SIJ lesions such as acute injury and chronic disease in working dog populations are poorly understood. A computational SIJ model would allow for complex modeling of the associated bone, ligaments and muscles; and unconstrained motion and modulation of the inherent anatomical and structural features. The 3-D model could be subjected to physiological and injury-inducing external loads for scrutiny of response; therein revealing underlying structural and mechanical properties inherent to the SIJ bones while under dynamic or kinematic simulation and ligament, muscle, and motions loads.

The current, preliminary, two-part study explored two methodologies for possible future use by researchers interested in quantifying effects of working tasks on canine sacroiliac joints: numbers of SIJ CT lesions/dog and SIJ ligament strain calculated using FEA. The first part of the study tested the hypothesis that numbers of SIJ CT lesions/dog would differ among work status groups in a sample of young Labrador retriever working dogs. The second part of the study tested hypotheses that a methodology for conducting FEA of SIJ ligament strain in the static canine pelvis using patient-specific CT data would be feasible and that models based on this methodology would yield repeatable measures of SIJ ligament strain. Long term goals were to lend insight into dynamic physiologic, anatomic, kinematic, and 3-D mechanical conditions that have the potential to lead to working dog SIJ lesions. Findings could also play a role in future predictive modeling and identifying critical conditions and mechanisms influencing the proclivity of SIJ injury particularly in working dogs that may have otherwise gone undetected, or worse been a source of morbidity leading to abbreviated service.

METHODS

Dogs

The study was a retrospective, secondary analysis, exploratory design. Labrador retriever working dogs, aged 11–48 months, that had undergone lumbosacral CT scanning for two, previous research studies were included (29, 30). Due to the retrospective nature of the study, institutional animal care, and use committee approval was not required. However, investigators in the previous research studies granted permission to retrieve and use the clinical and CT data for the current study. Scans had been acquired at two referral hospitals using 16-slice CT scanners and standardized protocols (Ryan Veterinary Hospital, PennVet, University of Pennsylvania, Brightspeed S, GE Medical systems, 0.625 mm slice thickness, body filter, bone convolution kernel; and the LTC Daniel E. Holland Memorial Military Working Dog Hospital, Lackland Joint Base, San Antonio, Lightspeed VCT, GE Medical Systems, 0.625 mm slice thickness, body filter, standard convolution kernel). All dogs had been placed in dorsal recumbency with the hips positioned in maximal flexion and maximal extension for scanning. To minimize outside sources of variation due to positioning, all interpretations and analyses for the current study were based on scans that had been acquired in the flexed hip position.

Part 1: Comparisons Between Numbers of SIJ CT Lesions/Dog and Dog Work Status

A total of 22 dogs met the inclusion criteria and were classified into the following groups for analyses: sex (male, female), neuter status (intact, spayed, neutered), age (younger, 11–30 months; and older, 31–48 months), and work status at the time of CT scanning (Detection, Breeder, and Other). Dogs listed as "in-training detection" or "detection" in their records were classified into the Detection work status group. Intact female dogs with no described work status in their records were classified into the Breeder work-status group. Dogs that were not intact female dogs and that had no described work status were classified into the Other work status group. An undergraduate research student (KO) and graduate student (MC) made consensus decisions for dog group classifications and clinical data recording, without knowledge of CT findings.

The CT scans for included dogs were retrieved and reviewed by an ACVR-Board Certified Veterinary Radiologist (JJ), who was unaware of clinical findings or dog group status at the time of data recording (Horos Version 2.0.2 on MacOS Sierra Version 10.12.3 IMac by Apple Inc). Scans were interpreted in random order and the SIJ to be evaluated first (left or right) was determined by a coin toss. Numbers of the following CT lesions were recorded for each SIJ, based on criteria described in a previous publication (16): subchondral sclerosis, subchondral cysts, subchondral erosions, subarticular clefts, intra-articular ankylosis, and para-articular ankylosis. The radiologist used 3D multiplanar reformatting and adjusted window/level settings as needed for making decisions. Lesions had to be detected in at least two orthogonal planes before they were recorded as present. Additional observations at the time of interpretation were also recorded.

In consultation with a statistician (WB), the undergraduate research student (KO) selected and performed statistical comparisons of mean numbers of SIJ CT lesions/dog among dog groups. All statistical analyses were performed using commercially available statistics software (JMP Pro Statistical Software). Analysis of Variance (ANOVA) was performed to determine if there were significant differences in the mean number of SIJ lesions/dog among the work status groups. Assumptions for ANOVA were assessed and found to be satisfied. If evidence of difference among the group means was found, Fisher's Protected Least Significant Difference (LSD) Test was used to compare the three, work status group means. This method was chosen because this was an exploratory study; Fisher's Protected LSD is appropriate to reduce the chance of missing important differences to be examined in the future. Effects of sex and age on the work status group mean comparisons were evaluated by repeating the group comparison analysis with an ANOVA that included the main effects of sex and age, as well as the interactions of work status group with sex and age. Also because this was an exploratory study, any

$p < 0.10$ was considered evidence of possible differences in the work groups.

Part 2: Development of an FEA Method for Modeling Canine SIJ Ligament Strain Using CT Data and Repeatability of Ligament Strain Measurements

Design and Inclusion Criteria

The CT scans were further evaluated for a dog meeting the following criteria: (1) the CT scan had to include images of the complete pelvis and had to demonstrate minimal or no SIJ CT lesions with the intention to minimize possible disease-based effects on SIJ ligament strain values; (2) the CT scan had to include both ischiatic tuberosities so that all sacroiliac ligament attachments could be modeled; (3) the CT scans also had to include a bone algorithm study with 0.625 mm slice thickness such that boney edges could be clearly distinguished for segmentation (31). These decisions for subject selection were made by an ACVR-certified Veterinary Radiologist (JJ).

The dog selected for this part of the study was a 25 kg, 20-month-old, female, purebred Labrador retriever working dog. The pelvic CT scan DICOM files from this dog were exported to a personal computer (Lenovo ThinkPad S1 Yoga, Intel Core i5-4200U CPU 1.60 GHz, 8.00 GB, Hong Kong, China; Microsoft Windows 10, Redmond, WA) and a series of pilot studies were conducted to develop the FEA methods described below (32).

Segmentation Procedures

Segmentation of the individual pelvic bones (sacrum, ilium, ischium, acetabulum, pubis, and first caudal vertebra) was done using transverse CT images (**Figure 1**) and a three-dimensional (3D) image analysis freeware program (3D Slicer, version 4.5.0, http://www.slicer.org). A semi-automated process using a tool called the "threshold effect" was first run throughout all the imported digital imaging and communications in medicine (DICOM) CT image files for segmentation of the majority of pelvic bones. A total of 226 CT images were used for the segmentation. Next, a "paint effect" tool was manually drawn on the remainder of bone not covered by the "threshold effect" tool. This was conducted on each image using the three different planes (transverse, sagittal, and dorsal). Each bone was segmented using a different color label for ease of manipulation. Special care was especially taken to manually trace margins in areas where the bones came in close contact with each other (the sacrum and ilia). Once the individual pelvic bones listed above were segmented, a surface model was created by exporting each bone segmentation as a stereolithography (.STL) file. Next, each STL surface model was individually imported into a mesh generation software (ICEM CFD 17.0) to create surface mesh based off the geometric .STL file. The surface mesh of each bone was set to roughly have the same size elements such that the small features of bones are reasonably modeled to the live specimen. The surface meshes were then exported to a .STL file once again.

To enable solid model operations of the bones (i.e., Boolean operations) of the bones, the surface mesh .STL file is then converted to an Initial Graphics Exchange Specification (.IGS) file using a 3D solid modeling tool (SolidWorks 2016, Waltham, MA). After the geometric model was in the correct file format, the engineering simulation software (ANSYS Workbench) was used for setting up the geometry, creating the ligaments and joints, generating the finite element mesh, assigning the material properties, and performing the finite element analysis.

Material Properties of Bone, Ligament, and Joints

The pelvic bones, sacroiliac joints, sacrocaudal joints, and sacroiliac joint ligaments were included in the model and all assumptions for the model were based on human pelvic tissue properties (21) (**Table 1**). The ligaments were modeled to resist only tension (i.e., there was no stiffness in compression).

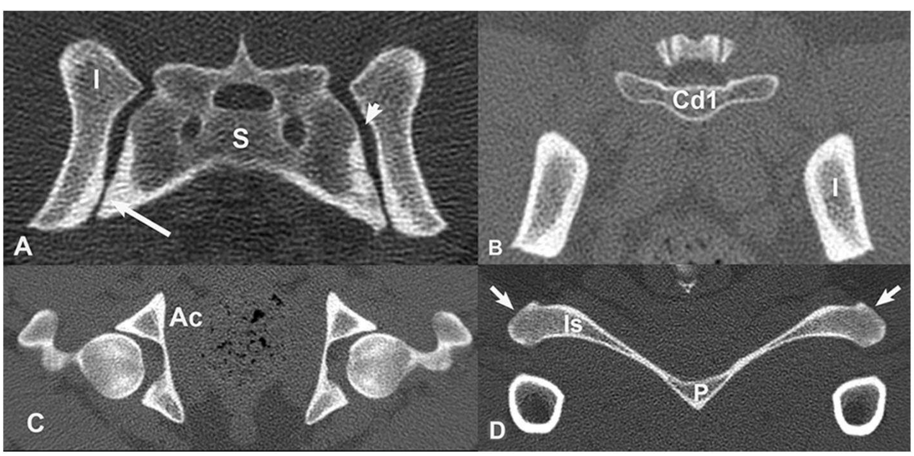


FIGURE 1 | Transverse, bone window CT images illustrating pelvic bone structures that were segmented for construction of the three-dimensional model. **(A)** S, sacrum; I, ilium; Large arrow = synovial component of sacroiliac joint space. Small arrow = fibrous component of sacroiliac joint space. **(B)** Cd1 = first caudal vertebra, I, ilium; **(C)** A, acetabulum; **(D)** Is, ischium; P, pubis; arrows = ischiatic tuberosities. Images are displayed with dorsal at the top and the patient's left to the viewer's right.

TABLE 1 | Isotropic elasticity properties that were used for modeling the pelvic bones in the current canine study.

Youngs modulus (Pa)	Posson's ratio	Bulk modulus (Pa)	Shear modulus (Pa)
1.7E+10	0.3	1.4167E+10	6.5385E+09

These were previously published for human pelvic bones at room temperature (27°C) (21).

TABLE 2 | Values of ligament stiffness as function of tensile strain that were used in the SIJ FEA model.

Ligament strain	<2.5%	<5%	<10%	>10%
Stiffness (N/mm)	39	55	103	100

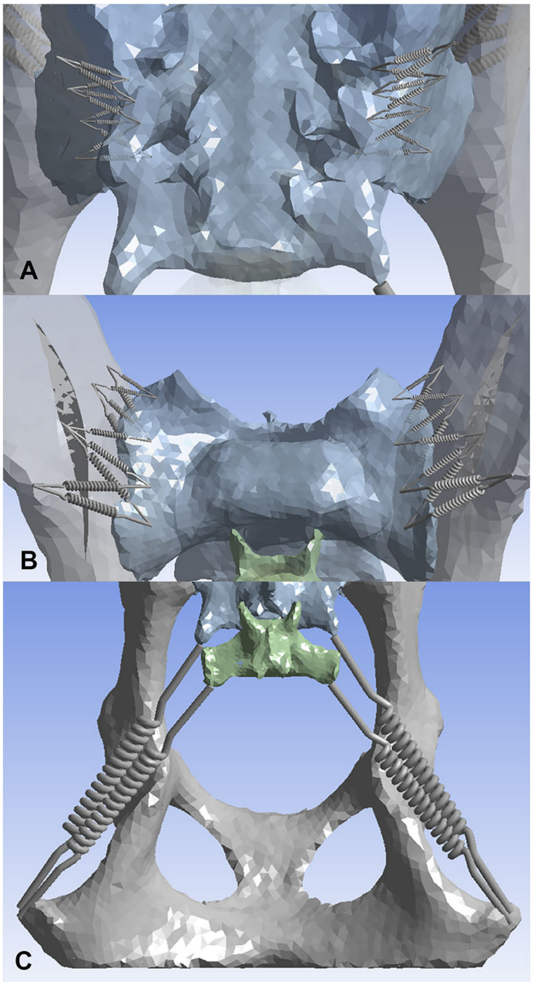


FIGURE 2 | Three-dimensional models displaying non-linear springs that were used to represent the sacroiliac joint ligaments. **(A)** dorsal sacroiliac ligaments, **(B)** ventral sacroiliac ligaments, **(C)** sacrotuberous ligament. Ten non-linear springs per side were used on both the dorsal sacroiliac ligament and the ventral sacroiliac ligament. Four points of interests were first picked on the sacrum and the ilium of both sides, followed by connecting the non-linear springs in a crisscross pattern. This pattern was used to mimic more of the ligaments sheet-like properties. For the sacrotuberous ligament, two non-linear springs per side were used. The ligament attachment sites were placed on the margins of the transverse processes of the third sacral and first caudal vertebrae, and the ischiatic tuberosities.

Ligaments in the pelvic region (dorsal sacroiliac ligament, ventral sacroiliac ligament, and sacrotuberous ligament) were modeled by using non-linear springs. Connection points of the ligaments on the bones were first identified based on a canine anatomy reference textbook (11). Then the non-linear springs were inserted to represent the ligaments. The non-linear tensile stiffness of each of the ligaments was defined as a piecewise linear strain-stiffness curve, using values listed in **Table 2**.

Ten non-linear springs were placed on each side (L, R) for the dorsal and ventral sacroiliac ligaments (**Figure 2**). To limit sliding motion and provide more of a sheet-like property, a crisscross pattern was developed. The 10 non-linear springs on each side for each ligament were all connected via a crisscross as shown in the figure. The sacrotuberous ligament was demonstrated using two non-linear springs and a "Y" configuration (**Figure 3**).

To define its non-linear tensile stiffness and zero compression stiffness, the sacroiliac joint was modeled using 10 non-linear springs per side. Each of the non-linear springs share the same properties as the non-linear springs that were used in the ligaments. Next, the joint between the sacrum and the first caudal vertebrae was modeled (sacrocaudal joint). The modulus of elasticity E was defined using the equation below

$$E = \begin{cases} 0 & for\ \varepsilon \geq 0 \\ 20.71\varepsilon + 234\varepsilon^2 & for\ \varepsilon < 0 \end{cases} \quad (1)$$

where ε is the strain (21). The above equation gives an increasing stiffness when a joint is compressed and a zero stiffness when it is stretched. This implies that the joint does not resist force separating the bones to which it is connected. Equation (1) can be rewritten in the form of stress-strain relation:

$$\sigma = \begin{cases} 0 & for\ \varepsilon \geq 0 \\ 10.355\varepsilon^2 + 78\varepsilon^3 & for\ \varepsilon < 0 \end{cases} \quad (2)$$

where σ is the normal stress with unit of MPa. In this work, the non-linear compressive stress-strain behavior of the joint is modeled as a hyperelastic material response. The 3-paramater Mooney-Rivlin model is employed and the model parameters are calculated by fitting the stress-strain curve defined by Equation (2). The solid model of the joint was constructed by manually creating a joint volume between the C1 vertebrae and the sacrum through a combination of drawing, extrusion, and Boolean operations.

Finite Element Analysis Settings and Loading Conditions

For each trial, the model was fixed in space with a fixed support applied to each side of the acetabulum. Eight different force loads (scenarios) were applied to the sacrum as shown in **Table 3**. The location where the forces were applied is illustrated in **Figure 4**.

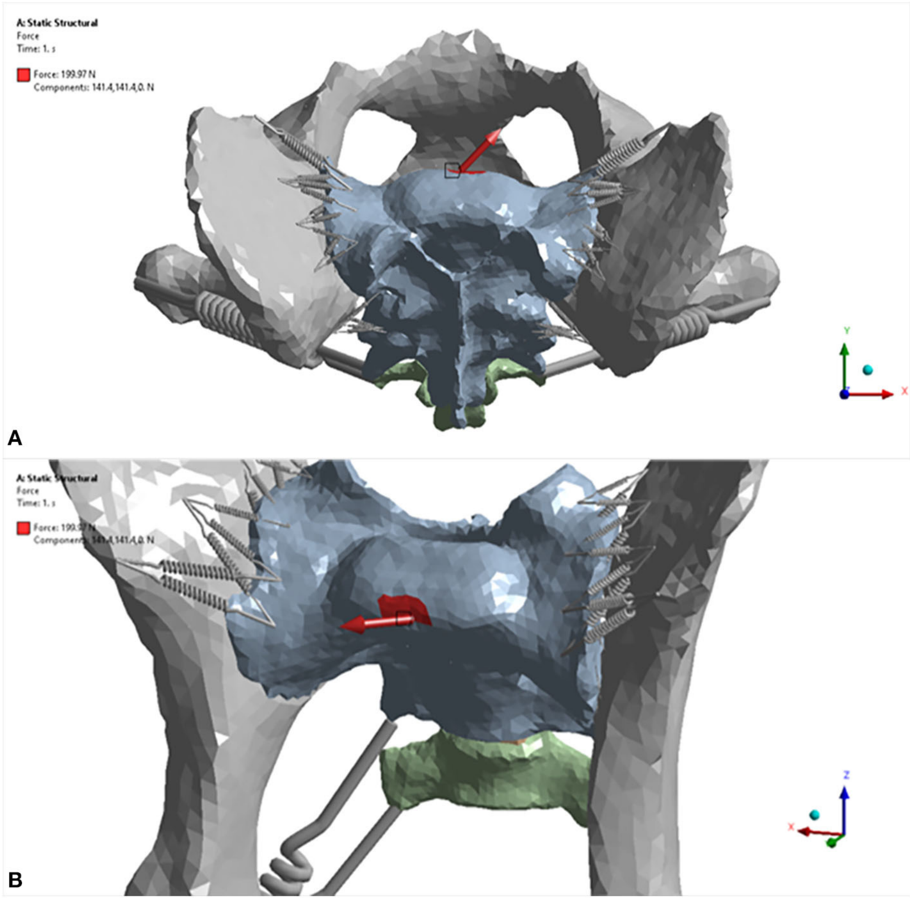


FIGURE 3 | Three-dimensional models displaying load scenario No. 5 for the dog pelvis (red arrow). **(A)** craniodorsal view; **(B)** cranioventral view. For this load scenario, 141.4 N was placed in the X direction and 141.4 N in the Y direction. The model was fixed in place at the level of the acetabulum.

Strain (elongation per unit length) in the ligaments caused by the loads was calculated and recorded for each ligament (non-linear spring) in each scenario. For each scenario, the strains of all non-linear springs in each ligament group on each side were averaged. The results obtained from the five trials were then compared and analyzed statistically. Deformation and equivalent stresses were also obtained from the analysis (**Figure 5**). The deformed position of the dog pelvis under the load scenario 7 is shown in **Figure 5B**. For comparison, the unreformed pelvis is also shown on the right. Notice that the displacement in the plot is exaggerated to demonstrate the deformation more clearly.

TABLE 3 | Eight loading scenarios that were applied to the sacrum in the model.

Scenario	Force in x-direction (*N*)	Force in y-direction (*N*)
1	0	200
2	0	−200
3	200	0
4	−200	0
5	141.4	141.4
6	−141.4	−141.4
7	141.4	−141.4
8	−141.4	141.4

Repeatability of Ligament Strain Measurements Using the FEA Method

Upon completion of the pathway from segmentation to model development to finite element analysis, five trials separated by exactly 1 week apart were conducted by a single observer (MC). Strain values of ligaments were recorded for each trial using different loading scenarios. The trials consisted of the steps discussed above and involved four software programs: a three-dimensional (3D) image analysis program (3D Slicer, version 4.5.0, http://www.slicer.org), mesh generation software (ICEM CFD 17.0, ANSYS, Canonsburg, PA), 3D computer-aided design program (SolidWorks 2016, Waltham, MA), and engineering simulation software (ANSYS Workbench 17.0, Canonsburg, PA). Software used was either free of charge or was provided under Clemson University's licensing. All trials were performed using one workstation (Lenovo ThinkPad S1 Yoga, Intel Core i5-4200U CPU 1.60 GHz, 8.00 GB, Hong Kong, China; Microsoft Windows 10, Redmond, WA).

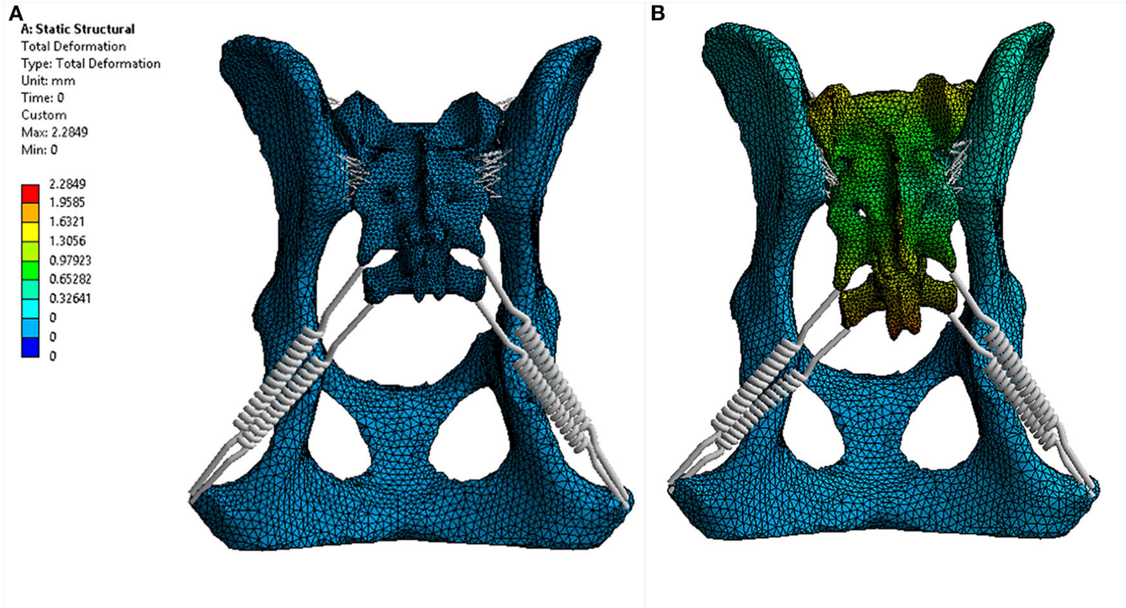


FIGURE 4 | Deformation of the proposed canine pelvis model due to a 200 N force applied to the sacrum (load scenario 7). **(A)** dorsal 3D view of the pelvis prior to applied force. **(B)** dorsal 3D view of the pelvis with the applied force (141.4 N in the X direction and −141.4 N in the Y direction).

FIGURE 5 | Transverse **(A)** and dorsal planar **(B)** CT images illustrating an intra-articular ankylosis lesion in the left sacroiliac joint (arrows). The left and right sacroiliac joints also appear asymmetrical in size and shape. The transverse images are displayed with dorsal at the top and the patient's left to the viewer's right. Dorsal planar images are displayed with cranial at the top and the patient's left to the viewer's right.

Finite element analysis was conducted on a different workstation (Apple MacBook Pro 17, 2.4 GHz quad-core Intel Core i7, 16.00 GB, Cupertino, California; Microsoft Windows 7, Redmond, WA).

Statistical analyses were selected and conducted in consultation with a statistician (JS). Strain values (y-axis) by side and ligament group, vs. trial (x-axis) were examined visually using commercially available statistics software (JMP, Version 14.0, SAS Institute Inc., Cary, NC, 1989-2007). A regression analysis was used to compare the strain value of each ligament (dependent variable) to determine if the values were similar throughout the five trials (independent variable). A total of 48 regression analyses were run (8 force load scenarios x 3 ligaments x 2 sides). Our null hypothesis was that the methods are repeatable (slope of the regression line is zero) while the alternative hypothesis was methods are not repeatable (slope of the regression line is not zero). Multiple comparison adjustments were not employed for this exploratory analysis.

RESULTS

Part 1: Comparisons Between Numbers of SIJ CT Lesions/Dog and Dog Work Status

Of the 22 included dogs, four were in training for future police or search and rescue work and 18 were military working dogs. Ten dogs were categorized as Detection, 6 dogs categorized as Breeder, and 6 dogs categorized as Other (**Table 4**). Thirteen dogs were classified into the younger age group (age 11–30 months) and nine dogs were classified into the older age group (age 31–48 months). Of the younger age group dogs, nine were intact males, and four were intact females. Within the younger age group, there were six Detection dogs (four males and two females), two Breeders, and five Others (all male). The older age group was composed of six females and three males. There was one neutered male and one spayed female and the rest of the dogs were intact. There were four Detection dogs (one intact male, one neutered male, one intact female, one spayed female), four Breeder dogs,

TABLE 4 | Descriptive summary of clinical and computed tomography findings for the 22 Labrador retriever working dogs included in the sample.

Work status category	Number of dogs	Age range (months)	Weight range (kg)	Sex and neuter status	Average # SIJ CT lesions/dog in each lesion category						
					SCy	SE	SS	PAA	IAA	SCl	IBS
Breeder	6	28–48	22.6–27.2	6 FI	0.42	1.58	1.17	0	1.33	0.5	0.5
Detection	10	11–41	26–34	2 FI 1 FS MI 1 MN	0.85	1.2	0.4	0	0.7	1.1	0.45
Other	6	14-32	19.1–29	1 FI 5 MI	1.58	1.75	0.67	0	0.58	1.17	1.08

kg, kilogram; CT, computed tomography; FI, female intact; FS, female spayed; MI, male intact; MN, male neutered; SCy, subchondral cyst; SE, subchondral erosion; SS, subchondral sclerosis; PAA, para-articular ankylosis; IAA, intra-articular ankylosis; SCl, subarticular cleft; IBS, intra-articular bone spur.

FIGURE 6 | Transverse **(A)** and dorsal planar **(B)** CT images illustrating a subarticular cleft lesion in the left sacroiliac joint (arrows). Subchondral erosion lesions are also evident in the right sacroiliac joint. The transverse images are displayed with dorsal at the top and the patient's left to the viewer's right. Dorsal planar images are displayed with cranial at the top and the patient's left to the viewer's right.

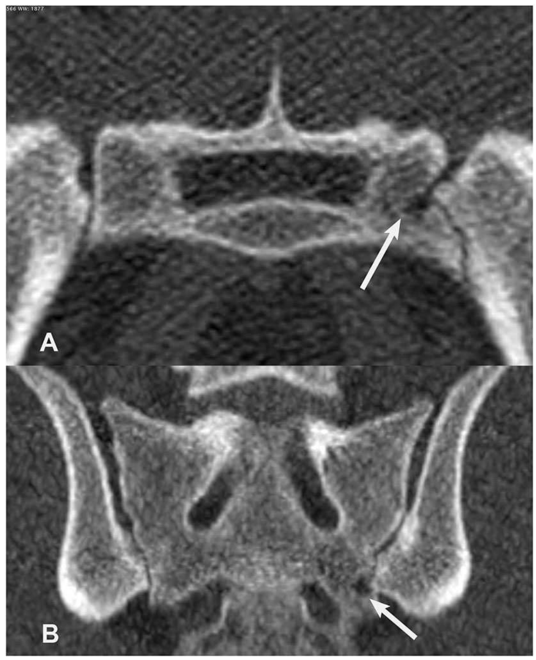


FIGURE 7 | Transverse **(A)** and dorsal planar **(B)** CT images illustrating a subchondral cyst lesion in the left sacroiliac joint (arrows). A subchondral sclerosis lesion is also evident in the dorsal planar view of the left sacroiliac joint. The transverse images are displayed with dorsal at the top and the patient's left to the viewer's right. Dorsal planar images are displayed with cranial at the top and the patient's left to the viewer's right.

and one Other dog (intact female). While neuter status may have implications for joint and ligament health, its effect could not be assessed in the present study due to the small number of neutered/spayed animals in this sample (33, 34). In addition to detecting SIJ CT lesions that have been previously described (16) (**Figures 6–9**), the veterinary radiologist detected a new type of CT SIJ lesion and termed it "intra-articular bone spur" (**Figure 10**). None of the dogs were found to have para-articular ankylosis lesions.

The ANOVA testing identified no significant difference in the mean number of SIJ CT lesions/dog across work status categories for subchondral erosions, subarticular clefts, intra-articular bone spurs, or intra-articular ankylosis lesions for the right SIJ, left SIJ, or in the total joint with both sides added together. Analysis of the total number of SIJ CT lesions found in each work status group yielded no significant difference in the means among the three categories.

No significant differences were found in the occurrences of subchondral sclerosis and subchondral cyst lesions among the work status groups. There was evidence that the mean number

FIGURE 8 | Transverse **(A)** and dorsal planar **(B)** CT images illustrating a subchondral sclerosis lesion in the left sacroiliac joint (arrows). A subchondral erosion lesion is also evident in the dorsal planar view of the right sacroiliac joint. The transverse images are displayed with dorsal at the top and the patient's left to the viewer's right. Dorsal planar images are displayed with cranial at the top and the patient's left to the viewer's right.

FIGURE 9 | Transverse **(A)** and dorsal planar **(B)** CT images illustrating a subchondral erosion lesion in the left sacroiliac joint (arrows). A subchondral sclerosis lesion is also evident surrounding the erosion. The transverse images are displayed with dorsal at the top and the patient's left to the viewer's right. Dorsal planar images are displayed with cranial at the top and the patient's left to the viewer's right.

of subchondral sclerosis lesions in the left SIJ differed among work status groups ($p = 0.0780$). Specifically, there appeared to be a greater number of subchondral sclerosis lesions identified in the left SIJ of Breeders than of Detection dogs ($p_{adj} = 0.0258$). There was not evidence of a difference between Other dogs and Breeders or Detection dogs in mean number of subchondral sclerosis lesions in the left SIJ. There was not evidence of a difference in the mean number of subchondral sclerosis lesions in the right SIJ among work status categories. The mean number of subchondral cysts in the right SIJ was different among the work status categories ($p = 0.0296$). The mean number of subchondral cysts in the right joints of Breeder dogs was significantly lower than for Other dogs ($p_{adj} = 0.0227/2$) and for Detection dogs ($p_{adj} = 0.0154/2$). There was not evidence of a difference in mean number of subchondral cysts between Other dogs and Detection dogs. These results were isolated to the right SIJ as no significant difference was seen among the work status categories for the mean number of subchondral cysts in the left side of the joint.

When introducing the variables of age and sex to the mean number of the subchondral sclerosis lesions seen in the left side of the SIJ across work status groups, female dogs in the "older" (31–48 months) group appear slightly different. Dogs in the Other work status group had a subjectively greater number of these lesions, followed by Breeders, while subjectively fewer lesions of this type were seen for Detection dogs.

When the same variables were introduced for the mean number of subchondral cysts in the right SIJ across work status groups, similar results were seen. Female dogs in age group 2 (i.e., older) working as Detection dogs and dogs categorized as Other showed a higher mean number of subchondral cysts in the right SIJ than Breeder females in the same age group.

Development of FEA Method for Modeling Canine Sacroiliac Joint Ligament Strain Using CT Data and Repeatability of Ligament Strain Measurements

Results of the 48 hypothesis tests for the selected dog are summarized in **Table 5**. We failed to reject the null hypothesis that the method is repeatable in approximately 98% of the tests at each combination of strain, ligament, and side (47 out of 48). We rejected the null hypothesis that the method is not repeatable in approximately 2% of the tests at each combination of strain, ligament, and side (1 out of 48). The hypothesis test that was not statistically repeatable was in force load (scenario) 8 for the left ventral sacroiliac ligament ($P = 0.0175$). Strain values under this scenario for this ligament ranged from −0.02343 (trial 1) to −0.01422 (trial 5).

DISCUSSION

The intentions of this two-part, preliminary study were to introduce two quantitative CT methods; with the long term

FIGURE 10 | Transverse **(A)** and dorsal planar **(B)** CT images illustrating an intra-articular bone spur lesion in the right sacroiliac joint (arrows). The left and right sacroiliac joints are asymmetrical in shape and size. Subchondral erosion lesions are visible in both joints in the transverse view. The transverse images are displayed with dorsal at the top and the patient's left to the viewer's right. Dorsal planar images are displayed with cranial at the top and the patient's left to the viewer's right.

goal of supporting future research studies characterizing effects of working tasks in SIJ. We explored application of these two methods in a small sample of young, Labrador retriever working dogs. We described a preliminary assessment of the presence of SIJ CT lesions in young working dog groups. We also introduced a new SIJ lesion, termed "intra-articular bone spur." Although no significant differences were found for the mean number of SIJ lesions/dog, significant differences in subchondral sclerosis and subchondral cyst lesions were observed among work status groups when comparing individual sides of the joints.

Part 1: Comparisons Between Numbers of SIJ CT Lesions/Dog and Dog Work Status

Findings from the first part of the study did not support the hypothesis that mean numbers of SIJ CT lesions/dog would differ among work status groups. However, when SIJ side (right or left) was examined, significant differences among work status groups were found for subchondral sclerosis and subchondral cyst lesions. Breeders were found to have a higher number of subchondral sclerosis lesions in the left SIJ when compared to that of Detection dogs while both Detection and Other dogs had higher numbers of subchondral cysts found in the right SIJ when each was compared to Breeder dogs. Age and sex also appeared to play a role in the number of lesions found in the SIJ with older (31–48 months) females having a greater number of both subchondral sclerosis and subchondral cyst lesions identified across all three categories when comparing the individual sides of the joints. These preliminary results suggest that multiple factors such as work status, age, and sex may contribute to the development of SIJ lesions in working dogs.

A previously unreported SIJ CT lesion was observed in some of the dogs and authors introduced the term "intra-articular bone spur." These bone spurs were considered to most likely be enthesiophytes that formed at the interosseous ligament attachment sites. Though not histopathologically confirmed in the dogs of the current study, there is histologic evidence of a transitional zone of fibrocartilage extending into the ligamentous portion of the sacroiliac joint (12). This zone can develop degenerative changes in dogs as young as 5 months of age, with histologic evidence of cartilage matrix splitting and chondrocyte proliferation at ligamentous attachment sites. In people, when placed under unusual tensile strain, ligamentous entheses can undergo degenerative changes that make them more likely to tear outright, pulling away a portion of the underlying subchondral bone resulting in enthesiophytes (35, 36). The presence of these intra-articular enthesiophytes in young Labrador retriever working dogs could be an indication of cumulative mechanical energy (or overuse injury); or indicators that intense, repetitive training techniques employed in puppies employed in puppies in preparation for various lines of duty could possibly be detrimentally placing strain on the joint and its ligaments (6).

In this study, Breeders were found to have a higher mean number of subchondral sclerosis lesions in the left SIJ when compared to that of Detection dogs. This was an unexpected finding because unspayed, female dogs are not typically trained or used for working tasks. Possible theories could be that dogs voluntarily performed repetitive upright postures in their kennels or that whelping and hormonal influences could have altered SIJ ligament rigidity. In people, changes in the pubic symphysis of women caused by degeneration or vertical displacement was positively correlated with SIJ pain, despite being not significantly associated with parous status (37). Estrogens and relaxins released in dogs during pregnancy have been shown influence the tensile properties of connective tissues by increasing the flexibility of both the pubic symphysis as well as the SIJ (38). The aforementioned puts the hip joints at a greater risk for instability and in turn affect the stability of the SIJ. Hormonally influenced ligament laxity combined with structural changes in the pubic bone and SIJ resulting from pregnancy, repeated estrus cycles in whelping intact females may contribute to SIJ instability. Based on Wolff's law, this could in turn predispose the subchondral bone to microfractures and subsequent sclerosis (39). Parity may also play a role in the degree of change within the SIJ, which may in turn influence the number of lesions seen, but further studies would be necessary to determine whether the number of litters a female carries is related to increased joint laxity when compared to females that have had fewer litters.

In the current study and in a previous canine cadaver study, subchondral cyst CT lesions appeared as discrete oval radiolucencies with surrounding sclerotic rims in the

TABLE 5 | Regression analysis results [slope estimate (95% confidence interval limits); *p*-value] for sacroiliac ligament strain values (dependent variable) vs. trial (independent variable) for each ligament (Dorsal Sacroiliac, Sacrotuberous, and Ventral Sacroiliac) on each side (L, R) for each of 8 scenarios.

Ligament	Scenario number	Slope estimate (95% confidence interval limits); *p*-value for SIJ ligament strain values (dependent variable)	
		Right side	Left side
Dorsal sacroiliac	1	−0.001 (−0.004, 0.003); 0.6841	0 (−0.002, 0.001); 0.8539
	2	−0.002 (−0.007, 0.003); 0.3029	0.006 (0, 0.011); 0.058
	3	−0.002 (−0.007, 0.004); 0.4442	0.006 (0, 0.011); 0.0511
	4	−0.001 (−0.009, 0.007); 0.6946	0.001 (−0.001, 0.003); 0.3748
	5	0 (−0.006, 0.005); 0.809	−0.002 (−0.004, 0.001); 0.1356
	6	−0.002 (−0.009, 0.005); 0.4267	0.002 (−0.002, 0.005); 0.2197
	7	−0.002 (−0.006, 0.003); 0.3843	0.006 (−0.001, 0.014); 0.0689
	8	0.001 (−0.002, 0.004); 0.2746	0 (−0.001, 0.001); 0.7566
Sacrotuberous	1	0.001 (−0.001, 0.002); 0.1631	0.001 (−0.001, 0.003); 0.2797
	2	0.001 (−0.004, 0.005); 0.572	0.001 (−0.003, 0.005); 0.6195
	3	−0.001 (−0.003, 0.002); 0.5198	0 (−0.003, 0.002); 0.5664
	4	0.002 (−0.001, 0.005); 0.1488	0.002 (−0.001, 0.005); 0.1347
	5	0 (−0.004, 0.005); 0.8486	0 (−0.004, 0.005); 0.7552
	6	0.002 (−0.003, 0.008); 0.2248	0.002 (−0.002, 0.007); 0.1914
	7	−0.001 (−0.004, 0.002); 0.4739	−0.001 (−0.004, 0.002); 0.4105
	8	0.001 (0, 0.001); 0.138	0.001 (−0.001, 0.002); 0.2084
Ventral sacroiliac	1	−0.001 (−0.005, 0.003); 0.5834	0 (−0.003, 0.003); 0.8687
	2	0 (−0.007, 0.007); 0.9036	0.004 (−0.01, 0.018); 0.396
	3	−0.002 (−0.008, 0.003); 0.2856	0.004 (−0.006, 0.015); 0.2922
	4	0.002 (−0.008, 0.011); 0.602	−0.002 (−0.007, 0.004); 0.4572
	5	−0.001 (−0.005, 0.003); 0.3987	−0.003 (−0.012, 0.007); 0.4644
	6	0.002 (−0.007, 0.011); 0.5244	0 (−0.009, 0.008); 0.8955
	7	−0.002 (−0.008, 0.004); 0.2908	0.004 (−0.004, 0.013); 0.1929
	8	0.002 (0.001, 0.004); ***0.0175**	−0.001 (−0.002, 0.001); 0.286

**and bold font indicates statistically significant difference; "Scenario" was defined as the different loading conditions (different forces placed on different axes). "Trial" was the entire process (segmentation, model creation, analysis including all 8 scenarios) each separated by a week.*

subchondral bone of the articular components of the SIJ, and subchondral sclerosis lesions appeared as focal areas of increased subchondral bone radioopacity (16). A previous canine study described histologic evidence of articular cartilage damage (including splitting of the cartilage matrix, proliferation of chondrocytes, and decreased glycosaminoglycan production in the cartilage matrix) in the synovial component of SIJ in dogs as early as 5 months of age (12). A "synovial lined recess" was identified in a 3-year-old dog. In the present study, there was a pattern for increased numbers of subchondral sclerosis and subchondral cyst lesions/dog in female dogs of the "older" (31–48 months) category. While not statistically significant, these differences could be preliminary evidence used to generate hypotheses in other studies that age and female sex could also be risk factors for injury or progression of microtraumas in the SIJ. All three work status categories were represented in this trend, with some variation in the specific frequencies amongst each category. There was significant lateralization in the mean numbers of subchondral sclerosis and subchondral cyst lesions/dog, with significantly more lesions/dog identified in one side of the joint but not the other. Breeder dogs were found to have a significantly higher occurrence of subchondral sclerosis lesions when the left side of the SIJ was compared to the left SIJ of detection dogs but no significant differences were found when comparing the number of lesions in the right SIJ or the mean number of lesions in the total joint of both groups. Additionally both Other and Detection dogs had a significantly higher number of subchondral cysts when comparing the right SIJ between Other vs. Breeder and Detection vs. Breeder but similarly, no significant differences were found between the pairings when comparing the left SIJ or the mean number of lesions in the total joint. Though possibly limited due to a lack of power, one theory for these observed side differences could be that dog handlers have a side preference when asking dogs to perform repetitive training and working tasks.

Part 2: Development of the FEA Method for Modeling the Sacroiliac Joint and Ligaments Using CT Data and Repeatability of Ligament Strain Measurements

We developed a methodology for constructing a computer-based model of the canine pelvis using CT images and four computer software applications; one of which was freely available, while the other three required commercial licenses. Our methodology was adapted from a previously published, CT-derived, human pelvis FEA model (21). In-depth details on the process of creating the model were not available in the previous publication and no repeatability testing was described. To evaluate intra-observer repeatability in the current study, we determined that quantifiable parameters were needed for each trial and case number. The decision to average all of the strain values for each of the non-linear springs to obtain an average SIJ ligament strain value was developed for this reason. Intra-observer repeatability of this methodology was then tested by using regression analyses on the SIJ ligament strain values of five models separated by a week apart and performing FEA on each of the models. All models were given the same bone and ligament properties as those used in the previous human study and eight different loading conditions were placed on each model. Findings from the current study indicated that our procedure is repeatable (47 out of 48 load scenarios). The one variable that was not repeatable between trials was in the left ventral sacroiliac ligament under scenario 8 (−141.4 N placed on the sacrum base in the X direction and 141.4 N in the Y direction). It is possible that this non-repeatable variable occurred because of operator-based variability and non-physiological loading conditions.

After we had initiated the current study, a thesis was published that also described development of a biomechanical model using CT scans of a canine pelvis (40). Similar to ours, the study modeled bone and ligament properties for the canine pelvis based on the human literature. The study also included assumptions based on previously published tissue properties of canine long bones (41). Ligament placements were based on a veterinary textbook (11). Muscles and their properties were incorporated into the model using a cadaveric model and a separate biomechanical test for model validation. However, strain values for the sacroiliac ligaments and repeatability testing were not performed. In another study, a CT-derived canine FEA determined the presence of micro-motion following the repair of sacroiliac separation by way of different surgical fixations (26). These authors used a simplified model of the canine pelvis and, similar to the current study, modeled the bones as isotropic and elastic. Authors noted that the loading conditions they used for magnitude and directionality in their model were not physiologically realistic to the dog. The current study shared a similar trait to the previous study in that a large value in magnitude was used to capture the micromovements during FEA.

We considered repeatability testing to be important to assess the FEA methodology because operator decision-making occurs during multiple steps of the process. We chose to perform each of the trials at exactly 1 week apart in order to minimize effects of retained operator muscle memory on repeatability results. The segmentation portion of the process had the largest potential for operator-based variability due to the large number of CT slices and hand-traced bone regions of interest. Segmentation was also the most time-consuming stage of the procedure. The segmentation was performed using a freeware program (3D Slicer), and some automated thresholding was possible. However, this was otherwise mostly a time-consuming manual process. The use of newer software products that allow more automated thresholding could possibly reduce the time required and increase the accuracy of the method in the future. Non-linear spring placement to provide ligaments and joints in the model was also another source of variability. Not only the length and location of these springs, but also the axial direction, were possible sources of variability for the method. The sacroiliac joint has very complex anatomy, therefore finding the same location for placement of the springs on each trial was a particular challenge. Another possible source of variability was the exact positioning of the boundary conditions applied to the model. This included the location where the load was placed and where the model was held in space.

Limitations and Conclusions

Several limitations need to be acknowledged for this preliminary, exploratory study. The sample sizes were small and these reduced the power of comparisons for work status, age, sex, and neuter status groups. The study sampled only one working dog breed, and therefore generalizability of findings for working dogs of other breeds remains unknown. The working tasks for dogs at these particular training centers may not be the same as those used at other training centers. The calculated ligament strain values for the single dog selected for Part 2 of the study were not validated based on actual bone and ligament properties in Labrador retrievers and were instead based on tissue properties published for humans. A previous review article described multiple limitations for using human tissue properties in canine musculoskeletal modeling (42). Muscles of the pelvis were not included in our model because we had intended to focus on adapting our methods to those described in a previous paper that was focused on FEA modeling of human SIJ ligament mechanical properties (21). A single dog was used for the 2nd part of the study and therefore effects of size, sex, breed, and age on ligament strain values were not tested. A single operator performed all trials and therefore inter-operator reliability of the technique was not tested. Histopathologic confirmations of the SIJ lesions and interosseous ligaments was not performed. This would have been unethethical in this otherwise healthy population of dogs. Finally, the FEA model was not validated with a cadaveric study.

In conclusion, this preliminary study introduced two quantitative CT measures for possible use in future research studies on effects of training and working tasks on SIJ in young working dogs. The first methodology was the number of SIJ CT lesions/dog and application of this method was explored using work status group comparisons in a small sample of young, Labrador retriever working dogs. The second methodology was SIJ ligament strain modeling based on patient-specific CT data and FEA. Intra-observer repeatability of SIJ ligament strain measurements using these models was tested in one dog and found to be very good. Further research is needed to examine specific training and working tasks required of young working dogs across multiple lines of duty and compare these using quantitative SIJ CT measures.

Future studies are also needed to improve the SIJ ligament strain FEA model. A validation study should be performed using actual bone, joint, and muscle properties from dogs of representative breeds. The long-term goals for these research efforts would be to develop more evidence-based strategies for minimizing early retirement and maximizing quality of life for working dogs.

DATA AVAILABILITY STATEMENT

The datasets for this article are not publicly available because of hospital patient data confidentiality requirements. Requests to access military working dog datasets should be directed to Army Public Health Center ATTN: MCHB-IP-V 8252 Blackhawk Rd. Aberdeen Proving Ground, MD 21010-5403; email: usarmy.apg.medcom-aphc.mbx.iph-vet@mail.mil.

ETHICS STATEMENT

The original, prospective military working dog study was reviewed and approved by LTC Daniel E. Holland Memorial Military Working Dog Hospital, Lackland Joint Base, TX; Institutional Animal Care and Use Committee # 2012–06; approval Aug. 12, 2012.

AUTHOR CONTRIBUTIONS

MC, JJ, GL, and JS: conception and design. MC, JJ, KO, and GL: acquisition of data. MC, JJ, GL, JS, KO, and WB: analysis and interpretation of data, revising article for intellectual content, and final approval of the completed article. MC, JJ, and KO: drafting the article. All authors contributed to the article and approved the submitted version.

FUNDING

Funding was provided by the South Carolina Translational Research Improving Musculoskeletal Health (SC-TRIMH) Center at Clemson University (NIH P20GM121342) and the Clemson Creative Inquiry Fund.

ACKNOWLEDGMENTS

The authors acknowledge the veterinarians and staff at the Ryan Veterinary Hospital and the LTC Daniel E. Holland Memorial Military Working Dog Hospital for their assistance in acquiring data used for the current study. Authors would also like to thank Mr. Josh Tan at the Wake Forest School of Medicine Clinical and Translational Science Institute for assistance with image analysis methodology development.

REFERENCES

1. Worth AJ, Sandford M, Gibson B, Stratton R, Erceg V, Bridges J, et al. Causes of loss or retirement from active duty for New Zealand police German shepherd dogs. *Animal Welfare.* (2013) 22:167–74. doi: 10.7120/09627286.22.2.167
2. Moody JA, Clark LA, Murphy KE. Working dogs: history and applications. In: Ostrander EA, Giger U, Lindblad-Toh K, editors. *The Dog and its Genome*. Vol. 44. Woodbury, NY: Cold Spring Harbor Laboratory Press (2006). p. 1–18.
3. Takara MS, Harrell K. Noncombat-related injuries or illnesses incurred by military working dogs in a combat zone. *JAVMA.* (2014) 245:1124–8. doi: 10.2460/javma.245.10.1124
4. Mey W, Schuh-Renner A, Anderson MK, Stevenson-LaMartina H, Grier T. Risk factors for injury among military working dogs deployed to Iraq. *Prev Vet Med.* (2020) 176:6. doi: 10.1016/j.prevetmed.2020.104911
5. Knaus I, Breit S, Kunzel W, Mayrhofer E. Appearance and incidence of sacroiliac joint disease in ventrodorsal radiographs of the canine pelvis. *Veterinary Radiol Ultrasound.* (2004) 45:1–9. doi: 10.1111/j.1740-8261.2004.04001.x
6. Zink MC, VanDyke JB, (editors). Conditioning and retraining the canine athlete. In: *Canine Sports Medicine and Rehabilitation.* Hoboken, NJ: John Wiley & Sons (2013). p. 176–200. doi: 10.1002/9781118783443.ch10
7. Marymont JV, Lynch MA, Henning CE. Exercise-related stress reaction of the sacroiliac joint - an unusual cause of low-back-pain in athletes. *Am J Sports Med.* (1986) 14:320–3. doi: 10.1177/036354658601400414
8. Harvey J, Tanner S. Low-back-pain in young athletes - a practical approach. *Sports Med.* (1991) 12:394–406. doi: 10.2165/00007256-199112060-00005
9. Connolly LP, Drubach LA, Connolly SA, Treves ST. Young athletes with low back pain: skeletal scintigraphy of conditions other than pars interarticularis stress. *Clin Nuclear Med.* (2004) 29:689–93. doi: 10.1097/00003072-200411000-00003
10. Carfagno DG, Hendrix JC III. Overtraining syndrome in the athlete: current clinical practice. *Curr Sports Med Rep.* (2014) 13:45–51. doi: 10.1249/JSR.0000000000000027
11. Evans HE, de Lahunta A. *Miller's Anatomy of the Dog
*. 4 ed. St. Louis, MO: Elsevier Saunders (2013). p. 113–23.
12. Gregory CR, Cullen JM, Pool R, Vasseur PB. The canine sacroiliac joint. preliminary study of anatomy, histopathology, and biomechanics. *Spine.* (1986) 11:1044–8. doi: 10.1097/00007632-198612000-00019
13. Edge-Hughes L. Hip and sacroiliac disease: selected disorders and their management with physical therapy. *Clin Techniq Small Animal Practice.* (2007) 22:183–94. doi: 10.1053/j.ctsap.2007.09.007
14. Mustafy T, Londono I, Moldovan F, Villemure I. High impact exercise improves bone microstructure and strength in growing rats. *Sci Rep.* (2019) 9:13128. doi: 10.1038/s41598-019-49432-2
15. Bourrin S, Genty C, Palle S, Gharib C, Alexandre C. Adverse effects of strenuous exercise: a densitometric and histomorphometric study in the rat. *J Appl Physiol.* (1994) 76:1999–2005. doi: 10.1152/jappl.1994.76.5.1999
16. Carnevale M, Jones J, Holaskova I, Sponenberg DP. CT and gross pathology are comparable methods for detecting some degenerative sacroiliac joint lesions in dogs. *Veterinary Radiol Ultrasound.* (2019) 60:378–89. doi: 10.1111/vru.12749
17. Breit S, Kunzel W. On biomechanical properties of the sacroiliac joint in purebred dogs. *Ann Anatomy.* (2001) 183:145–50. doi: 10.1016/S0940-9602(01)80036-4
18. Dalstra M, Huiskes R, Vanerning L. Development and validation of a 3-dimensional finite-element model of the pelvic bone. *J Biomech Eng-Trans Asme.* (1995) 117:272–8. doi: 10.1115/1.2794181
19. Oonishi H, Isha H, Hasegawa T. Mechanical analysis of the human pelvis and its application to the artificial hip joint–by means of the three dimensional finite element method. *J Biomech.* (1983) 16:427–44. doi: 10.1016/0021-9290(83)90075-1
20. Phillips ATM, Pankaj P, Howie CR, Usmani AS, Simpson A. Finite element modelling of the pelvis: inclusion of muscular and ligamentous boundary conditions. *Med Eng Phys.* (2007) 29:739–48. doi: 10.1016/j.medengphy.2006.08.010
21. Eichenseer PH, Sybert DR, Cotton JR. A finite element analysis of sacroiliac joint ligaments in response to different loading conditions. *Spine.* (2011) 36:E1446–52. doi: 10.1097/BRS.0b013e31820bc705

22. Hammer N, Steinke H, Lingslebe U, Bechmann I, Josten C, Slowik V, et al. Ligamentous influence in pelvic load distribution. *Spine J.* (2013) 13:1321–30. doi: 10.1016/j.spinee.2013.03.050
23. Kim YH, Kim JE, Eberhardt AW. A new cortical thickness mapping method with application to an *in vivo* finite element model. *Comput Methods Biomech Biomed Eng.* (2014) 17:997–1001. doi: 10.1080/10255842.2012.731594
24. Shi DF, Wang F, Wang DM, Li XQ, Wang QG. 3-D finite element analysis of the influence of synovial condition in sacroiliac joint on the load transmission in human pelvic system. *Med Eng Phys*. (2014) 36:745–53. doi: 10.1016/j.medengphy.2014.01.002
25. Lim TH, Goel VK, Weinstein JN, Kong WZ. Stress-analysis of a canine spinal motion segment using the finite-element technique. *J Biomech.* (1994) 27:1259–69. doi: 10.1016/0021-9290(94)90279-8
26. McCartney WT, Comiskey D, MacDonald B. Use of transilial pinning for the treatment of sacroiliac separation in 25 dogs and finite element analysis of repair methods. *Vet Comp Orthop Traumatol.* (2007) 20:38–42. doi: 10.1055/s-0037-1616586
27. Shahar R, Banks-Sills L, Eliasy R. Stress and strain distribution in the intact canine femur: finite element analysis. *Med Eng Phys.* (2003) 25:387–95. doi: 10.1016/S1350-4533(03)00002-X
28. van Rietbergen B, Weinans H, Huiskes R, Odgaard A. A new method to determine trabecular bone elastic properties and loading using micromechanical finite-element models. *J Biomech.* (1995) 28:69–81. doi: 10.1016/0021-9290(95)80008-5
29. Mukherjee M, Jones JC, Holaskov I, Raylman R, Meade J. Phenotyping of lumbosacral stenosis in labrador retrievers using computed tomography. *Vet Radiol Ultrasound.* (2017) 58:565–80. doi: 10.1111/vru.12520
30. Mukherjee M, Jones JC, Yao J. Lumbosacral stenosis in Labrador retriever military working dogs - an exomic exploratory study. *Canine Genet Epidemiol.* (2017) 4:12. doi: 10.1186/s40575-017-0052-6
31. Schwarz T, Saunders J. *Veterinary Computed Tomography*. Hoboken, NJ: John Wiley & Sons (2011). doi: 10.1002/9781118785676
32. Carnevale M. *An Exploratory Repeatability Study Quantifying the Strain in Sacroiliac Joint Ligaments in the Static Canine Pelvis*. Clemson, SC: Clemson University (2020).
33. Anderson KL, Zulch H, Neill DG, Meeson RL, Collins LM. Risk factors for canine osteoarthritis and its predisposing arthropathies: a systematic review. *Front Vet Sci*. (2020) 7:220. doi: 10.3389/fvets.2020.00220
34. Slauterbeck J, Pankratz K, Xu K, Bozeman S, Hardy D. Canine ovariohysterectomy and orchiectomy increases the prevalence of ACL injury. *Clin Orthop Relat Res.* (2005) 2004:301–5. doi: 10.1097/01.blo.0000146469.08655.e2
35. Dimitrakopoulou A, Schilders E, Bismil Q, Talbot JC, Kazakos K. An unusual case of enthesophyte formation following an adductor longus rupture in a high-level athlete. *Knee Surg Sports Traumatol Arthrosc.* (2010) 18:691–3. doi: 10.1007/s00167-009-1029-0
36. Maertens M, Mielants H, Verstraete K, Veys EM. Evaluation of the involvement of axial entheses and sacroiliac joints in relation to diagnosis: comparison among diffuse idiopathic skeletal hyperostostis (DISH), osteoarthrosis and ankylosing spondylitis. *Clin Rheumatol.* (1992) 11:551–7. doi: 10.1007/BF02283118
37. Becker S, Capobianco R, Seita M. Is sacroiliac joint pain associated with changes in the pubic symphysis? A radiographic pilot study. *Eur J Orthop Surg Traumatol.* (2015) 25(Suppl. 1):S243–9. doi: 10.1007/s00590-014-1575-0
38. Steinetz BG, Williams AJ, Lust G, Schwabe C, Bullesbach EE, Goldsmith LT. Transmission of relaxin and estrogens to suckling canine pups via milk and possible association with hip joint laxity. *Am J Vet Res.* (2008) 69:59–67. doi: 10.2460/ajvr.69.1.59
39. Stewart HL, Kawcak CE. The importance of subchondral bone in the pathophysiology of osteoarthritis. *Front Vet Sci.* (2018) 5:178. doi: 10.3389/fvets.2018.00178
40. Reed SA. *A Preliminary Musculoskeletal Model of the German Shepherd Lumbosacral Spine : A Thesis Presented in Partial Fulfilment of the Requirements for the Degree of Master of Science in Animal Science at Massey University, Manawatu, New Zealand* [Masters]: Massey University, New Zealand (2017).
41. Pressel T, Bouguecha A, Vogt U, Meyer-Lindenberg A, Behrens BA, Nolte I, et al. Mechanical properties of femoral trabecular bone in dogs. *Biomed Eng Online.* (2005) 4:17. doi: 10.1186/1475-925X-4-17
42. Dries B, Jonkers I, Dingemanse W, Vanwanseele B, Vander Sloten J, van Bree H, et al. Musculoskeletal modelling in dogs: challenges and future perspectives. *Vet Compar Orthop Traumatol.* (2016) 29:181–7. doi: 10.3415/VCOT-15-08-0133

Disclaimer: The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Army, Department of Defense, nor the United States Government.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

BRIEF RESEARCH REPORT
published: 18 August 2020
doi: 10.3389/fvets.2020.00525

A Lesson in Standardization – Subtle Aspects of the Processing of Samples Can Greatly Affect Dogs' Learning

*Claire M. Guest[1], Rob Harris[1], Iqbal Anjum[2], Astrid R. Concha[3] and Nicola J. Rooney[1,4]**

[1] Medical Detection Dogs, Greenway Business Park, Milton Keynes, United Kingdom, [2] US Army Research Office - Texas Tech University, Lackland, TX, United States, [3] Milton Keynes University Hospital, Milton Keynes, United Kingdom, [4] Animal Welfare and Behaviour Group, Bristol Veterinary School, University of Bristol, Bristol, United Kingdom

Training new medical odors presents challenges in procuring sufficient target samples, and suitably matched controls. Organizations are often forced to choose between using fewer samples and risking dogs learning individuals or using differently sourced samples. Even when aiming to standardize all aspects of collection, processing, storage and presentation, this risks there being subtle differences which dogs use to discriminate, leading to artificially high performance, not replicable when novel samples are presented. We describe lessons learnt during early training of dogs to detect prostate cancer from urine. Initially, six dogs were trained to discriminate between hospital-sourced target and externally-sourced controls believed to be processed and stored the same way. Dogs performed well: mean sensitivity 93.5% (92.2–94.5) and specificity 87.9% (78.2–91.9). When training progressed to include hospital-sourced controls, dogs greatly decreased in specificity 67.3% (43.2–83.3). Alerted to a potential issue, we carried out a methodical, investigation. We presented new strategically chosen samples to the dogs and conducted a logistic regression analysis to ascertain which factor most affected specificity. We discovered the two sets of samples varied in a critical aspect, hospital-processed samples were tested by dipping the urinalysis stick into the sample, whilst for externally sourced samples a small amount of urine was poured onto the stick. Dogs had learnt to distinguish target aided by the odor of this stick. This highlights the importance of considering every aspect of sample processing even when using urine, often believed to be less susceptible to contamination than media like breath.

Keywords: dog, training, confounder, odor, standardization, processing, prostate cancer

OPEN ACCESS

Edited by:
Wendy Irene Baltzer, Massey University, New Zealand

Reviewed by:
Jennifer Lynn Essler, University of Pennsylvania, United States
Federica Pirrone, University of Milan, Italy

****Correspondence:***
Nicola J. Rooney
nicola.rooney@bristol.ac.uk

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences, a section of the journal Frontiers in Veterinary Science

Received: *16 April 2020*
Accepted: *08 July 2020*
Published: *18 August 2020*

Citation:
Guest CM, Harris R, Anjum I, Concha AR and Rooney NJ (2020) A Lesson in Standardization – Subtle Aspects of the Processing of Samples Can Greatly Affect Dogs' Learning. Front. Vet. Sci. 7:525. doi: 10.3389/fvets.2020.00525

INTRODUCTION

When starting to train dogs to detect new emerging target odors, organizations are often faced with a challenge of procuring adequate numbers of both target samples, and suitably matched controls (1, 2). As a consequence, they are often forced to choose between using small numbers of training samples and risking the dog's learning to identify individual samples (rather than the target vs. control distinction) or sourcing samples from multiple places. The latter choice, whilst aimed at increasing the possibility of dogs' learning to discriminate the target odor, risks the possibility that, if samples come from limited sources, and are not processed identically, dogs may learn to distinguish target from control, based on a confounding factor. To mitigate this, organizations training dogs, aim to standardize all aspects of the collection, processing and sample presentation, however even with the greatest care, there may remain subtle factors which differ and which dogs can potentially learn to use to discriminate during training. This can lead to artificially inflated performance rates during training, which are not replicable when novel samples from another source are presented for training or blind testing.

There is a real challenge when training dogs to learn complex odors in complex environments, especially when those training them are not aware what the odor signature is, as is the case for many medical detection tasks [e.g., (1)]. What's more, there is variation in both targets and controls, and the aim of training is to ensure that the animal's identification responses are being controlled by disease-related Volatile Organic Compounds (VOCs) rather than other volatiles unique to the individual who provided the sample (2). It is important to ensure that training conditions are conducive to this "concept formation."

Studies of explosives have shown that increasing the variation in training samples of TNT (3) and gunpowder (4) improved generalization by the dogs and hence increased the likelihood of "concept formation." Odor profiles associated with disease are complex and are presented within numerous background odors. Gas chromatography-mass spectrometry studies of urine collected from prostate patients, identified over 500 potentially relevant VOCs amongst a total of over 9,000 (5). As Edwards et al. (2) advise, ideally a wide variety of positive samples with a single commonality: positive disease status and likewise, a wide variety of negative samples with a single commonality – negative disease status – is needed. However, Edwards et al. (2) also point out, one of the largest challenges in olfactory detection of human disease is sample availability.

Since dogs can identify individual human odors [e.g., (6)] and retain individuals in their memory (1), if one starts to train on a small number of samples, they run the risk of dogs simply learning to recognize whether each person's samples are rewarded or not. Hence, training organizations may be forced to seek larger numbers of samples and controls that may not be sourced identically for example using more than one hospital or other sources. It is commonly acknowledged that sourcing human target and control samples from non-matched sources presents the risk that dogs learn to discriminate based on a confounder or cross contamination (7). Whilst such issues are widely acknowledged when using breath as a medium (7), the risk with more stable media such as urine is less widely known, and therefore believed to be less of a risk.

This was the situation when Medical Detection Dogs (MDD) the UK's leading medical detection dog charity, first started to train dogs to detect prostate cancer from urine. Proof of principle studies had suggested that dogs can be trained to detect prostate cancer from urine (8, 9). But when MDD started to train for this, they were faced with the challenge of having small initial numbers of samples, especially controls, supplied by a single source. Below we describe the training that was carried out, the issue that emerged and our logical and systematic effort to identify and overcome this.

METHODS

The study received Ethical Approval from North West - Lancaster Research Ethics Committee (Ref:15/NW/0527).

Sample Collection

Samples were collected from Milton Keynes University Hospital (MKUH), both positive samples and age- and symptom-matched controls from men attending urological outpatients' clinics and to supplement control sample numbers, men and women from external Medical Detection Dog (MDD) events. All participants were over 18 years, had no previous history of malignancy (urological or non-urological), were not undergoing dialysis, nor had a diagnosis of HIV or Hepatitis (except Hepatitis A).

All urine samples were believed to be collected and processed in the same way. Participants were provided with a collection pot and plastic gloves and asked to urinate directly into the pot. Samples were then handled by the experimenter or nurse who tested for urine composition (presence of UTI, diabetes, and kidney disorders) using a urinalysis stick (Siemens Multistix 10SG), labeled the sample and placed it in a portable freezer, before being frozen in the hospital's freezer or, in the case of external samples, MDDs freezer. Samples taken at the hospital, were stored for up to 6 months, and then following the patient's diagnosis by biopsy, cystology or MRI, were classified as positive for prostate cancer or negative controls. Since these control patients likely had other urological conditions, they were classified as "unhealthy" and were only used in the most advanced stage of discrimination training (Stage 4) when they were age-matched and symptom-matched to cancerous targets.

Sample Processing

A Standard Operating Procedure was followed to avoid cross contamination. Each consenting participant's whole sample was defrosted and spun in a vortex machine for 10 s; separated into several 1 m samples, each decanted into a 1.75 glass vial and marked with an anonymised code. All aliquots of the same code were stored in the same zip-lock bag in the -20C freezer. Samples were selected and defrosted on the day of training and then placed in a refrigerator for no longer that 1 h, before being decanted into 60 ml polystyrene pots for training. Each aliquot was used once during only one training session.

Dogs

The six dogs were all female, there were two Labrador Retrievers, two Labrador crosses, one Cocker Spaniel and one Wire-haired Hungarian Vizsla. At the start of training, dogs ranged in age from 14 to 54 months old. Their training involved four stages which involved gradually increasing the number of samples presented and the subtlety of the difference between target and control. So, dogs started with a small number of controls that were all healthy, the diversity of controls was gradually increased and ultimately included "unhealthy" controls, that may have had a urological condition other than prostate cancer (as described above). Whilst the positive samples were all from men, the controls included females. Although females never occurred in the targets, in the early stages of training we included wide ranging control samples varying in multiple aspects, to encourage the dog to learn the important discriminatory cue (9).

Dog Training – Stage 1

All training was performed in a dedicated room in the Bio-Detection building at Medical Detection Dogs, UK. Dogs were initially taught to recognize the target scent using search games. The target scent was paired with a food or play reward. Gradually the dogs were trained to follow a more formalized search pattern, when-upon samples were presented in either a four-stand line-up or an eight-position carousel into which stainless steel plates each containing a polystyrene pot were placed.

Training used a 100% reward protocol. Dogs were encouraged to search all vials and when a target sample was encountered to show a trained alert behavior (sit and stare), but to show no response to control samples. When a dog showed a correct positive response, it was rewarded with an audible clicker (as a secondary reinforcer) followed by food or a play reward, whilst a dog showing an incorrect alert was ignored, and encouraged to keep searching. Dogs were also trained to carry out blank runs in which no target samples were presented, whereupon they were rewarded for searching all the apparatus but not showing any alerting behaviors.

Due to a paucity of control samples initially dogs were trained over a 70-day period (individual dogs ranged from 53 to 70 days), using 21 positive samples [100% male; aged 28–80 years; all confirmed prostate cancer positive of Gleason score 3+3 to 4 + 3; (10)] collected from Milton Keynes Hospital Urological clinic and 215 control samples collected from external events, by self-declared healthy volunteers (65% male aged 50–80 years). Over this period, dogs on average received 312 (± 75.5) presentations of positive samples (ranging from 230 to 419) and on average 1,088 (± 182.6: range 768–1,260) of controls. Each presentation was a separate aliquot decanted from a sample, and each aliquot was used during only one training session, although during this session, multiple dogs were usually presented with the same sample numerous times. New controls and targets were gradually introduced throughout this training phase.

Dogs performance in all training sessions was recorded using a computer data base, and sessions were filmed using CCTV for later analysis, if required. Whenever presented with a sample, the dog's response was classified as correct (trained alert to a target and no response to a control) or incorrect. When dogs exhibited a hesitation when encountering a sample, but no full alert, since the dogs were in the training phase, this was treated as an alert so in response to a target was classified as a true positive and rewarded whilst when in response to a control it was classified as a false positive and the behavior was ignored.

Over the initial Stage 1, all six dogs were seen to be performing well, with sensitivities (% of positive samples correctly identified) ranging from 92.2 to 94.5% (averaging 93.5%) and specificities (% of the control samples that were ignored) ranging from 78.2 to 91.9% (averaging 87.9%; **Table 1**).

Dog Training – Stage 2

The dogs then progressed to Stage 2, when-upon 79 new control samples and 13 target samples were added to the training pool. The controls were samples that had been collected from volunteers (staff, relatives and friends), 48 male and 23 female and 8 unknown, ranging in age from 18 to 79 years, attending the same clinic as the initial targets and internal hospital recruitment events. All volunteers were self-declared healthy. These new samples were presented in combination of the external MDD controls over a 10-week training period, averaging 482 (± 188.9: range: 221–707) control and 148 (± 123.9 range 122–192) target sample presentations per dog. When training progressed to include these hospital-sourced controls, a noticeable decrease in performance, was seen particularly in specificity which now averaged only 67.3% (43.1–83.3%; **Table 2**).

Examination of the training data showed that the drop in performance was specific to the new control samples, in response to which the dog showed a large number of false positive responses, leading to a reduction in measured specificity. Novel

TABLE 1 | Demographics and training performance of each of the six female dogs.

Dog name	Breed	Stage 1: Initial training		Stage 2: Training including new healthy MKUH control samples		Stage 4: Post investigation re-training using MKUH healthy and unhealthy controls	
		% Sensitivity	% Specificity	% Sensitivity	% Specificity	% Sensitivity	% Specificity
Florin	Labrador retriever	94.2	87.5	93.4	67.8	92.5	81.4
Karry	Labrador cross	94.5	89.9	92.4	83.3	86.5	77.9
Kim	Labrador cross	92.2	91.9	87.1	77.1	83.3	71.9
Kiwi	Labrador retriever	93.9	89.3	75.3	65.5	57.1	88.9
Martha	Cocker spaniel	92.7	78.2	91.9	43.1	Dog Rejected	
Midas	Wire-haired hungarian vizsla	93.7	90.5	90.7	67.1	85.7	79.7

Shaded cells highlight those dogs rejected from the programme due to training issue mid- or post study.

TABLE 2 | Seven urine samples procured for the systematic investigation of confounding factors.

Sample	Vesicle storage site	Sample collection site	Processing methods	Sex	Age in years
1	MDD	MKUH	MKUH	Female	38
2	MKUH	MKUH	MKUH	Female	38
3	MKUH	MKUH	MDD	Male	25
4	MKUH	External	MKUH	Male	19
5	MDD	MKUH	MKUH	Female	34
6	MKUH	MKUH	MKUH	Female	34
7	MDD	MKUH	MDD	Female	30

samples of both controls and targets had been gradually added throughout training so it was unlikely a response to novel samples [e.g., (11)]. Since the new control samples came from healthy volunteers of a similar age range to the initial external samples, we had no reason to assume that they were any harder to discriminate than those used in Stage 1. This suggested that in Stage 1 dogs had learnt to distinguish the original targets from control samples, on the basis of a factor other than disease state.

Now, in order to rectify the dog's training and ensure optimal sample collection and processing and hence training and performance in the future, we aimed to identify which confounding factor the dogs had used. Alerted to a potential training issue, we carried out a methodical and sequential investigation into all factors which could potentially vary between the hospital and external samples. We used a small number of carefully chosen samples to complete this investigation, avoiding wasting precious training samples.

The training team suspected that the processing or storage at the two sites may have differed. We therefore embarked upon an investigative phase. We observed the processes at both sites from collection to delivery to the dog and discussed the procedures with the hospital nursing team to obtain any clues as to systematic differences between sites. We were assured that there were no systematic differences in: type of gloves used to handle the sampling vesicles; disinfectant used to clean areas, or length of time for which samples were stored in the cool box, prior to being placed in the freezer between the two sites.

However, we identified three potential differences:

Vesicle storage site - location where storage pot was long-term stored (MDD or MKUH);

Sample collection site– place where sample was collected (MDD external events or MKUH);

Processing method: at the two sites (MDD or MKUH)

We next meticulously and systematically investigated which of these factors was the causal issue using new samples and all six training dogs.

Stage 3 - Investigative Stage

We recruited five control human volunteers to provide urine: three females and two males. Four gave urine samples at the hospital, one at an external venue. Two participants gave two samples each, one in MKUH and one MDD stored collection vesicles. In total, we procured seven samples (**Table 2**) presenting different combinations of the suspected confounding factors.

We presented these samples to each of the dogs (within an assortment of other targets and the control samples) a number of times (between 11 and 91 presentations per sample) in order to identify which factor was most linked to high rates of false positive alerts. By recording the number of incorrect alerts to each sample (false positives) performed by each dog, we could carry out statistical analysis to identify which factor was most responsible and hence the major confounder. The effects on specificity were estimated from logistic regression models including these three factors and allowing for differences in performance between dogs. The effects were expressed as odds ratios, and least squares means were estimated for each factor.

RESULTS

The results showed that individual dogs vary widely in their specificity (**Table 3**; $p < 0.0001$). Storage site had a marginally significant effect, with samples in MKUH storage vesicles resulting in significantly lower specificity than MDD samples, but that the effect of processing method had the biggest impact (Chi squared = 14.4 $p = 0.0001$). Control samples which underwent Medical Detection Dog's (MDD) processing, were more likely to be correctly ignored than samples undergoing Milton Keynes University Hospital's (MKUH) processing (OR = 4.32), as were those placed in vesicles stored at MDD (OR = 2.11), whilst externally sourced samples were slightly less likely to be ignored (OR = 0.55). The response to each of these factors varied between individual dogs (**Table 4**).

Once the processing sites was implicated as the most important factor, the team watched the sample handling post-patient, the cleaning of equipment and observed that they varied only in a subtle aspect of their processing; whilst hospital-sourced samples were tested using a urinalysis stick dipped into the sample, externally-sourced samples were tested by pouring a small amount the sample onto the stick. Therefore, only the hospital-sourced training samples contacted the urinalysis stick and hence, the dogs had likely learnt to distinguish target from non-target aided by the odor of this stick.

Dog Training: Stage 4

Based on this knowledge, we modified our subsequent training (Stage 4) and processing to ensure standardization (e.g., all samples were decanted and applied to a urinalysis stick externally). Also, having identified the main confounder, the four remaining dogs (two were rejected prior to this stage due to ongoing training issues) were now trained intensely, with a large number of sample presentations per day and concentrating only on teaching the distinction of malignant vs. non- malignant whilst ignoring the previously learnt processing factor. Matched controls from individuals sampled at the same MKUH clinic, but subsequently diagnosed as having non-malignant organ-specific conditions and no history of cancer, were also included at this stage. With very large numbers of presentations of both controls (664–1,016 per dog) and targets (196–376 per dog), including 143 familiar and 217 novel samples, we progressively trained three of the dogs to ignore the processing method and alert based only on disease state, the fourth dog failed to respond to this training and was therefore also rejected.

It is noteworthy, that whilst this rehabilitation training served to teach three of the six dogs to categorize samples based on disease state and they achieved high levels of performance (71.9–88.9% specificity: **Table 1**), three dogs failed either during Stage 3 (Kiwi and Martha), or during Stage 4 (Kim) to be retrained and required to be rejected from future trials.

TABLE 3 | Results of Logistic regression exploring effect of number of factors on specificity of dog's response to control samples.

Effect	DF	Odds ratio	Wald chi-square	*P*
Dog ID	5		26.92	< 0.0001
Vesicle storage site (base category MKUH)	1	2.11 (1.03, 4.33)	4.17	0.0411
Sample collection site (base category MKUH)	1	0.55 (0.25, 1.2)	2.25	0.1332
Processing method (base category MKUH)	1	4.32 (2.03, 9.21)	14.39	0.0001

DISCUSSION

This study supports previous findings [e.g., (8, 9)], that dogs can be trained to detect prostate cancer, but shows that even within a population of all female, similarly selected and trained dogs, individuals showed very different levels of both sensitivity and specificity. Once dogs were seemingly well-trained on initial samples, we saw a decrease in specificity when new control samples were added. This demonstrates that even when trained on target and control samples that were apparently identically collected, processed and stored, dogs had learnt to discriminate targets from non-target, not by the intended disease state, but by a confounding factor.

By using carefully chosen samples with each combination of potential confounders and employing statistical analysis, we were able to identify the most likely causal factor. The results of our logistic regression of training data indicated that the biggest effect on performance at the discrimination task was dog ID, highlighting individual differences between the dogs, each varying widely in their specificity. Collection vesicle storage site had a marginally significant effect, but the sample processing method had by far the greatest impact. This showed that control samples which underwent Medical Detection Dog's (MDD) processing, were more likely to be correctly ignored than samples undergoing Milton Keynes University Hospital's processing, as were those placed in vesicles stored at MDD, whilst externally sourced samples were slightly less likely to be ignored.

The analysis highlighted that a confounder associated with the processing was likely inflating the dogs' overall specificity during the initial training. The dogs appeared to have learnt to use a cue to discriminate samples from one another, and this cue was not only disease state, but something associated with the site at which the processing occurred. The actual reason was not obvious as the Standard Operating Procedure was believed to be identical in

TABLE 4 | Percentage specificity of dog's response when each factor is compared (hesitations without a full alert were classified as alerts and hence constituted a true positive on a target sample and a false positive on a control).

Dog's name	Vesicle storage site - location where storage pot was long-term stored		Sample collection site– place where sample was collected		Processing method	
	MDD	MKUH	External	MK	MDD	MKUH
Florin	88.9	45.0	14.3	80.0	94.7	53.6
Karry	85.7	83.3	66.7	89.3	100	75.6
Kim	95.2	76.2	76.9	87.3	90.0	83.3
Kiwi	52.4	72.7	40.0	63.0	66.7	40.0
Martha	66.7	15.4	15.4	66.7	83.3	27.3
Midas	92.1	59.4	50.0	83.9	96.2	65.9

all cases. It was only by watching and discussing with the onsite healthcare team, that the most critical elements of the process were identified. We discovered the two sets of samples varied in a subtle aspect of their processing; hospital-sourced samples were tested by dipping the urinalysis stick into the sample, for MDD processed samples, a small amount of urine was poured onto the stick. We conclude that dogs had likely learnt to distinguish target aided by the odor of this stick.

It can be argued, that if samples need to be sourced from more than one location, ideally clinical processes should be replicated and an external person should watch the processing and minimize the possibility of confounders pre-training. Potential factors should be identified and eliminated from the outset. However, this can be an onerous task, especially in medical settings when samples which would normally be collected from consented patients by health care professionals who may rotate daily. However, this case study highlights the importance of considering and monitoring every aspect of sample collection, processing and delivery when using a limited number of collection locations, even when using urine for dog training. Although it may have seemed trivial to clinical health care professionals (involved in patient consent and sample collection) whether a dipstick was placed into the fresh urine sample after patient production, or a drop is taken from the urine, we have shown that for dogs working on a highly complex discrimination task, this aspect had a significant effect. Studies of dog training show that given multiple possible cues by which dogs can solve a training task, dogs will learn to use the cue(s) most salient and accessible to them [e.g., (12)], and here it appears that the altered odor created by the dipstick was that cue. Urine was previously believed to be less susceptible to cross contamination and processing effects than more volatile media such as breath (13). It is widely acknowledged that ambient VOC's can contaminate breathe samples [e.g., (7)], but here we demonstrate that even for a liquid medium there is significant risk of cross contamination so standardized processing of samples is essential. It is not known exactly what effect the dipstick inclusion had on the urine sample or how it changed the odor, but the canine performance indicated that it was a significant factor in learning discrimination. Interestingly the extent of the effect varied between dogs.

There is currently limited research examining factors that affect a dog's propensity to generalize or discriminate odors. The balance between generalization and discrimination in odor recognition is affected by target odor molecular structure (14), as structurally similar molecules compete and activate overlapping receptors, making these compounds harder to discriminate (15) and the olfactory threshold may vary for different compounds (16). The tendency to discriminate may also vary with the individual dog's olfactory acuity (8), and with training and reward protocols [e.g., (12, 17)]; and here we suggest also individual personality differences in the dog. This is an area important for future study.

Our results show a moderate significant effect of vesicle storage site, suggesting that ambient atmosphere may have contaminated the storage pots and had some effect on the dogs, but since sampling site did not exert a significant effect, we have no evidence that changes such as time pre- freezing, freezer temperature or differences in procedures when samples are moved from the clinic to the freezer (which varied between sites) were used by dogs to categorize the samples. However, this may be because in our case, the processing methods, and odor of the urinalysis stick was the most salient cue, and we cannot rule out that if the processing cue were absent (due to standardization), the dogs would not have learnt to discriminate based on sample collection site or vesicle storage location to a greater extent. Given the potential for subtle aspects to affect training, we suggest future studies should aim to standardize all aspects and that papers reporting dog detection results should state clearly where and how all samples have been collected and how audits are carried out to ensure that internal and external sites achieve identical processing. Historically this has not always been the case [see (6)].

As pointed out by Edwards et al. (2) the validity of performance and results are threatened when systematic differences between positive and control samples (other than disease status) are present during training phases. Here there was a systematic difference which was pinpointed by a systematic investigation. When training a complex signature in a complex background, we need to ensure dogs learn to accurately discriminate disease state only. This is best achieved by using completely matched samples from a single site. Controls should be from the same clinical environment as the targets, and ideally collected at the same time since ambient VOCs may vary from time to time even within the same environment. If such standardization is impossible and confounders are unavoidable, then we need to maximize the variation in them e.g., by using multiple collection methods, locations and processing methods for both targets and controls to ensure the dogs learn to categorize based only on the target factor: disease state (7). But, when training for novel diseases presented only in a limited number of patients or with a paucity of initial control sample as seen here, it is often not possible and, in such cases, we have demonstrated how when training anomalies arise, a thorough investigative stage is extremely valuable.

Here due to a lack of initial controls we needed to source control samples from additional sites. This was important, in order to avoid the issue identified by Elliker et al. (1) when training two dogs to detect prostate cancer, dogs appeared to memorize the samples and hence not generalize to new samples when presented in double-blind testing. Canine memory is an important consideration when training with a limited number of samples and again points to the necessity for larger training sets collected from the same source or if not, multiple varied sources (1, 18).

The study also demonstrated the value of continually monitoring performance throughout training, in order to be able to rapidly identify if a training problem develops or the performance of dogs is being affected by a confounding factor. Ideally this should be accompanied by rigorous blinding throughout training, so that human cues do not present an additional confounder (7). The electronic monitoring system at MDD allowed us to continuously monitor performance and to analyse individual accuracy on a rolling basis. All training sessions were recorded using an internally developed database system MDD-Olfactory Performance Recording Application (OPRA). Each session was filmed using CCTV and stored on a protected drive for later analysis. Use of this footage, and the analytical methods described here, allowed us to identify that there was a problem, and a systematic and methodical investigative phase allowed the route of the problem to be pinpointed and subsequent remedial training carried out. The technology also aided objective performance measurement and decision-making regarding individual dogs. Trainers often become heavily invested in the dogs with which they work and are challenged to make objective decisions about performance and accuracy of individuals. Being able to review and collaboratively discuss footage, can also allow consensus decisions e.g., before withdrawing a dog from training.

It is noteworthy that of the six dogs starting this trial, three were rejected as a result of the them learning to distinguish based on a confounding factor and trainers being unable to re-train the correct categorization within a reasonable time frame. Two dogs were rejected after Stage 3, and one during and one after Stage 4. Although showing great initial aptitude for the task, having learnt the incorrect discrimination cue, in spite of large numbers of presentations and positive reward-based training, these dogs failed to learn the correct association, and systematic search errors and behavioral issues ensued. In such cases, trainers often find it easier to start with a new dog than to rectify the problem, which highlights the potentially great costs of issues in initial odor training. Interestingly, the training data shows that the extent to which the confounder was used to discriminate samples varied with dog. Whilst **Table 3** suggests that Kim predominantly learnt the disease vs. control distinction, as intended, Martha for example relied heavily on the difference in processing in her decision making. Further research into these individual differences is required.

The importance of minimizing potential confounding cues which the dogs can use in place of the intended categorization feature, is obvious in this medical context. However, this concept applies equally well when training dogs for narcotics and explosives and other targets when-upon dogs often learn, for example, that training and hence rewards only ever occur outside an operational scenario or when seniors trainers as well as handlers are present (19). However, our study shows that when subtle differences apparently lead to training issues, systematic analytic methods can be employed to identify and subsequently rectify the problem.

DATA AVAILABILITY STATEMENT

Data are available at the University of Bristol data repository, data.bris, at https://doi.org/10.5523/bris.3mw2y33y7a1j32v3r7r77t953z.

ETHICS STATEMENT

Retrospective routine dog training data collected during Medical Detection Dogs usual training operations was analyzed. Ethical review and approval for animal use was not required according to institutional guidelines. Owner consent was not required as the dogs were the property of Medical Detection Dogs at the time.

AUTHOR CONTRIBUTIONS

CG, RH, and IA: concept formation. AC, CG, and RH: conducted investigative work. RH and NR: analysis. NR, CG, and RH: manuscript preparation. All authors: edits and approval.

FUNDING

This work was not externally funded but carried out during routine Medical Detection Dogs Operations. The charity relies totally on donations from the public and charitable trusts.

ACKNOWLEDGMENTS

We would like to thank all the study participants, the dog trainers and handlers, Sophie Aziz for her help with the MS and two reviewers for their useful comments.

REFERENCES

1. Elliker KR, Sommerville BA, Broom DM, Neal DE, Armstrong S, Williams HC. Key considerations for the experimental training and evaluation of cancer odour detection dogs: lessons learnt from double-blind, controlled trail of prostate cancer detection. *BMC Urol.* (2014) 14:22. doi: 10.1186/1471-2490-14-22
2. Edwards TL, Browne CM, Schoon A, Cox C, Poling A. Animal olfactory detection of human diseases: guidelines and systematic review. *J Vet Behav.* (2017) 20:59–73. doi: 10.1016/j.jveb.2017.05.002
3. Goldblatt A, Gazit I, Grinstein D, Terkel J. *The Olfactory System and Olfaction: Implications for REST.* (2011). Available online at: https://www.gichd.org/fileadmin/GICHD-resources/rec-documents/REST-Nov2011.pdf (accessed 03 August 2020)
4. Oxley JC, Waggoner LP. Detection of explosives by dogs. In: Marshall, M, Oxley J, editors. *Aspects of Explosives Detection*. Amsterdam: Elsevier (2009). p. 27–40.
5. Gao Q, Su X, Annabi MH, Schreiter BR, Prince T, Ackerman A, et al. Application of urinary volatile organic compounds (VOCs) for the diagnosis of prostate cancer. *Clin Genitouri Cancer.* (2019) 17:183–90. doi: 10.1016/j.clgc.2019.02.003
6. Schoon GAA, De Bruin JC. The ability of dogs to recognize and cross-match human odours. (1994) *Forensic Sci Int*. 69:111–118 doi: 10.1016/0379-0738(94)90247-X
7. Jezierski T, Walczak M, Ligor T, Rudnicka J, Buszewski B. Study of the art: canine olfaction used for cancer detection on the basis of breath odour. Perspectives limitation. *J Breath Res*. (2015) 9:027001. doi: 10.1088/1752-7155/9/2/027001
8. Cornu J-N, Cancel-Tassin G, Ondet V, Girardet C, Cussenot O. Olfactory detection of prostate cancer by dogs sniffing urine: a step forward in early diagnosis. *Eur Assoc Urol*. (2010) 59:197–201. doi: 10.1016/j.eururo.2010.10.006
9. Taverna G, Tidu L, Grizzi F, Torri V, Mandressi A, Sardella P, et al. Olfactory system of highly trained dogs detects prostate cancer in

urine samples. *J Urol.* (2015) 193:1382–7. doi: 10.1016/j.juro.2014.09.099
10. Gleason DF. Classification of prostatic carcinomas. *Cancer Chemother Rep.* (1966) 50:125–8.
11. Pirrone F, Albertini M. Olfactory detection of cancer by trained sniffer dogs: a systematic review of the literature. *J Vet Behav.* (2017) 19:105–17. doi: 10.1016/j.jveb.2017.03.004
12. DeChant MT, Ford C, Hall NJ. Effect of handler knowledge of the detection task on canine search behavior and performance. *Front Vet Sci.* (2020) 7:250. doi: 10.3389/fvets.2020.00250
13. Amundsen T, Sundstrøm S, Buvik T, Gederaas OA, Haaverstad R. Can dogs smell lung cancer? First study using exhaled breath and urine screening in unselected patients with suspected lung cancer. *Acta Oncol.* (2014) 53:307–15. doi: 10.3109/0284186X.2013.819996
14. DeGreeff LE, Simon AG, Peranich H, Holness HK, Frank K, Furton KG. Generalization and discrimination of molecularly similar odorants in detection canines and the influence of training. *Behav Proces.* (2020) 177:104148. doi: 10.1016/j.beproc.2020.104148
15. Mershin A, Wassie A, Maguire Y, Kong D, Zhang S, Moran P, et al. *Methods and Apparatus for Artificial Olfaction*. U.S Patent: US9140677B2 (2015).
16. Walker DB, Walker JC, Cavnar PJ, Taylor JL, Pickel DH, Hall SB, et al. Naturalistic quantification of canine olfactory sensitivity. *Appl Anim Behav Sci.* (2006) 97:241–54. doi: 10.1016/j.applanim.2005.07.009
17. Walczak M, Jezierski T, Górecka-Bruzda A, Sobczynska M, Ensminger J. Impact of individual training parameters and manner of taking breath odor samples on the reliability of canines as cancer screeners. *J Vet Behav.* (2012) 7:283–94. doi: 10.1016/j.jveb.2012.01.001
18. Williams M, Johnston JM. Training and maintaining the performance of dogs (*Canis familiaris)* on an increasing number of odor discriminations in a controlled setting. *Appl Anim Behav Sci.* (2002) 78:55–65. doi: 10.1016/S0168-1591(02)00081-3
19. Gazit I, Goldblatt A, Terkell J. The role of context specificity in learning: the effects of training context on explosives detection in dogs. *Anim Cogn.* (2005) 8:143–50. doi: 10.1007/s10071-004-0236-9

Conflict of Interest: CG, RH, AC, and NR were all employed by Medical Detection Dogs the charity which trained and own the dogs. However, since the study described training issues and their resolution. We do not see this as a conflict of interest.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

REVIEW
published: 02 September 2020
doi: 10.3389/fvets.2020.00597

Selecting Dogs for Explosives Detection: Behavioral Characteristics

Lucia Lazarowski[1], Lowell Paul Waggoner[1], Sarah Krichbaum[1,2], Melissa Singletary[1,3], Pamela Haney[1], Bart Rogers[1] and Craig Angle[1]*

[1] Canine Performance Sciences Program, College of Veterinary Medicine, Auburn University, Auburn, AL, United States, [2] Department of Psychology, College of Liberal Arts, Auburn University, Auburn, AL, United States, [3] Department of Anatomy, Physiology, and Pharmacology, College of Veterinary Medicine, Auburn University, Auburn, AL, United States

Detection dogs are widely considered the most effective and adaptive method for explosives detection. Increases in emerging sophisticated threats are accelerating the demand for highly capable explosives detection, causing a strain on available supplies of quality canines worldwide. These strains are further compounded by rigorous behavioral standards required to meet mission-specific capabilities, leading to high rates of dogs disqualified from training or deployment. Ample research has explored the behavioral characteristics important for assistance, guide, and other traditional working roles, while those corresponding to more specialized tasks such as detection of explosives are not as well-understood. In this review we aim to identify the behavioral characteristics important for operational tasks of explosives detection dogs, contrasting with that of other working roles and highlighting key differences between explosives and other types of detection dogs. Further, we review the available research on methods for assessing and selecting candidate detection dogs and make recommendations for future directions and applications to the industry. Improvements and standardization in assessment technology allowing for the identification and enhancement of behavioral characteristics will be key to advancing canine detection technology in general.

Keywords: detection dogs, detection dog evaluation, explosives detection dogs, working dogs, selection, canine

OPEN ACCESS

Edited by:
Wendy Irene Baltzer, Massey University, New Zealand

Reviewed by:
Mia L. Cobb, The University of Melbourne, Australia
James Serpell, University of Pennsylvania, United States

****Correspondence:***
Lucia Lazarowski
lucia.lazarowski@auburn.edu

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences, a section of the journal Frontiers in Veterinary Science

Received: *16 March 2020*
Accepted: *27 July 2020*
Published: *02 September 2020*

Citation:
Lazarowski L, Waggoner LP, Krichbaum S, Singletary M, Haney P, Rogers B and Angle C (2020) Selecting Dogs for Explosives Detection: Behavioral Characteristics. Front. Vet. Sci. 7:597. doi: 10.3389/fvets.2020.00597

INTRODUCTION

Increasing recognition of the detection dog as the most capable and adaptable method for real-time detection of explosives has led to a world-wide increase in their use in security and military operations, which is straining the supply of dogs capable of performing explosives detection (1). The U.S. Congress has stated that U.S. dependence on foreign procurement and a lack of domestic production of explosive detection dogs (EDD) presents a critical security gap [(2), 115th U.S. Congress]. Military and security officials from numerous nations attending the 2019 *International Working Dog Conference* of the *International Working Dog Breeding Association* noted the dwindling supply of suitable candidate EDDs from traditional private sources. Moreover, EDD tasks are increasingly specialized and sophisticated, further constraining the availability of dogs with the behavioral, physiological, and structural characteristics necessary to perform those tasks.

EDDs are primarily sourced from populations of dogs that have been selectively bred for hundreds of years for hunting, herding, and protection (3, 4). Substance detection tasks, which mostly occur in the context of intense human activity such as urban landscapes, are a relatively recent application of dogs for which there has been very limited directed selective breeding.

Increasing evidence indicates that behavioral characteristics have a greater influence on detection dog success than sensory or morphological differences (5, 6). However, standardized and reliable methods for identifying suitable candidates are lacking, resulting in low rates of dogs achieving operational status and high levels of "behavioral wastage," which has obvious implications for program efficiency as well as concerns regarding animal welfare (7). Given the extensive time involved and economic investment in the preparation of a working dog, as well as a lack of reliable predictors of success, identification and valid measurement of the expression of behavioral characteristics important to EDD performance are essential for accurate selection and, especially, the purpose-breeding of potential EDDs (7, 8) Therefore, better defining and communicating of the key behavioral characteristics of successful EDDs is critical to enhancing the supply of dogs capable of performing contemporary EDD tasks.

While characteristics have been fairly well-defined by research for assistance, guide, and some other working dogs, information about EDD characteristics is largely siloed within and varied across programs and has not been subject to much scientific examination and validation. Recent reviews have examined detection dog characteristics for wildlife/conservation dogs (3, 9), but there have been few systematic examinations and standardization of behavioral characteristics important to EDD performance (10, 11). In this review, we aim to identify behavioral characteristics that by general consensus, our experience in breeding and preparing dogs for explosives detection tasks over the last 20+ years, and pertinent research are important to EDD performance. We also review research on available methods for assessing and identifying candidate detection dogs and make recommendations for future directions and applications to the industry.

TYPES OF EXPLOSIVES DETECTION DOGS

EDD is a general vocation defined by the class of targets to be detected (explosives) that implies some, but not all or even the predominant, capabilities necessary for performing the range of different EDD tasks. EDD tasks are more specifically defined by the parameters of the context and details of the search task. In many cases, the characteristics necessary for these varying EDD search tasks are the same but may vary significantly in the needed degree of expression. Traditionally, the most general EDD (sometimes referred to as "standard" EDD) is a dog that searches an array of areas at the immediate direction of a handler, most often but not exclusively on lead. Such areas include the interior and exterior of varied types of buildings, road vehicles, limited open areas (e.g., a park), and articles such as luggage and boxed goods. More specialized EDD applications are extensions of these general tasks and are often more focused on a particular search task, the context in which that search occurs, or the mode by which that search is performed.

As mentioned, specialization is increasing as a consequence of the growing sophistication of EDD applications with some fairly well-defined specialties. For example, Person-Borne Improvised Explosive Device (PBIED) EDDs interrogate persons or their aerodynamic wake, such as TSA *Passenger Screening* and *Vapor Wake*® canines, respectively. *Specialized Search Dogs* (SSD) and similar variants are remotely-directed dogs working off-lead and down-range, primarily but not exclusively for military applications in detecting IEDs. Variants of the SSD include specific main route (i.e., roadway) clearance and land-mine EDDs. Some specializations are less well-recognized at this time, but becoming increasingly defined by the search task to be performed, such as cargo screening. Of course, many military working dogs (MWD) and law enforcement EDDs are dual-purpose or multi-purpose canines employed for multiple tasks including protection/apprehension or tracking, but these additional tasks are beyond the scope of this review; suffice to say that some specialty EDD tasks, such as the screening of persons, may be less compatible with dogs having the propensities necessary to perform these additional non-EDD tasks.

Specialization may be contextual in nature requiring that dogs exhibit characteristics particularly well-suited for working in particular conditions. This may be the case, for example, of EDDs for maritime operations working on and transferring between vessels, working in the confined spaces of those vessels, and in the loud environment of the engine rooms of those vessels. Some dogs that may not be behaviorally well-suited for particular tasks may possess very suitable characteristics for other tasks. For example, dogs unable to work in large crowds of people may be capable of specializing in interrogating cargo, where other characteristics, such as the ability to search vigilantly for long durations, has primacy. Some search tasks can also be delineated by the required concentration of explosive odor to be detected. For instance, for EDDs in the aviation security sector, trace levels of explosives are important to detect as compared to the SSD dog in a combat theater that may, in some circumstances, need to be conditioned to ignore trace levels of explosives. Such parameters may translate to the intrinsic propensity of a dog to engage in meticulous sniffing without which it may be difficult to condition it to detect trace levels of explosives or the PBIED EDD that may require an intrinsic propensity for air-scenting behavior as contrasted with ground- or object-scenting to perform the task successfully.

BEHAVIORAL CHARACTERISTICS

The behavioral characteristics required of EDD tasks can be generally broken down into three broad categories: detection characteristics, trainability/tractability, and environmental characteristics. Detection characteristics are those related to the style and intensity of interrogation and search for explosive odor. Trainability/tractability relate to the various cognitive, behavioral, and social characteristics necessary to be trained to perform the particular search requirements. Environmental characteristics refer to the collection of traits enabling a dog to work effectively in the particular search context, such as the high-stimulus settings of a large event venue, crowded mass transit stations, or military combat. In the following sections we aim to review these characteristics, as well as methods for evaluating the degree of expression of the various characteristics.

Odor-Guided Behavior

Olfaction is undoubtedly a critical aspect of explosives detection. Olfaction is considered a primary sense for canines, but a large degree of variation in olfactory acuity exists due to differences in olfactory receptor genes and conformation based on selective breeding practices for morphological features designed to enhance olfactory ability. For example, differences in nose shape and population of odor receptor cells differ greatly between breeds selected for olfactory-based tasks and non-scenting breeds (9), leading to differences in olfactory threshold (12). However, sensory and morphological characteristics are considered secondary to behavioral characteristics in determining suitability as an operational detection dog (13). Rather, the specific type of odor-guided behavior used to identify and locate a scent, and a dog's propensity to use olfaction in general, is essential for effective operational search performance.

Search Technique

Different types of searching involve different search techniques, and thus the type of search technique desired will depend on the type of task. For example, air-scenting involves sampling odor molecules in the air, as opposed to on the ground or from objects, where the dog searches for the target odor by sampling the air currents in order to identify and work an odor to its source (9, 14). PBIED EDDs use air-scenting to detect airborne odor molecules, following the path of odor of the moving person. Due to the bilateralism of canine olfaction which allows dogs to determine the direction of an odor source by differential sniffing with each nostril, air-scenting is likely performed by detecting airborne scents in open areas without a scent trail to follow (9). Thus, an advantage of air-scenting is the ability to cover more ground in a shorter amount of time (15), and also allows dogs to locate some static targets more efficiently and directly by using air currents instead of following a path. There is a continuum, of course, between air-scenting and ground-/object-scenting as all odors are generally airborne; the difference is in the degree to which a dog attends to open air-space vs. its tendency to attend to and interrogate the ground or objects for target odor. Such tendencies are the result of both intrinsic qualities resulting from breeding and experience/explicit training. In selecting dogs for any detection task, the predominance of odor-guided behavior over other stimuli influencing behavior is a key foundational characteristic.

While all dogs are capable of air-scenting, which can be further fostered through training, breeds that naturally exhibit air-scenting (e.g., dogs selectively bred for upland game hunting) are often selected for such tasks. Potential advantages of using natural air-scenting breeds for detection work have not been systematically explored, and current evaluation methods do not typically account for these natural preferences (9). However, in a recent examination of the behavioral characteristics associated with dogs bred and trained for PBIED tasks, air-scenting ability assessed at 6 months was predictive of dogs' future placement as a PBIED EDD vs. traditional EDD, and was the only behavioral measure that distinguished the two types of outcomes at this age (16). This finding is consistent with the notion that air-scenting has a genetic basis, with such predispositions appearing earlier in development before extensive training may obscure differences. Thus, selecting dogs based on a natural propensity for the desired type of search method will likely result in reduced training time.

Propensity to Hunt

In addition to search technique, a dog's propensity for olfactory-based searching in general is an important characteristic in its success as an EDD. While most dogs can be trained to perform searches, as evidenced by the popularity of Nose Work® as a sport for pet dogs of various breeds, some dogs are more naturally inclined to hunt with their nose without having to be trained to do so due to the intrinsically reinforcing nature of engaging in the hunt itself. Engaging in intrinsically reinforcing behavior in a non-functional context is thought to underpin the behavior of most working dogs, such as sled dogs racing as a form of play, or border collies showing of eye (17). A related but somewhat different example is that of the pointer, for which the stalk and pointing behavior has been greatly exaggerated and is presumed to be so strongly genetically controlled that external reinforcement is likely not necessary to maintain the behavior (18). Thus, the desire to hunt can be a powerful motivator for sustaining endurance and engagement during long searches where the probability of encountering a target odor, and thus receiving a reward, is low (3). Further, selecting dogs for which hunting is intrinsically reinforcing is likely to significantly reduce training time.

Even within hunting breed groups or within breeds, differences may exist as a function of the modern utility of the dog. For example, retrievers bred for hunting upland and non-waterfowl bird species (i.e., retrievers that also serve the function of a pointer and a flushing spaniel) hunt using their nose, detecting cryptic avian species (i.e., concealed or camouflaged). On the other hand, retrievers bred and trained for waterfowl hunting locate downed prey using visual cues and memory, relying less on their nose, in order to maximize the efficiency of retrieving the downed bird and minimize disturbance that may deter other birds in the area. This visual-based searching is even more enhanced in dogs bred and trained for competitions such as field trials and hunt tests where the primary focus is waterfowl hunting, which are further removed from the traditional utility of the breed, as is also seen in herding and sled competition trial dogs (17).

Evaluating a dogs' natural hunt ability can be measured by observing the pattern, efficiency, and intensity of the dogs' search. For example, when searching in a complex environment, efficiency will be improved if dogs ignore visual targets and only attend to odor cues in the air, using air currents to locate the source odor directly. In a test developed as a measure of search ability known as the Brownell-Marsolais scale (19), dogs are tested for their willingness to search for an object thrown into thick brush after varying intervals of time between the hiding of the object and when the dog is released to search. Higher scores are given for dogs that enthusiastically search the area without hesitation to enter the brush and locate the object. However, it is likely that other factors may influence performance on this test such as desire for the reward, memory for where the object was thrown, and sensitivity to the environmental aspects (i.e., entering and walking through the brush). In order to isolate natural hunt ability from other, likely interrelated, characteristics, attention should be paid to dogs' ability to hunt methodically using an instinctive pattern, efficiently searching areas with no encouragement or direction from the handler required. One could envision research that related wind current conditions, dog movement and sniffing, and ultimately, target detection that could enable the development of a standardized means of testing candidate EDDs for their relative efficiency and accuracy in using air currents to detect an explosive target.

Reward Value

As discussed above, while some dogs find the opportunity to hunt reinforcing in itself, training a dog to perform a specific type of search task and to locate specific target odors typically involves using some type of reward (i.e., reinforcement) for performing the correct behavior. For example, teaching a dog to use a precise search pattern, respond to directional cues, detect an artificial chemical odor with no biological relevance, and communicate a find by performing a trained alert all require the use of operant conditioning to teach the desired behaviors. In order for something to function effectively as a reinforcer for a behavior (i.e., the behavior will be repeated in the future as a function of that reinforcer), the dog must regard the reinforcer as a high value reward- or at least more rewarding than competing sources of reinforcement available in the environment.

Food is a primary reinforcer for all organisms, meaning that an animal will work to obtain a primary reinforcer with no prior learning required due to its biological importance. Food rewards can be highly effective in training a dog, and is the preferred method used by the Bureau of Alcohol, Tobacco, Firearms, and Explosives (ATF) (20). However, the efficacy of food as a reinforcer can be influenced by individual differences in preferences and genetics, and may be impractical for use in some operational contexts. Furthermore, the reinforcing value of food decreases as satiety increases, though it can be increased through food deprivation. As such, training employing food rewards is often conducted prior to feeding using the meal ration during training, but satiety will eventually limit the number of repetitions that can be performed.

Toys are a popular reward used in detection dog training and are the preferred reward method of the military (20). The opportunity to chase, possess, and play with a toy is highly intrinsically reinforcing for some dogs. The reinforcing nature of playing with an inanimate object likely taps into the canine predatory motor sequence, a sequence of innate behaviors engaged in during the pursuit and apprehension of prey. Wolves engage in the full predatory sequence beginning with orienting toward the prey, triggered by its movement, and ending with dissection and consumption (21). Through selective breeding, the presence and intensity of parts of the sequence have been modified in domestic dogs, and in particular in working breeds which have and continue to experience strong selection for the expression of these patterns (21, 22). For example, dogs originally bred for assisting hunters (e.g., Spaniels and retrievers) exhibit exaggerated portions of the sequence related to chasing and grabbing, but not killing, as this part of the sequence would be counterproductive to the hunter. In herding breeds (e.g., border collies), the stalking and chasing portions of the predatory sequence are more greatly exaggerated (22). However, selective breeding has led to differences even within herders, such as German shepherds bred for protection (i.e., Shutzhund) which exhibit the orient, chase, and grab-bite (17).

For some breeds, such selection has led to engaging in these actions toward non-edible objects being reinforcing in itself, manifest as play behavior and an obsession-like desire for object-play (3). Moreover, the act of performing the behavior appears to be intrinsically rewarding and is unrelated to satisfying nutritional needs (23). The desire for object-play is then harnessed as a potent reinforcer allowing for the repetition of hundreds of trials without the risk of satiety (3), and functions as a powerful motivator to work over extended periods of time. In a comparison of three breed groups, retrievers were more likely to engage in solitary play with an object than livestock guarding dogs (which show no portion of the predatory sequence) and herders (22). Given such robust breed differences, object-play likely has a strong genetic basis and is likely to be evident early in development. Indeed, the tendency to retrieve an object has been shown to be predictive of future police dog suitability as early as 8 weeks of age (24). Similarly, a factor identified as "attitude toward predation" (comprised of willingness to chase, catch, and fetch a tennis ball as well as follow a dragged object) was predictive of police dog success as early as 7 weeks, and measures of fetching in 8-week old German shepherd puppies in a MWD program has been reported to be highly heritable (25).

The desire to maintain physical possession of an object is widely considered an important trait for a successful detection dog and likely reflects a high degree of intrinsic reward value (10, 11, 16). This characteristic, often termed "physical possession" or "object possession," is often assessed by engaging the dog in tug-of-war play and measuring the dog's force and determination in maintaining its grip on the toy (26). However, when procuring dogs from gundog populations, in particular waterfowl dogs, it is important to differentiate natural possession (i.e., resulting from genetic selection) from a conditioned retrieve with a "soft-mouth" hold (i.e., resulting from negative reinforcement training). A dog with trained possession will typically only exhibit the beginning portions of the predatory sequence (i.e., retrieving and holding), which is less natural and likely to decay over time. A dog with a strong prey-related desire to possess the item will, if left to its own devices, engage in further behaviors such as thrashing and chewing.

A misconception in traditional working dog assessments is that a dog that relinquishes its reward after a target find has low reward value (i.e., lack of possession), but it may be that dogs with a high propensity to hunt, especially those that have been conditioned to perform multiple searches consecutively, may give up the reward for the opportunity to return to searching. It is also important to consider that such tests may be measuring multiple and potentially overlapping constructs; for example, a dog's engagement in a game of tug-of-war may be influenced by its desire to gain possession of the item, its desire to interact with the person playing, or both. Furthermore, it may be an inherent trait for dogs such as retrievers to have a propensity to return and drop a thrown toy at a handlers' feet in order to gain access to the opportunity to again retrieve the thrown object.

An analysis of tests used by the US Transportation Security Administration (TSA) for assessing dogs' suitability for explosives detection found that dogs' willingness to carry an object absent any external input reflected an underlying trait termed "independent possession," which appeared to measure a different underlying construct than that termed "dominant possession," the latter which was characterized by the duration and strength of grip during a game of tug. "Dominant possession" accurately predicted dogs' selection outcome whereas "independent possession" did not, indicating that the interactive nature of the tugging game may have reflected traits more important than independently possessing the object, such as a desire to interact with a person (26). Similarly, an analysis of the tests used to measure suitability of MWDs for the Swedish Armed Forces revealed that physical and social engagement were interrelated and both were predictive of training outcomes (27). Possession of an object and corresponding engagement with a handler during a tug game has also been shown to be predictive of suitability as an EDD as early as 6 months of age, indicating that this trait may be relatively genetically influenced and stable across development (16).

Some tests have differentiated between physical possession and "mental possession," defined as the tendency to focus on an object or on the location where an object was hidden, and maintain focus over a period of time or despite distractions (6, 10, 26). This test is similar to the classic delayed-search task (28), and is likely a measure of sustained attention or memory. The more desirable the object is, the more motivated the dog will be to attend to its location over an extended period despite distractions or to remember the location of its placement after a delay. Thus, this test is probably influenced by a number of factors including reward value, attention, and arousal. Indeed, MacLean and Hare (29) found that performance on a delayed-search problem-solving task in which dogs were required to remember the location of a hidden reward after varying intervals was predictive of detection dog outcomes.

Another way to measure reward value is to assess a dog's persistence in attempting to obtain a reward. Persistence in dogs can be assessed by measuring the amount of time a dog spends attempting to gain access to an unattainable reward (e.g., a toy locked inside a container) before giving up, a test known as the "Unsolvable Task," which has recently been used as a measure of detection dog suitability (29, 30). Persistence is generally considered a desirable trait as it likely reflects motivation for the reward, but in some cases, persistence can be a sign of learning difficulties related to an inability to flexibly respond to changing contingencies (31). For example, Dalal and Hall (32) found that greater persistence, measured as continued responding after reinforcement was discontinued (i.e., extinction), was associated with poorer olfactory discrimination learning. This suggests that high levels of persistence could be associated with an increased tendency to commit false alarms due to a decreased sensitivity to extinction. Further, Lazarowski et al. (30) found that glancing back and forth between the inaccessible reward and a nearby person during the unsolvable task as if requesting help was predictive of future placement as a detection dog, as opposed to dogs that persisted independently. Therefore, it may be that a balance between a strong desire to work for a reward with the ability to shift strategies when a response becomes ineffective is most desirable, which may be a sign of trainability (discussed below).

Task Engagement

Many tests of working dog suitability include assessments of a dog's willingness and ability to stay engaged while performing a search (8). However, this general behavioral characteristic likely reflects a number of underlying traits rather than a unitary construct. For example, the desire to obtain the reward for completing the task as well as the reinforcing nature of the task itself is likely to influence individual willingness to work. Maejima et al. (33) found that drug detection dog success could be predicted by a general factor termed "Desire to Work" which consisted of several seemingly disparate underlying traits including increased general activity, ability to obey commands and concentrate during training, a greater degree of anxiety, and interest in a dummy object. Sinn et al. (6) identified a factor termed "Search Focus" which consisted of dogs' ability to search vigorously without handler input or interruption, using olfaction rather than vision, combined with physical stamina during the search (i.e., ability to search over large areas and long periods without physical signs of fatigue); despite the subtests reflecting a global construct related to search ability, the measure was not predictive of odor detection certification.

Task engagement may be best characterized as the dogs' level of independent engagement while searching. This is distinguished from propensity to hunt, which specifically refers to the dogs' willingness and ability for olfactory-based investigation, though a dog with a high propensity to hunt is likely to remain highly engaged in the task. However, task engagement also takes into account the level of the dogs' independence and ability to work without handler guidance or encouragement. A dog with low task engagement may require excessive handler direction in order to engage and remain engaged in the search, or may stay close to the handler and become easily distracted. A dog with high task engagement will remain engaged in the search independent of the handler until the target is located, immediately returning to continue searching after being rewarded.

Detection dog suitability tests also often measure distractibility, which is inherently part of a dogs' ability to stay engaged in a task. However, distractibility is thought to be a multifaceted construct and is not well-characterized in dogs (34). For example, one study found that in a population of drug detection dogs, aggression toward other dogs, low obedience, and a desire to play with humans were all related to a common construct thought to reflect distractibility (33), yet distractibility was not predictive of training success. Dogs may become distracted for a number of reasons, which may underlie different phenotypic traits. For example, disengaging during a search task could be due to impulsivity, or due to a general lack of interest (e.g., low hunt or reward value) (34). On one hand, whether a dog is able to be easily distracted from a task is important to know regardless of the cause of the distraction. However, the cause of the distraction may also be important. For example, a detection dog easily distracted by people may be able to work effectively in an environment without people, but knowing whether the distraction is due to a fear of people or an attraction of people would be important for accurate phenotypic characterization.

Sociability

Detection dogs work as a team with a handler, and thus must be able to work effectively with humans. Responsiveness to human commands, e.g., influences the ease in which a dog is trained and is critical for the ability to be directed by the handler while working. For dogs that work down-range at a long distance from their handler, such as directionally-controlled Improvised Explosive Device Detector Dogs (IDD), the ability to respond to handler commands is imperative to the team's success and safety. Indeed, a recent study found that the ability to utilize human gestures in a problem-solving task was associated with desirable IDD outcomes (29). For these reasons, breeds originally selected for working cooperatively with humans (e.g., herding dogs, gundogs), selected for their ability to work while maintaining visual contact with their human partner and taking commands from a distance (35), tend to be favored for a variety working roles involving working as a team with a person. Conversely, "independent worker breeds" (e.g., scent hounds, livestock guarding dogs) were bred for working independently with minimal human interaction. For example, bloodhounds were selected for a steadfast persistence in pursuing an odor trail independently and over long distances, making them excellent tracking dogs. However, the trade-off is that this "single-mindedness" which allows them to focus entirely on the odor and ignore distractions can make them stubborn, disobedient, and difficult to train (3). For this reason, scent hounds are rarely used in explosives detection despite their purported superior olfactory acuity.

The ability to cooperate also relates to a willingness to please. Wilsson and Sundgren (36) described this trait as "the tendency to be influenced by the handler without being given a direct command or sign," and found breed-related differences in scores for this trait. Scores were higher in Labrador retrievers compared to German shepherds, and was the most heritable trait for the labs (37). Cooperability/willingness to please was found to be a separate behavioral trait than a willingness to make contact with people, termed affability, which was also found to be higher in Labradors than German shepherds. The authors attributed these breed differences to the genetic history of the Labrador, originally used as hunting dogs that worked closely with their human partners, while German shepherds were used for herding and livestock guarding and more recently as police and protection dogs. Despite the importance of dogs' desirability to interact and work with people in their effectiveness as a team, EDDs should not be so attracted to people to the point that they become distracted (11). While play with the handler is likely to be reinforcing and highly effective for training, the task itself must be more rewarding in order for dogs to work effectively.

Similarly, dependence on the handler is an undesirable characteristic as detection dogs need to be able to work independently without constant guidance (11). Dogs are incredibly sensitive to human body language and other social cues, and so too much attention to the handler could interfere with the dogs' ability to make independent decisions (38). In a recent study, adolescent candidate explosives detection dogs that ignored a human's inaccurate pointing gesture that conflicted with olfactory information were more likely to be selected as EDDs in the future than dogs that followed the deceptive gesture (39). On the other hand, dogs that are too independent may be stubborn and difficult to train or control. For example, as mentioned above, candidate EDDs that independently persisted longer on an unsolvable task (attempting to obtain a reward from a locked container) were less likely to be selected for working roles in the future than those that looked to the handler, considered a sign of soliciting help, suggesting that some degree of social sensitivity to people is important (30). The degree of desired independence likely depends on the type of task, where dogs trained to work off-leash need to be able to range far ahead of the handler but be responsive to directional controls given by the handler at a distance.

Trainability

The speed and ease in which a dog learns a new behavior or task, or trainability, is clearly an important characteristic as rapid learning will lead to faster and thus more efficient training. A greater desire for the reward will enhance attention to the task, motivation to remain engaged, and thus increase training efficiency (9). Thus, trainability is likely a multifaceted construct involving several other processes like attention, as learning also requires sensitivity to changing contingencies. Trainability may therefore be a difficult construct to measure using a single test. Trainability in pet dogs as assessed by The Canine Behavioral Assessment and Research Questionnaire (C-BARQ), a widely used assessment of dog behavior with established reliability and validity, is defined as an aggregate score including attention to the owner, obedience to simple commands, fetching objects, responding positively to correction, and ignoring distractions, and has demonstrated high heritability (40). Trainability measured by the CBARQ has also been shown to predict MWD success rate, and correlated with a behavioral test of "physical engagement" which consisted of tug of war, chasing, interest in object, and persistence in searching for a hidden tennis ball (41). Trainability is likely a combination of the various traits described thus far, which, when evaluated in conjunction, provide a strong predictor of working dog success.

Emotional Reactivity

Arousal

Detection dogs are typically selected for high activity levels, and high energy is thought to result in strong motivation and willingness to work. Handlers of actively working police dogs in the UK reported higher levels of "energy and interest" observed in their dogs compared to reports of those withdrawn from service for behavioral reasons as well as a population of pet dogs (34), suggesting that higher energy levels may be associated with desirable working dog characteristics. However, high energy and activity is often associated with increased arousal. For example, Belgian Malinois bred as MWDs exhibit high levels of excitement often resulting in spinning behavior when kenneled, which has been reported to be higher in individuals with better work performance due to a stronger desire to work (42). Maejima et al. (33) found that higher levels of anxiety were associated with a stronger desire for work and successful certification as a drug detection dog. Given that physiological arousal is closely associated with stress (e.g., increased cortisol) (43), it may be that selecting for high levels of energy and arousal carries with it an increased general reactivity.

While arousal can reflect either negative (e.g., stress) or positive (e.g., excitement) emotional states, high levels of arousal regardless of the underlying affective mechanism can interfere with the ability to perform a task. The phenomenon known as the Yerkes-Dodson law (44) has been well-established in humans, and recently in dogs (45), demonstrating that there is an optimal level of arousal for successful performance on a task where increasing arousal can improve performance on a task up to a certain point, after which performance begins to decline. This effect varies by individual baseline arousal level; for example, the inhibitory control abilities of service dogs with low baseline levels of arousal benefited from a boost in arousal, whereas, the performance of pet dogs with higher baseline arousal suffered when arousal was increased (by exciting the dog) (45). Further, increased arousal can impair learning, memory, and decision making (43). Increased arousal is characterized by activation of physiological responses, such as increased breathing and heart rate. Thus, increasing arousal may result in heavy panting, which reduces olfactory ability as dogs are not able to sniff and pant at the same time (43, 46). While high energy levels are important for sustaining motivation during long searches, excessive arousal may interfere with endurance. For dogs with high baseline levels of arousal, the excitement of searching and being rewarded with play could increase arousal to suboptimal levels and interfere with stamina. Therefore, on-task arousal should be evaluated while working over a period of time, assessing cumulative effects. Adverse signs of arousal that may interfere with performance include open-mouth searching, whining, salivating, frantic searching, agitation, decreased ability to safely navigate the search area, and difficulty handling, increasing over the course of the task. Importantly, low arousal can also be suboptimal if dogs are uninterested or unmotivated, and thus arousal should be evaluated in conjunction with other measures like task engagement.

The anticipation of beginning a task can also lead to an increase in arousal, which can interfere with subsequent performance. For example, if a team must wait before entering a building to start a search, the anticipation during the wait can be stressful for some dogs. Anticipatory arousal can be measured using the same tests traditionally used to measure dogs' ability to locate a thrown object after a delay, but by measuring the dogs' behavior during the delay while restrained such as whining, barking, spinning, aggressing toward the handler, and other behavioral signs of stress indicating the dogs' inability to manage the frustration of anticipation. For example, how much time a dog spent running and restless while restrained by the handler was indicative of "energy management," considered an important trait in selecting dogs for explosives detection (26).

A dogs' ability to manage arousal levels while off-duty, referred to as "off-duty calmness," has been reported by handlers as an important but often overlooked characteristic for EDDs (10). This may be due, in part, to a larger systemic problem related to a lack of feedback to procurement teams from trainers and operationally deployed teams. For dogs living in homes with their handlers, the ability to adapt to the home environment and remain calm while off-duty is clearly important to the dogs' ability to adapt to the handler's home life and ease of management. For dogs living in kennels, the ability to relax when not working may be indicative of effective energy management or general anxiety. One study examined the behavior of guide dog candidates in the kennel and found that a greater amount of time spent resting was predictive of certification (47). The authors speculated that resting during the evening allows for better concentration during training the following day. Further, as in humans, sleep has been shown to be important for memory consolidation and learning in dogs (48, 49). Unpublished data indicated a similar pattern in a population of EDDs, in which dogs that had been successfully selected for service spent a greater proportion of time resting in the kennel than dogs that had been rejected, and that the amount of time spent moving in the kennel was associated with poorer reactions to novel objects, visual startles, and people during a behavioral test (50). In this case, a lack of resting in the kennel likely reflects underlying anxiety and hypervigilance that interferes with the ability to relax. Thus, selecting dogs that are able to "turn off" and appropriately channel arousal is critical.

Fearfulness

Detection dogs are exposed to a range of unpredictable stimuli in the environments in which they work. Therefore, an aspect of behavior that is critically important to their success is resilience toward potential stressors, referred to by a number of terms including environmental soundness (16), environmental sureness (27), environmental stability (26), nerve strength (19), emotional reactivity (51), courage (36), and sensitivity to aversives (8), and are commonly reported as primary reasons for rejection or disqualification across working dog programs (16, 52, 53). Fearfulness can be detrimental for most working dogs, but especially for EDDs working in mass-transit areas with large crowds of people, noisy ambiance, or urban environments with a large variety of novel stimuli to be encountered. For this reason, the level of soundness required is likely comparable to that of guide dogs; however, guide dogs must be wary of potential dangers in order to safely navigate their handler and such wariness has been shown to be predictive of future guide dog selection (54), whereas EDDs do not have this level of responsibility.

Evaluating environmental soundness typically involves presenting dogs with a series of anxiety-provoking situations, with the goal of identifying behaviors during the tests that may reflect the dogs' ability to work effectively in a range of environments (8). Fearfulness in dogs is often characterized by *approach-withdrawal* tendencies, including avoidance of a stimulus, exploratory behavior, and reactions toward stimuli (8, 27, 47, 51, 55–57). Other measures include identifying the presence and severity of specific behavioral indicators of fear (e.g., posture, tail position, lip licking, freezing) in response to stimuli as a measure of *sensitivity to aversives* (8). Both initial reaction and subsequent recovery may yield important information, as the initial startle response likely reflects general autonomic nervous system sensitivity, whereas recovery reflects the ability to cope with the stressor (8). Repeating an exposure can also be informative, as decreased sensitivity upon repeat exposures may be indicative of adaptability and desensitization, and increased sensitivity indicates an inability to cope. For example, Tomkins et al. (47) found that the longer it took a dog to settle after the third repetition of an acoustic startle (a metal plate hitting a concrete floor), the less likely it was to succeed as a guide dog. This suggests that a transient sign of fear with immediate recovery may be acceptable and indicative of resilience (58). The degree of acceptable fearful behavior likely depends on the nature of the dogs' role, or the goals of the selection. For example, greater fear responses may be acceptable for a dog that will be deployed in lower-intensity situations and if the specific fear is likely amenable to overcoming with training. If selecting for breeding, a lower degree of fearfulness will be more important as fearful behavior is known to be heritable (53). It is also important to measure behavior using a wide range of tests rather than a single component, as fearful behavior in isolated incidents may be less problematic than consistent fearfulness across a range of contexts which could be indicative of a more stable fearfulness trait (59). Further, aggregate scores of canine behavior assessments have been shown to be more predictive of future behavior than single measures (24, 54, 60). Indeed, fearfulness has been characterized as a Fear/Reactivity personality dimension in dogs (61), and influences a range of important working dog outcomes (52). For example, Svartberg (57) found that a personality dimension of shyness-boldness reflected fearfulness as well as general learning ability, indicating that boldness (i.e., more exploratory and outgoing) likely facilitates learning due to encouraging interaction with the environment, persisting against challenges, and being less distracted or inhibited.

Social fears

Studies have indicated the presence of two separate and distinct aspects of fearfulness, one relating to social fears (e.g., toward unfamiliar humans and dogs) and another to non-social fears (e.g., inanimate objects) (52, 62, 63). For EDDs that will be expected to work in environments where they may encounter people or other animals, such as in airports, public venues, and mass transit areas, social fears may interfere with the ability to work effectively. For dogs that work in close contact with people, such as passenger screening dogs, friendliness toward people may be an important consideration regarding public perception and level of comfort (11).

A common way to measure social fear toward people is to evaluate dogs' greeting behavior toward an unfamiliar person, such as willingness to approach a stranger, as well as body posture and other fear behaviors during the approach or during interaction with the stranger. Other stimuli have been used when the safety of the human is a concern, such as human-like dolls or dummies (64). Fearfulness toward other dogs has been measured by evaluating the dogs response to another "stimulus dog," a fake model dog, a picture of a dog, or a mirror reflection, though any artificial representation of a dog will not provide social information (e.g., odor, movement) that could influence responses (64). Social fears have been shown to elicit a greater fear response than inanimate objects in dogs, with dogs only vocalizing in response to social fears suggesting a communicative intent of the behavior (65). It can then be speculated that any vocalizations toward inanimate objects may be due to the dog perceiving it as a person or animal.

Non-social fears

Detection dogs also encounter a variety of non-social stimuli in their working environments that could potentially interfere with their ability to complete task. Non-social stimuli are typically characterized as mobile/animated, immobile/inanimate, acoustic, and visual (64) and reflect a separate category of fear than social fears (63). Below are various types of non-social fears relevant for EDDs and common ways to test for them.

Tactile Detection dogs must be able to continue searching without hesitation across a variation of surface textures (19). For example, EDDs completing a building search must be comfortable walking across varying surfaces such as slick flooring, and searches in urban areas may require traversing open grates or unstable footing. In addition to underfootings, detection dogs must also not show sensitivity to body contact, which could inhibit the ability to search in tight spaces, navigate over an obstacle, or under objects hanging overhead. In this regard, an explosives detection dog's confidence encountering a variety of tactile stimuli is similar to that required by search and rescue dogs required to navigate over rubble piles and unstable structures.

Elevation Explosives detection dogs must be able to climb tall structures and search on elevated surfaces if necessary. While a fear of heights in most mammals represents an innate and evolutionary-based aspect of self-preservation (59), fear that prevents a detection dog from searching an area of concern can be a performance-limiting factor. Fear of elevation can be assessed by observing dogs' willingness to approach a ledge or to jump off of a raised surface (e.g., out of a truck), or more formally using a catwalk or elevated maze (59). For example, King et al. (59) found that dogs spent significantly more time in the closed arms of an elevated (1.5 m high) plus maze than the open arms, suggesting that the open arms were somewhat aversive.

Stairs Behavioral tests for working dogs also typically involve testing dogs' willingness to ascend and descend stairs, which may or may not reflect fear of elevation. A variety of types of stairs should be tested including open-backed or open-grate stairs, which may invoke a greater elevation-based response than closed stairs. Approach of the stairs as well as behavior while on the stairs should be assessed as the ability to traverse the stairs in itself may not be indicative of a lack of fear; for example, a dog fearful of stairs may rush over them. Fear of stairs has been shown to be unrelated to other non-social fears, suggesting that a fear of stairs may represent a different underlying fear that develops separately from other fears that appear to be related (e.g., unusual or unfamiliar noises or objects) (63). This would suggest that if fear of stairs is not related to an underlying fearfulness trait, that with sufficient experience and training on a variety of types of stairs the fear can be diminished. However, in contrast to this, Wilsson and Sinn (27) found that fearful behavior on a metal staircase was associated with fearful behavior in a dark room as well as in response to an acoustic startle test, all of which when combined appeared to reflect a trait considered to measure "environmental sureness."

Auditory Noise sensitivity is a common reason for release from working dog programs (66). For detection dogs, such sensitivity can be detrimental to the ability to work effectively as noises from machinery, traffic, blasts, gunfire, and general urban noises (e.g., loudspeakers, people talking) can be distracting or produce anxiety. A popular test used widely by working dog programs is the acoustic startle test, in which dogs' response to a sudden and loud noise is measured. The acoustic startle response is a fast contraction of the muscles elicited by a sudden and intense sound and is present across all mammals (67). Common tests of acoustic startle in working dogs include response to a gunshot, metal objects being dropped on hard surfaces, or other sound blasts (24, 55). Fearful reactions to noises such as fireworks and thunderstorms are typically evidenced by freezing behavior (68), while the latency to recover from an acoustic startle may be a reliable predictor of future success (47, 69). Reduced fearfulness and greater exploration in response to noise at 7 weeks of age was predictive of success as a police dog as an adult (25). Another study found that responses in a gunshot test were not predictive of future police dog outcome, which was attributed to likely prior desensitization because all puppies tested had been exposed to gunfire as part of their socialization (24).

While acoustic startle can be greatly diminished through proper desensitization during early development (70), evidence suggests that this response has a genetic basis. For example, (37) found breed differences in working dogs' responses to frightening situations and gunfire in which Labrador retrievers scored higher (less reactive) than German shepherds. Again, this difference was attributed to the breed history of the retriever, selected for working closely with hunters and withstanding gunfire at close range. Indeed, breeds commonly used for sport-hunting (e.g., Labrador retrievers, Cocker Spaniels, Springer Spaniels) have a reduced tendency to show an acoustic startle response. Researchers have speculated that genetic variations associated with hearing loss may be responsible for reduced startle responses in hunting breeds, but confounds of exposure resulting in habituation or possible damage to the auditory system cannot be ruled out (71). Thus, it is possible that selection for dogs that are less responsive to gunfire has actually modified physiological or anatomical characteristics making some sub-populations of dogs less sensitive to loud sudden noises.

Visual Similar to acoustic startles, visual startles in which an object suddenly appears are commonly used in tests of working dog suitability, such as an umbrella opening, a dummy popping up, or a bag falling in front of the dog. The severity of the startle response, time to recover, and exploration of the object are then measured. A visual startle test measuring the dogs' reaction to a person suddenly jumping out in front of the dog was found to be predictive of future police dog performance as young as 12 weeks of age (24). Reactions including running away and avoidance were associated with poorer outcomes whereas not attempting to run away, and even barking or trying to attack the stranger, was associated with successful outcomes. The type of reaction is clearly related to the nature of the role, where in dogs with a protection role confronting a potential threat aggressively is more desirable. Similarly, Foyer et al. (58) found that stronger emotional reactions and higher levels of cortisol in response to potentially fear-inducing stimuli was predictive of placement as an MWD. For explosives detection dogs, an unremarkable reaction would be most desirable. In particular, for EDDs working around crowds of people, this type of reaction would be undesirable and potentially dangerous. Further, behavior that may be perceived as overly confident, such as moving toward a threat, may actually reflect the dog attempting to actively control the situation driven by fear as fearfulness is sometimes exhibited as an active reaction or agitation (54, 55).

Novelty Because EDDs will likely encounter novel situations on a daily basis, behavior in unfamiliar environments or toward odd or unfamiliar objects such as statues, animated objects (e.g., race car), and large or oddly-shaped items (e.g., beach ball, umbrella, rocking horse) are commonly used to assess working suitability (53). Novel object tests differ from visual startle tests in that they are not intended to elicit a startle response, but rather measure the dogs' willingness to approach an ambiguous object. Indeed, King et al. (59) found that responses to novelty and responses to startles appeared to measure two different types of fearful behavior. When encountering an object that is novel, a dog must make an appraisal as to whether the object is benign or potentially dangerous. In this sense, novel object tests may be similar to the cognitive bias test which assesses animals' responses toward ambiguity and serves as a measure of positive or negative expectancy (72). In this task, approaching an ambiguous stimulus in the same manner as a stimulus with a positive association (i.e., previously rewarded) is indicative of a positive expectancy, whereas approaching in a manner similar to a stimulus with a negative association (i.e., previously unrewarded or punished) is indicative of a negative expectancy. In dogs, negative cognitive bias is associated with negative emotional states (73). Thus, how a dog approaches a novel object may be indicative of its bias toward expecting positive or negative outcomes.

It is likely that the appraisal of the object as a threat or not involves perception of the object as a predator. Thus, novel object tests that use animal statues or objects with facial features (e.g., large eyes) likely tap into predator avoidance responses. Avoidance or defensiveness toward novelty is a well-established fear response in many animals and is of clear adaptive value, and novel objects with intense characteristics, movement, and unpredictability are likely to elicit predator-related fear responses (59). While predator-related fear is considered an innate and adaptive response across mammals, it must necessarily be reduced in detection dogs that will be expected to work effectively regardless of what is encountered. In this sense, the increased behavioral requirements for EDDs may result in selecting dogs with reduced self-preservation behavior; the consequences of which may need to be taken into consideration for the safe operation of an EDD.

Puppy Tests

The ability to test a puppy in order to reliably predict future behavior has been called a "holy grail' of dog research (13). For working dog programs, the ability to make decisions about breeding, training, or career paths as early as possible would significantly reduce the amount of time and costs involved. Unfortunately, research on the reliability of puppy testing is mixed, with many studies failing to find any consistent relationships between puppy and adult behavior [56, (37)]. In general, adult behavior is difficult to predict in puppyhood due to the continued interactions between neurological, environmental, and genetic influences across development (56). However, a few studies have found that certain aspects of behavior can be predicted in puppyhood (25, 54, 60). Behaviors exhibited during puppy tests that hold predictive validity are likely to be strongly genetically based, as such stability would indicate strong resistance to change due to environmental influence and maturation (13). For example, fearfulness is considered a core personality trait in dogs that can be identified early in working dogs and is relatively stable across development (52).

In general, the predictive power of puppy tests increase with age (53) and may be able to more accurately predict behavior when combined with other measures compared to a single measure (60). Aspects of the early environment such as proper socialization have been shown to be strongly associated with adult behavior (74, 75). Thus, rather than testing puppies for specific reactions, it may be more informative to know what kind of experiences and exposure the puppy received during early development. Further, tracking a puppy's behavior across development rather than testing at one time may be valuable. For example, McGarrity et al. (26) found that a construct that combined multiple measures including the ability to focus on a location where an object had been hidden, carrying a toy without handler engagement, grip of a toy during a game of tug, and performance during a search task was predictive of selection as an EDD, but only when assessed over time. Specifically, the overall score was not predictive but rather an increase in the score across the first year of life was, suggesting that tracking improvements in performance across development may be more informative than focusing on a single time point (26). Routinely evaluating behavior may also be useful for identifying deficiencies in order to develop targeted training interventions and to monitor progress in response to changes in training and breeding practices (56).

Types of Assessments

Behavioral Assessments

The majority of working dog assessments, and the majority of those reviewed above, utilize traditional behavioral assessments consisting of a series of sub-tests aimed to measures aspects of temperament. However, a lack of consensus and standardization in regard to terminology, test quality, stimuli used, and variables measured makes the ability to make meaningful comparisons across groups and extracting results challenging (64).

Behavioral tests are often scored using subjective ratings methods in which an observer makes a judgement of global behavior on a Likert-type scale to indicate the degree of a behavior, e.g., in terms of its frequency, desirability, or strength (26). Alternatively, coding methods involve quantifying specific instances of behaviors that occur during the test (e.g., barking, cowering, jumping) which may be more objective than rating methods (26). While some studies have found consistencies across rating and coding methods, the optimal method may depend on the situation (26, 27).

Questionnaires

A number of questionnaires and handler surveys have been developed as alternatives or adjuncts to behavioral testing, which is time consuming, less cost effective, and may not accurately capture behavior that occurs outside of the test situation (56). Handler reports of dog behavior using the C-BARQ has been successfully used to predict working dog suitability (41) and guide dog training success (56, 76, 77). The Positive and Negative Affect Scale (PANAS), developed as a way to measure sensitivity to rewards and punishers, and the Dog Impulsivity Assessment Scale (DIAS) (78), assessing factors related to impulsivity in dogs such as behavioral regulation, aggression, and responsiveness, have both been successfully used to identify important characteristics in police dogs (34). An advantage of a survey such as this is that data can be collected without having to expose dogs to potentially stressful situations, which could subsequently affect future behavior (79). A disadvantage of surveys is that accurate reporting requires sufficient knowledge of the dog (e.g., its trainer or handler), which is not always feasible when assessing large numbers of previously unseen dogs in a mass procurement activity, in which even the vendor of those dogs may not have prior experience with those dogs, and also requires honesty and accuracy by the person reporting (8). Though survey reports may not be the most feasible selection tool alone, they may serve as a valuable measure for validating existing behavioral tests.

Cognitive Measures

Recently, researchers have applied measures of cognitive ability (i.e., problem-solving and information processing) to the assessment of working dog suitability. Many aspects of cognition are likely involved in working dog tasks including memory, behavioral flexibility, mental representation, self-control, and communication, and therefore may represent quantifiable and objective metrics for evaluating individual differences in detection dog success (29, 80). For example, as discussed above, measures of socio-cognitive abilities have indicated that social communicative behaviors are related to detection dog trainability (29, 30, 39). Other studies have found that measures of non-social cognition are predictive of working dog performance, such as inhibitory control, problem-solving, and short-term memory (29, 66, 81, 82). Thus, assessments of cognitive abilities identified as contributing to working dog success show promise as valuable complementary measures to traditional evaluations for improving the selection process (29).

CONCLUSIONS

In this review, we highlight the behavioral characteristics that appear to be critical in the selection of EDDs. Many of these characteristics are similar to those desired in other types of working dogs such as search and rescue, conservation, protection/patrol, and even assistance and guide dogs, which also require high levels of motivation, trainability, and the ability to work in potentially stressful environments. However, we suggest that there is an ideal constellation of characteristics for EDDs that is somewhat unique, which is likely also the case for other types of working dogs. The constellation of desirable characteristics for explosive detection is defined by the degree and suitable balance of the expression of particular characteristics, and the optimal balance of this expression will vary based on the specialized explosive detection application (**Figure 1**). Specialization is rapidly becoming more normative than the general or "standard" EDD, which is indicative of technological advancement of explosive detection dog capabilities.

The behavioral characteristics examined in this review may be divided into three broad categories, detection characteristics, trainability, and environmental soundness. Detection-related characteristics include odor-guided behavior such as the innate propensity for hunting and the type of search technique. Trainability comprises multiple, likely overlapping traits that will influence a dogs' ability to learn, such as reward value and sociability. Environmental-related characteristics include emotional reactivity subcategories of arousal and fearfulness. These are likely not unitary categories and there are probably significant interactions within and between performance and environmental characteristics.

Explosive odors do not have any particular biological relevance to dogs and thus are not of any intrinsic interest for them to detect. Therefore, it is imperative that candidate explosives detection dogs exhibit a strong combination of odor-guided behavior and very high reward value in order to build the contingency between detecting explosive odors and obtaining a desired reward. Although food is inherently reinforcing and is used by some organizations for reinforcing explosive detection performance, managing satiety and diet complicates its use and the opportunity to play with a toy and/or the handler is the predominant reinforcer used in operating explosives detection dogs. Furthermore, it would appear that dogs for which delivery of toys is a highly effective reinforcer have concomitant characteristics, such as a propensity to hunt with their nose, important to detection dog performance. The degree of possession of a toy appears to be a useful metric of the potential of a dog for being successfully trained and employed as an EDD, but the nature of such possession (e.g., "independent" vs. "dominant") may reflect multiple traits that have differential predictive value. Once the contingency between detecting explosives odors and obtaining a desired reward is established, it is desirable in EDDs that such conditioning makes the opportunity to engage in searching a preferred activity.

A dog's willingness to stay engaged or focused on searching and vigilant for alerting to target odors in the midst of distracting stimuli is also an important metric in assessing its potential as an EDD. Resilience in searching in the absence of handler encouragement and prompts to work, which can lead to a dependence on handler cues that can increase false alerts, is a particularly important characteristic for explosive detection performance because of the relatively low rate of encountering targets.

It is important in EDDs that independent task engagement and lack of distractibility does not come at the expense of trainability. For example, although it is essential that PBIED detection dogs remain engaged in searching around people, it would appear that social sensitivity is an important characteristic for following handler directions (39) and being successfully trained as an EDD. Thus, a balance of underlying traits, such as social sensitivity and co-operability, which seem to contribute to trainability, with that of independent task engagement, is needed for optimal EDD performance. Though the focus of this review is on dog characteristics, it is important to consider the impact the human side of the dog-human dyad can have on the performance and success of a detection dog. For example, search errors are often due to handler error, such as handler-induced false alerts (38, 83) or missed targets due to the handler interfering with the dogs' ability to adequately search an area (84). Handler stress can also influence dogs' search performance (85, 86), and working with an unfamiliar handler can be stressful for dogs leading to reduced search accuracy (87). Other characteristics related to the experience and skill of the handler, such as timing of reinforcement delivery, consistency in their interaction with the dog, and training methods will also be critical to the efficiency of training and ultimately the success of the dog, though research in this area is lacking (88). Future studies are needed to identify the attributed of effective handlers and optimal methods for their selection and education.

Equally important and arguably of greater difficulty to define and assess than the performance-related characteristics summarized above is emotional reactivity, often classified as "environmental" characteristics. The difficulty in defining and assessing emotional reactivity in relation to the potential of a dog to perform explosive detection is in large part due to high levels of energy or activity associated with arousal being both a positive, necessary, attribute to performing detection work and such levels of arousal carrying along with it increased generalized reactivity to stimuli that interfere with performance. High levels of general

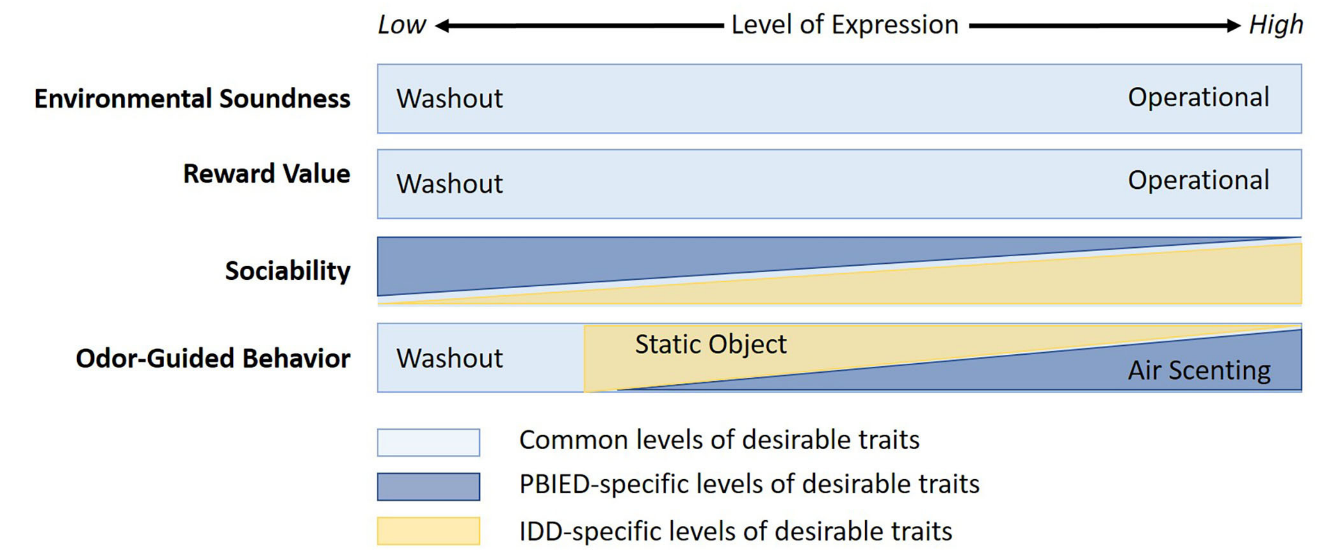


FIGURE 1 | Theoretical representation of the critical behavioral characteristics and optimal levels of expression necessary for explosives detection dogs, where Operational refers to dogs deployed in the field and Washout refers to dogs unsuitable for such roles.

arousal may also interfere with learning and performing a task. Low levels of general arousal are associated with low levels of activity and motivation to perform a task. Thus, it is not a balance of high and low arousal that appears to be ideal but rather arousal organized about and directed toward engaging in searching and detecting odor targets. Anticipatory pre-work and off-duty high arousal detract from performance as does arousal associated with reactivity to stimuli that may manifest as fearfulness of both social (e.g., negative – fearful reactivity to novel people) and non-social (e.g., negative – fearful reactivity to loud sudden noises) stimuli. Such fearfulness is decidedly incompatible with performing explosive detection in the most frequent context for such work, the modern urban terrain.

In our experience, evaluating environmental soundness in highly motivated dogs is further complicated by such dogs often lacking awareness to stimuli in the environment when they are engaged in searching. This effect may be the result of inattentional blindness, which refers to the failure to notice unexpected stimuli when engaged in a task demanding high levels of attention (89). Although inattentional blindness has not yet been explored in dogs, a recent study found that horses trained to expect a reward in a particular location show a reduced startle response to a novel stimulus compared to those that were not given such experience (90). Therefore, we suggest it more useful to examine emotional reactivity while dogs are not and have not immediately been engaged in searching, and to conduct such evaluations in an area not associated with expectation of reward. This is more difficult than it may first seem because, for dogs engaged in training, the context of transportation, location, and presence of trainers/handlers all tend to predict the opportunity to engage in searching, find target odor, and obtain their reward. Although environmental interest outside the context of searching may at first seem to *not* be critical, in operational deployment there is considerable downtime, which becomes readily discernable to the dog, between searches in operational venues when emotional reactivity to stimuli may occur that sensitizes the dog to those stimuli such that it interferes with performance. Future research is needed to examine such contextual factors that may influence performance during behavioral assessments as well as the effectiveness of environmental socialization.

In selectively breeding dogs for detection, performance characteristics appear to be more readily enhanced by selective breeding than environmental characteristics, with lack of environmental soundness being the predominant reason for failure of dogs to succeed in explosives detection (16). Anecdotal reports in our contact with the explosive detection dog industry suggest that, particularly at the extreme of the performance continuum such as PBIED and SSD (i.e., off-lead IED detection) specializations, the normative reason for failure are emotional reactivity issues. Although this may be the result of differences between performance and environmental characteristics' sensitivity to selective pressure, we posit that it is much more likely the result of the difficultly in defining and disentangling positive and negative aspects of arousal with current assessment techniques.

Understanding how particular behavioral characteristics are related to explosive detection performance is key to a technology for assessing dogs for such service. The foundation of such an assessment technology will be the extension of the important research efforts examined in this review evaluating this relationship. Unlocking the potential of such assessments will further depend on a deeper look into the phenomenology of the characteristics themselves in order to better define, disentangle in several cases, and measure those characteristics. Development of a standardized and accurate assessment technology that can be applied across programs will be critical to increasing the supply of suitable detection dogs and improving detection technology overall.

Puppy assessments predictive of potential explosive detection performance would allow for significant efficiencies in managing resources and optimizing the value of every dog bred for working purposes by directing it toward a successful career path. Research suggests that some characteristics, such as fearfulness, observed in puppies are relatively stable across time suggesting strong genetic determinants (52). Characteristics that are emergent and likely significantly influenced by experience may be better assessed by multiple observations across time and trends, such as stability or improvement, and may be more predictive than any single time-point observation. Our review suggests that there is a useful historical base and some momentum for more targeted research in early assessment activities that promises to advance early prediction of the potential of dogs for explosive detection tasks.

Current procurement of candidate dogs for explosive detection usually utilizes behavioral assessments based on traditional conceptualizations of working dog characteristics for which subjective ratings are assigned. Different organizations use different assessments, which in part logically reflects the parameters of the particular organization's explosive detection mission. However, there is also considerable variability in terminology, testing techniques, and subjective ratings making it difficult to make meaningful comparisons across assessments.

Validated questionnaires such as the C-BARQ and PANAS have been shown to predict, among other behavioral outcomes, working dog suitability. Questionnaires are unlikely, however, to replace direct immediate behavioral assessments for selection of candidate EDDs because they require that the respondent has historical knowledge of the dog's behavior and the impracticality of such a respondent's judgement being impartial or being perceived as such to the receiving/procuring party. Nonetheless, elements of these validated questionnaires in combination with direct observation of behavior might be combined to potentially enhance the predictive value of assessments for selecting candidate explosive detection dogs.

Finally, recent research suggests that the incorporation of measures of problem solving, information processing, memory, and inhibitory control adapted from the cognitive sciences hold significant promise of providing more objective and quantifiable metrics indicative of suitability for explosive detection tasks. Another potential advantage of cognitive measures is that they may access underlying, immutable characteristics related to performance.

There is a need of a selection technology to advance the production and utilization of explosive detection dogs. Our review suggests that this technology is dependent upon ongoing efforts to further refine the identification, definition, and measurement of the constellation of characteristics important to different specializations of explosive detection tasks. Standardization of traditional working dog assessment techniques, incorporation of elements of proven behavioral questionnaires, and continuing evolution of the use of cognitive measures promise to advance selection technology. Finally, accurate identification and validation of selection measures will rely on continual feedback on dogs' operational performance post-selection.

AUTHOR CONTRIBUTIONS

LL and LW wrote the first draft of the manuscript. BR developed concepts and contributed content. MS created **Figure 1**. SK, MS, PH, and CA contributed to the development of the paper and revised content. All authors read and approved the final draft.

FUNDING

Funding for preparation of this review was provided by the Auburn University College of Veterinary Medicine Canine Performance Sciences Program through the Richard G. and Dorothy Metcalf Endowment, Allen Kalter and Dr. Chris Lezottee Health Excellence Fund, and James M. Hoskins Endowment.

ACKNOWLEDGMENTS

We thank the canine training professionals of Canine Performance Sciences and other programs who have generously shared their experience and expertise with us over many years, which has been invaluable in our understanding of detection dog behavioral selection.

REFERENCES

1. Leighton EA, Hare E, Waggoner LP, Otto CM, Thomas S. A solution for the shortage of detection dogs: a detector dog center of excellence and a cooperative breeding program. *Front Vet Sci*. (2018) 5:284. doi: 10.3389/fvets.2018.00284
2. FAA and Reauthorization Act of 2018. (2018). Available online at: https://www.congress.gov/bill/115th-congress/house-bill/302/text#toc (accessed February 1, 2020).
3. Jamieson LTJ, Baxter GS, Murray PJ. Identifying suitable detection dogs. *Appl Animal Behav Sci.* (2017) 195:1–120. doi: 10.1016/j.applanim.2017.06.010
4. Rooney NJ, Bradshaw JWS. Breed and sex differences in the behavioural attributes of specialist search dogs - A questionnaire survey of trainers and handlers. *Appl Animal Behav Sci.* (2004) 86:123–35. doi: 10.1016/j.applanim.2003.12.007
5. Graham LT, Gosling SD. Temperament and personality in working dogs. In: Helton WS, editor. *Canine Ergonomics: The Science of Working Dogs*. Boca Raton, FL: CRC Press/Taylor & Francis (2009). p. 63–81.
6. Sinn DL, Gosling SD, Hilliard S. Personality and performance in military working dogs: Reliability and predictive validity of behavioral tests. *Appl Animal Behav Sci.* (2010) 127:51–65. doi: 10.1016/j.applanim.2010.08.007
7. Cobb M, Branson N, McGreevy P, Lill A, Bennett P. The advent of canine performance science: offering a sustainable future for working dogs. *Behav Proc.* (2014) 110:96–104. doi: 10.1016/j.beproc.2014.10.012
8. Brady K, Cracknell N, Zulch H, Mills DS. A systematic review of the reliability and validity of behavioural tests used to assess behavioural characteristics important in working dogs. *Front Vet Sci.* (2018) 5:103. doi: 10.3389/fvets.2018.00103
9. Beebe SC, Howell TJ, Bennett PC. Using Scent detection dogs in conservation settings: a review of scientific literature regarding their selection. *Front Vet Sci.* (2016) 3:96. doi: 10.3389/fvets.2016.00096

10. Rocznik D, Sinn DL, Thomas S, Gosling SD. Criterion analysis and content validity for standardized behavioral tests in a detector-dog breeding program. *J Forensic Sci.* (2015) 60:S213–21. doi: 10.1111/1556-4029.12626
11. Rooney NJ, Bradshaw JWS, Almey H. Attributes of specialist search dogs–a questionnaire survey of UK dog handlers and trainers. *J Forensic Sci.* (2004) 49:300–6. doi: 10.1520/JFS2003048
12. Polgár Z, Kinnunen M, Újváry D, Miklósi Á, Gácsi M. A test of canine olfactory capacity: comparing various dog breeds and wolves in a natural detection task. *PLoS ONE.* (2016) 11:e0154087. doi: 10.1371/journal.pone.0154087
13. Miklási A. *Dog Behaviour, Evolution, and Cognition* 2nd edn. Oxford: Oxford Univeristy Press (2015).
14. Scientific Working Group on Dog and Orthogonal detector Guidelines (SWG- DOG). *SC1-abcdefghijk-Terminology*. (2011). Available online at: http://swgdog. fiu.edu/approved-guidelines; http://swgdog.fiu.edu/approved-guidelines/sc1_terminology_abcdefghijk.pdf (accessed January 21, 2019).
15. Jones KE, Dashfield K, Downend AB, Otto CM. Search-and-rescue dogs: An overview for veterinarians. *J Am Vet Med Assoc.* (2004) 225:854–60. doi: 10.2460/javma.2004.225.854
16. Lazarowski L, Haney P, Brock J, Fischer T, Rogers B, Angle C, et al. Investigation of the behavioral characteristics of dogs purpose-bred and prepared to perform Vapor Wake® detection of person-borne explosives. *Front Vet Sci.* (2018) 5:50. doi: 10.3389/fvets.2018.00050
17. Lord K, Schneider RA, Coppinger R. Evolution of working dogs. In: Serpell J, editor. *The Domestic Dog: Its evolution, Behavior, and Interactions With People.* Cambridge University Press. (2016). p. 42-66.
18. Spady T, Ostrander E. Canine behavioral genetics: pointing out the phenotypes and herding up the genes. *Am J Human Genet.* (2008) 82:10–8. doi: 10.1016/j.ajhg.2007.12.001
19. Brownell DA, Marsolais M. The Brownell-Marsolais scale: a proposal for the qualitative evaluation of SAR/disaster K9 candidates. *Adv Rescue Technol.* (2002) 5:57–67. Available online at: www.operationtakemehome.org/sar/Canine%20Downloads/Forms/AllItems.aspx
20. Oxley J, Waggoner L. Detection of explosives by dogs. In: Marshell M, Oxley J. editors. *Aspects of Explosives Detection*. Amsterdam: Elsevier Science (2009). p. 27–40.
21. Udell MAR, Ewald M, Dorey NR, Wynne CDL. Exploring breed differences in dogs (*Canis familiaris*): does exaggeration or inhibition of predatory response predict performance on human-guided tasks? *Animal Behav.* (2014) 89:99–105. doi: 10.1016/j.anbehav.2013.12.012
22. Mehrkam LR, Hall NJ, Haitz C, Wynne CDL. The influence of breed and environmental factors on social and solitary play in dogs (Canis lupus familiaris). *Learn Behav.* (2017) 45:367–77. doi: 10.3758/s13420-017-0283-0
23. Burghardt GM, Albright JD, Davis KM. Motivation, development and object play: comparative perspectives with lessons from dogs. *Behaviour.* (2016) 153:767–93. doi: 10.1163/1568539X-00003378
24. Slabbert J, Odendaal JS. Early prediction of adult police dog efficiency—a longitudinal study. *Appl Animal Behav Sci.* (1999) 64:269–88. doi: 10.1016/S0168-1591(99)00038-6
25. Svobodová I, Vápeník P, Pinc L, Bartoš L. Testing German shepherd puppies to assess their chances of certification. *Appl Animal Behav Sci.* (2008) 113:139–49. doi: 10.1016/j.applanim.2007.09.010
26. McGarrity ME, Sinn DL, Thomas SG, Marti CN, Gosling SD. Comparing the predictive validity of behavioral codings and behavioral ratings in a working-dog breeding program. *Appl Animal Behav Sci.* (2016) 179:82–94. doi: 10.1016/j.applanim.2016.03.013
27. Wilsson E, Sinn DL. Are there differences between bahaviral measurement methods? A comparison of predictive validity of two rating methods in a working dog program. *Appl Animal Behav Sci.* (2012) 141:158–72. doi: 10.1016/j.applanim.2012.08.012
28. Fiset S, Beaulieu C, Landry F. Duration of dogs' (*Canis familiaris*) working memory in search for disappearing objects. *Animal Cognit.* (2003) 6:1–10. doi: 10.1007/s10071-002-0157-4
29. MacLean E, Hare B. Enhanced selection of assistance and explosive detection dogs using cognitive measures. *Front Vet Sci.* (2018) 5:236. doi: 10.3389/fvets.2018.00236
30. Lazarowski L, Strassberg LR, Waggoner LP, Katz JS. Persistence and human-directed behavior in detection dogs: ontogenetic development and relationships to working dog success. *Appl Animal Behav Sci.* (2019) 220:104860. doi: 10.1016/j.applanim.2019.104860
31. Hall NJ. Persistence and resistance to extinction in the domestic dog: Basic research and applications to canine training. *Behav Proces.* (2017) 141:67–74. doi: 10.1016/j.beproc.2017.04.001
32. Dalal S, Hall NJ. Behavioral persistence is associated with poorer olfactory discrimination learning in domestic dogs. *Behav Proces.* (2019) 162:64–71. doi: 10.1016/j.beproc.2019.01.010
33. Maejima M, Inoue-Murayama M, Tonosaki K, Matsuura N, Kato S, Saito Y, et al. Traits and genotypes may predict the successful training of drug detection dogs. *Appl Animal Behav Sci.* (2007) 107:287–98. doi: 10.1016/j.applanim.2006.10.005
34. Brady K, Cracknell N, Zulch H, Mills DS. Factors associated with long-term success in working police dogs. *Appl Animal Behav Scie.* (2018) 207:67–72. doi: 10.1016/j.applanim.2018.07.003
35. Gácsi M, McGreevy P, Kara E, Miklósi A. Effects of selection for cooperation and attention in dogs. *Behav Brain Funct BBF.* (2009) 5:31. doi: 10.1186/1744-9081-5-31
36. Wilsson E, Sundgren PE. The use of a behaviour test for selection of dogs for service and breeding. II Heritability for tested parameters and effect of selection based on service dog characteristics. *Appl Animal Behav Sci.* (1997) 54:235–41. doi: 10.1016/S0168-1591(96)01175-6
37. Wilsson E, Sundgren PE. The use of a behaviour test for the selection of dogs for service and breeding, I: method of testing and evaluating test results in the adult dog, demands on different kinds of service dogs, sex and breed differences. *Appl Animal Behav Sci.* (1997) 53:279–95. doi: 10.1016/S0168-1591(96)01174-4
38. Lit L, Schweitzer JB, Oberbauer AM. Handler beliefs affect scent detection dog outcomes. *Animal Cognit.* (2011) 14:387–94. doi: 10.1007/s10071-010-0373-2
39. Lazarowski L, Rogers B, Waggoner LP, Katz JS. When the nose knows: ontogenetic changes in detection dogs' (*Canis familiaris*) responsiveness to social and olfactory cues. *Animal Behav.* (2019) 153:61–8. doi: 10.1016/j.anbehav.2019.05.002
40. MacLean EL, Snyder-Mackler N, VonHoldt BM, Serpell JA. Highly heritable and functionally relevant breed differences in dog behavior. *BioRxiv.* (2019). doi: 10.1101/509315
41. Foyer P, Bjällerhag N, Wilsson E, Jensen P. Behaviour and experiences of dogs during the first year of life predict the outcome in a later temperament test. *Appl Animal Behav Sci.* (2014) 155:93–100. doi: 10.1016/j.applanim.2014.03.006
42. Cao X, Irwin DM, Liu YH, Cheng LG, Wang L, Wang GD, et al. Balancing selection on CDH2 may be related to the behavioral features of the Belgian malinois. *PLoS ONE.* (2014) 9:e110075. doi: 10.1371/journal.pone.0110075
43. Troisi CA, Mills DS, Wilkinson A, Zulch HE. Behavioral and cognitive factors that affect the success of scent detection dogs. *Compar Cognit Behav Rev.* (2019) 14:51–76. doi: 10.3819/CCBR.2019.140007
44. Yerkes RM, Dodson JD. The relation of strength of stimulus to rapidity of habit-formation. *J Comp Neurol Psychol.* (1908) 18:459–82. doi: 10.1037/h0073415
45. Bray EE, MacLean EL, Hare BA. Increasing arousal enhances inhibitory control in calm but not excitable dogs. *Animal Cognit.* (2015) 18:1317–29. doi: 10.1007/s10071-015-0901-1
46. Gazit I, Terkel J. Explosives detection by sniffer dogs following strenuous physical activity. *Appl Animal Behav Sci.* (2003) 81:149–61. doi: 10.1016/S0168-1591(02)00274-5
47. Tomkins LM, Thomson PC, McGreevy PD. Behavioral and physiological predictors of guide dog success. *J Vet Behav Clin App Res.* (2011) 6:178–87. doi: 10.1016/j.jveb.2010.12.002
48. Iotchev I, Kis A, Róbert Bódizs R, van Luijtelaar G, Kubinyi E. EEG transients in the sigma range during non-REM sleep predict learning in dogs. *Sci Rep.* (2017) 7:12936. doi: 10.1038/s41598-017-13278-3
49. Kis A, Szakadát S, Gácsi M, Kovács E, Simor P, Török C, et al. The interrelated effect of sleep and learning in dogs (Canis familiaris). An EEG and behavioural study. *Nat Sci Rep.* (2017) 7:41873. doi: 10.1038/srep41873
50. Cox E, Lazarowski L, Krichbaum S, Waggoner LP, Katz JS. *Kennel Behavior and Enrichment in Working Dogs.* Phoenix, AZ: Canine Science Conference (2019).
51. Sherman BL, Gruen ME, Case BC, Foster ML, Fish RE, Lazarowski L, et al. A test for the evaluation of emotional reactivity in labrador retrievers used for explosives detection. *J Vet Behav.* (2015) 10:94–102. doi: 10.1016/j.jveb.2014.12.007

52. Dollion N, Paulus A, Champagne N, St-Pierre N, St-Pierre É, Trudel M, et al. Fear/reactivity in working dogs: an analysis of 37 years of behavioural data from the mira foundation's future service dogs. *Appl Animal Behav Sci.* (2019) 221:104864. doi: 10.1016/j.applanim.2019.104864
53. Goddard ME, Beilharz RG. A factor analysis of fearfulness in potential guide dogs. *Appl Animal Behav Sci.* (1984) 12:253–65. doi: 10.1016/0168-1591(84)90118-7
54. Asher L, Blythe S, Roberts R, Toothill L, Craigon PJ, Evans KM, et al. A standardized behavior test for potential guide dog puppies: Methods and association with subsequent success in guide dog training. *J Vet Behav.* (2013) 8:431–8. doi: 10.1016/j.jveb.2013.08.004
55. Beerda B, Schilder MBH, Van Hooff JARAM, De Vries HW, Mol JA. Behavioural, saliva cortisol and heart rate responses to different types of stimuli in dogs. *Appl Animal Behav Sci.* (1998) 58:365–81. doi: 10.1016/S0168-1591(97)00145-7
56. Harvey ND, Craigon PJ, Blythe SA, England GCW, Asher L. An evidence-based decision assistance model for predicting training outcome in juvenile guide dogs. *PLoS ONE.* 12:e0174261. doi: 10.1371/journal.pone.0174261
57. Svartberg K. Shyness-boldness predicts performance in working dogs. *Appl Animal Behav Sci.* (2002) 79:157–74. doi: 10.1016/S0168-1591(02)00120-X
58. Foyer P, Svedberg AM, Nilsson E, Wilsson E, Faresjö Å, Jensen P. Behavior and cortisol responses of dogs evaluated in a standardized temperament test for military working dogs. *J Vet Behav.* (2016) 11:7–12. doi: 10.1016/j.jveb.2015.09.006
59. King T, Hemsworth PH, Coleman GJ. Fear of novel and startling stimuli in domestic dogs. *Appl Animal Behav Sci.* (2007) 104:71–84. doi: 10.1016/j.applanim.2006.03.021
60. Riemer S, Müller C, Virányi Z, Huber L, Range F. The predictive value of early behavioural assessments in pet dogs - A longitudinal study from neonates to adults. *PLoS ONE.* (2014) 9:e101237. doi: 10.1371/journal.pone.0101237
61. Fratkin JL, Sinn DL, Patall E, Gosling SD. Personality consistency in dogs: a meta-analysis. *PLoS ONE.* (2013) 8:e54907. doi: 10.1371/journal.pone.0054907
62. Riemer S, Müller C, Virányi Z, Huber L, Range F. Individual and group level trajectories of behavioural development in border collies. *Appl Animal Behav Sci.* (2016) 180:78–86. doi: 10.1016/j.applanim.2016.04.021
63. Serpell JA, Hsu Y. Development and validation of a novel method for evaluating behavior and temperament in guide dogs. *Appl Animal Behav Sci.* (2001) 72:347–64. doi: 10.1016/S0168-1591(00)00210-0
64. Diederich C, Giffroy JM. Behavioural testing in dogs: A review of methodology in search for standardisation. *Appl Animal Behav Sci.* (2006) 97:51–72. doi: 10.1016/j.applanim.2005.11.018
65. Stellato AC, Flint HE, Widowski TM, Serpell JA, Niel L. Assessment of fear-related behaviours displayed by companion dogs (*Canis familiaris*) in response to social and non-social stimuli. *Appl Animal Behav Sci.* (2017) 188:84–90. doi: 10.1016/j.applanim.2016.12.007
66. Bray EE, Sammel MD, Seyfarth RM, Serpell JA, Cheney DL. Temperament and problem solving in a population of adolescent guide dogs. *Animal Cognit.* (2017) 20:923–39. doi: 10.1007/s10071-017-1112-8
67. Geyer MA, Braff DL, Swerdlow NR. Startle-response measures of information processing in animals: relevance to schizophrenia. In: Haug M, Whalen RE editors. *Animal Models of Human Emotion and Cognition.* American Psychological Association (1999), 103–16. doi: 10.1037/10335-007
68. Tiira K, Sulkama S, Lohi H. Prevalence, comorbidity, and behavioral variation in canine anxiety. *J Vet Beh.* (2016) 16:36–44. doi: 10.1016/j.jveb.2016.06.008
69. Batt LS, Batt MS, Baguley JA, McGreevy PD. Factors associated with success in guide dog training. *J Vet Beh.* (2008) 3:143–51. doi: 10.1016/j.jveb.2008.04.003
70. Chaloupková H, Svobodová I, Vápeník P, Bartoš L. Increased resistance to sudden noise by audio stimulation during early ontogeny in German shepherd puppies. *PLoS ONE.* 13:e0196553. doi: 10.1371/journal.pone.0196553
71. Kim J, Williams FJ, Dreger DL, Plassais J, Davis BW, Parker HG, et al. Genetic selection of athletic success in sport-hunting dogs. *Proc Natl Acad Sci USA.* (2018) 115:E7212–21. doi: 10.1073/pnas.1800455115
72. Mendl M, Burman OHP, Parker RMA, Paul ES. Cognitive bias as an indicator of animal emotion and welfare: emerging evidence and underlying mechanisms. *Appl Animal Behav Sci.* (2009) 118:161–81. doi: 10.1016/j.applanim.2009.02.023
73. Polgár Z, Blackwell EJ, Rooney NJ. Assessing the welfare of kennelled dogs—A review of animal-based measures. *Appl Animal Behav Sci.* (2019)213:1–13. doi: 10.1016/j.applanim.2019.02.013
74. Puurunen J, Hakanen E, Salonen MK, Mikkola S, Sulkama S, Araujo C, et al. Inadequate socialisation, inactivity, and urban living environment are associated with social fearfulness in pet dogs. *Sci Rep.* (2019) 10:3527. doi: 10.1038/s41598-020-60546-w
75. Vaterlaws-Whiteside H, Hartmann A. Improving puppy behavior using a new standardized socialization program. *Appl Animal Behav Sci.* (2017) 197:55–61. doi: 10.1016/j.applanim.2017.08.003
76. Duffy DL, Serpell JA. Predictive validity of a method for evaluating temperament in young guide and service dogs. *Appl Animal Behav Sci.* (2012) 138:99–109. doi: 10.1016/j.applanim.2012.02.011
77. Bray EE, Levy KM, Kennedy BS, Duffy DL, Serpell JA, MacLean EL. Predictive models of assistance dog training outcomes using the canine behavioral assessment and research questionnaire and a standardized temperament evaluation. *Front Vet Sci.* (2019) 6:49. doi: 10.3389/fvets.2019.00049
78. Wright HF, Mills DS, Pollux PMJ. Development and validation of a psychometric tool for assessing impulsivity in the domestic dog (*Canis familiaris*). *Int J Compar Psychol.* (2011) 24:210–25. Retrieved from: https://escholarship.org/uc/item/7pb1j56q
79. Rooney NJ, Clark CCA, Casey RA. Minimizing fear and anxiety in working dogs: A review. *J Vet Behav.* (2016) 16:53–64. doi: 10.1016/j.jveb.2016.11.001
80. Lit L. Evaluating learning tasks commonly applied in detection dog training. In: Helton WS, editor, *Canine Ergonomics: The Science of Working Dogs.* Boca Raton, FL: CRC Press (2009). p. 99–114. doi: 10.1201/9781420079920.ch6
81. Lazarowski L, Krichbaum S, Waggoner LP, Katz JS. The development of problem-solving abilities in a population of candidate detection dogs (*Canis familiaris*). *Animal Cognit.* (2020) 23:755–68. doi: 10.1007/s10071-020-01387-y
82. Tiira K, Tikkanen A, Vainio O. Inhibitory control – important trait for explosive detection performance in police dogs? *Appl Animal Behav Sci.* (2020) 224:104942. doi: 10.1016/j.applanim.2020.104942
83. Curran AM, Prada PA, Furton KG. Canine human scent identifications with post-blast debris collected from improvised explosive devices. *Forensic Sci Int.* (2010) 199:103–8. doi: 10.1016/j.forsciint.2010.03.021
84. Engeman RM, Vice DS, York D, Gruver KS. Sustained evaluation of the effectiveness of detector dogs for locating brown tree snakes in cargo outbound from Guam. *Int Biodeteri Biodegrad.* (2002) 49:101–6. doi: 10.1016/S0964-8305(01)00109-3
85. Jezierski T, Adamkiewicz E, Walczak M, Sobczyńska M, Górecka-Bruzda A, Ensminger J, et al. Efficacy of drug detection by fully-trained police dogs varies by breed, training level, type of drug and search environment. *Forensic Sci Int.* (2014) 237:112–8. doi: 10.1016/j.forsciint.2014.01.013
86. Zubedat S, Aga-Mizrachi S, Cymerblit-Sabba A, Shwartz J, Leon JF, Rozen S, et al. Human–animal interface: the effects of handler's stress on the performance of canines in an explosive detection task. *Appl Animal Behav Sci.* (2014) 158:69–75. doi: 10.1016/j.applanim.2014.05.004
87. Jamieson LTJ, Baxter GS, Murray PJ. You are not my handler! Impact of changing handlers on dogs' behaviours and detection performance. *Animals.* (2018) 8:176. doi: 10.3390/ani8100176
88. McGreevy P, Starling M, Payne E, Bennett P. Defining and measuring dogmanship: A new multidisciplinary science to improve understanding of human–dog interactions. *Vet J.* (2017) 229:1–5. doi: 10.1016/j.tvjl.2017.10.015
89. Mack A, and Rock I. *Inattentional blindness: perception without attention.* Cambridge, MA: MIT Press (1998).
90. Gabor V, Wall S, Gerken M, Brinkmann L. Does inattentional blindness exist in horses (*Equus caballus*)? *Appl Animal Behav Sci.* (2019) 216:45–51. doi: 10.1016/j.applanim.2019.04.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 10 September 2020
doi: 10.3389/fvets.2020.00612

Search Dog Handlers Show Positive Bias When Scoring Their Own Dog's Performance

Corinna C. A. Clark[1], Nicola J. Sibbald[2] and Nicola J. Rooney[3]*

[1] Department of Life Sciences, University of Warwick, Coventry, United Kingdom, [2] Defence Science and Technology Laboratory, Salisbury, United Kingdom, [3] Animal Welfare and Behaviour Group, Bristol Veterinary School, University of Bristol, Bristol, United Kingdom

OPEN ACCESS

Edited by:
Nathaniel James Hall,
Texas Tech University, United States

Reviewed by:
Janice Lauren Baker,
Veterinary Tactical Group,
United States
Joseph Wakshlag,
Cornell University, United States
Helen Zulch,
Dogs Trust, United Kingdom

***Correspondence:**
Nicola J. Rooney
Nicola.Rooney@bristol.ac.uk

Specialty section:
This article was submitted to
Veterinary Humanities and Social Sciences,
a section of the journal
Frontiers in Veterinary Science

Received: 24 March 2020
Accepted: 29 July 2020
Published: 10 September 2020

Citation:
Clark CCA, Sibbald NJ and Rooney NJ (2020) Search Dog Handlers Show Positive Bias When Scoring Their Own Dog's Performance. Front. Vet. Sci. 7:612.
doi: 10.3389/fvets.2020.00612

Self-assessments of performance are commonly used in the human workplace, although compared to peer or supervisor ratings, they may be subject to positive biases or leniency. The use of subjective ratings scales in animal sciences is also common, although little consideration is usually given to possible rater bias. Dog handlers, work very closely and form strong relationships with their dogs and are also best placed to monitor dog performance since they often work in isolation. Previous work found ratings of search dog performance correlated well between experienced dog trainers, instructors, and scientists; but until now, there has been no investigation into ratings made by a dog's own handler. We compared handlers' subjective assessment of their own dog's search performance to scores given by other handlers and in a second study, to scores made by impartial raters. We found that handlers generally showed leniency; for example scoring their own dogs more favorably for Control (responsiveness to commands) and Strength of Indication. But the degree of bias varied with the trait being scored and between raters. Such differences may be attributable to greater desirability or importance of favorable scores for certain traits, or a lack of clarity of their precise meaning. Handlers may vary in susceptibility to bias due to differing levels of experience and the extent to which they view their dog's ability as dependent on their own. The exact causes require further investigation. We suggest working dog agencies provide rater-training to overcome leniency, improve reliability and validity, and to increase handler's motivation to provide accurate assessments. This study represents one of a series of steps to formulate robust, validated and evidence-based performance rating systems and has relevance to any situation where raters assess their own performance or others (particularly where they may have a vested interest in, or loyalty toward, the ratee).

Keywords: bias, rating, working dog, leniency, reliability, validity

INTRODUCTION

Search dog teams perform a vital role in law enforcement agencies, search and rescue teams and in the military, searching for targets as diverse as people, drugs, money, weapons, and explosives. Many consider them to be, if not the most effective method [e.g., (1)], the fastest and most versatile method of detecting explosives (2). Specialist medical detection dogs are also effective in aiding the control of chronic life-threatening conditions by alerting their owners to physiological changes, such as hypoglycaemia in patients with diabetes (3, 4). To maintain standards of performance and maximize search/detection capability, organizations often monitor, and record various aspects of daily operational performance. This monitoring process is essential to address short-term training needs and for longer-term planning and changes in policy. For example longitudinal data can help to formulate optimal rearing and training protocols, to assess the impact of changes in care or operational procedures (e.g., work rest patterns, climate acclimatization), as well as answering questions such as whether certain breeds are better suited to a particular type of work or environment. If critical decisions are to be based on such data, it is imperative that the collection is robust, validated and evidence-based. Hence we have embarked in a multi-stage process to develop an optimal data collection tool for military search dogs, which rates the most critical behavioral traits and presents them in the optimal way (5). One stage of development is described in this paper.

Due to the nature of the tasks being performed (e.g., military dogs working in combat zones and search and rescue dogs working in remote or unstable terrain) handlers frequently work singly and so the responsibility for measuring performance will largely fall to the handlers themselves. For performance monitoring to be effective in these circumstances, methods need to be practically feasible and provide timely, accurate, and reliable information. Suitable behavioral measures of search performance have been derived by Rooney et al. (6) and it has been demonstrated that subjective scales rating these characteristics can be an effective method of point sampling search performance in an experimental setting and monitoring performance longitudinally using trainers' ratings (7). Subjective ratings of this kind have the advantage of being relatively quick and easy, so not only can they be completed in the field but feedback on changes in performance can be instantaneous.

Subjective assessments can, however, be subject to rater error or bias [e.g., (8–11)]. There are many forms and sources of bias, but particularly relevant in this setting are biases where raters provide more positive scores than reflect real performance, because they are either rating themselves or a colleague/friend. Positive response biases to survey questions and subjective rating scales are commonly reported, although they can have differing motivations, being described using terms such as leniency (12, 13), acquiescence (14, 15) and satisficing (11, 16). In surveys, more than 50% of respondents frequently believe themselves to be "above average" in respect to whatever question is being asked (17–19). Self-report bias, or leniency, when raters assess and score their own performance in a task is well-documented [e.g., (19–21)]. Although dog handlers are not scoring their own performance, they are working closely with their dog as a single search team and so it is conceivable that they may be reluctant to give poor scores to their own dog, for fear that it will reflect badly on their own performance, or out of "loyalty" to the dog akin to the friendship bias or "own-group" bias seen in peer assessments (22, 23). Handlers may also be influenced by an a-priori expectation of the dog's capabilities based on previous experience (e.g., my dog would usually perform better than this, so I will give him the benefit of the doubt on this occasion), leading to more lenient marking. This is important if procedural decisions are to be made on the basis of handlers' performance ratings.

We examined potential leniency in a group of operational dog handlers, testing whether subjective assessment of their own search performance was more favorable by comparing the scores they gave to their own dog's search performance to scores given by other handlers (Study 1) and independent raters (Study 2). Initially, 12 arms and explosives search (AES) dog handlers were divided into pairs and asked to rate the performance of their own and their partner's dog. As it would not be practically feasible to carry out experiments in an operational environment and having previously demonstrated that observers can reliably rate dogs' ability from video recordings (7), we used video recordings of training searches. Handlers scored several performance measures relevant to AES dogs (7), as well as giving a score for Overall Ability. This allowed us to test whether handlers showed rank order consistency (i.e., best to worst performance) over a series of searches and whether scores given to their own searches were more favorable than those given by the other handler in the pair. As any differences in overall ratings between own and other scores could be attributable to bias in either party, in Study 2 we compared own and other handler scores to independent expert ratings to clarify which were more accurate (i.e., are closer to the true score), the expert ratings were assumed to be an unbiased reflection of actual performance. We hypothesized that handlers would rate their own searches differently, and that in general, score their own dog more favorably (or leniently) compared to other handers and experts' scores.

METHODS

Study 1

Subjects and Training Searches

Twelve trainee arms and explosives search (AES) dog handlers were recruited. Each had between 1 and 10 years (average 4 years) experience of handling dogs and all were in their final week of a 15-weeks training course with their AES dog. Each handler had been filmed (using a hand-held video camera with a wide-angle lens) performing training searches on ten occasions during the previous 7 weeks. A section of each search was selected which was clear to see and which together showed a wide variety of different levels of performance. Cropped sections varied from 6 to 16 min long (average = 12 min). The majority of searches included the dogs encountering an explosives training sample (104 searches with, 16 without), and the recording ended after the dog alerted to or missed the hide.

Video Observations

Each of the 12 handlers was randomly paired with another handler at the same stage of their training, but who had been in a different training group and hence they had rarely seen each other search. Handlers watched and rated the videos in these pairs, with observations split into two sessions on consecutive days to reduce fatigue. In each session, handlers watched five of their own and five of their partner's searches, so they scored

10 own and 10 other searches in total. Thus, each of 120 video clips were watched by two handlers; generating one "own" handler score and one "other" handler score per dog, per characteristic.

The order of searches was randomized, but alternated between the two handlers. Immediately prior to observations, pairs were advised that they should: observe the entire video before rating performance; base their ratings only upon what they had seen on the video; and try to use the full range of the rating scales if appropriate. An experimenter was present throughout to ensure that handlers did not talk to each other about their ratings. There was a pause after each film to allow handlers as much time as they needed to complete the ratings form and short breaks after the third, sixth and eighth films.

Performance Measures

Previous work had prioritized the most relevant dimensions for current AES performance (7), from which the following seven characteristics were chosen:

Control, or response to commands;
Motivation to search;
Stamina throughout the search;
Confidence in the environment;
Independence or ability to search without direction;
Distraction from searching;
Strength of (behavioral) **Indication** when the dog locates a hide.

Handlers rated the characteristics on a 1 to 5 scale: 1 = very low level of the characteristic; 2 = low; 3 = intermediate; 4 = high; 5 = very high, recording this on a pre-printed sheet (tick boxes). Beyond a brief instruction on the meaning of these terms, no further descriptors, or guidance for marking was given. The scales were explicitly not valenced (very low to very high, as opposed to very poor to very good) and handlers were instructed that their scores should reflect the amount of the trait present and not how well the dog was performing. However, in general, high scores for a particular trait would indicate a well-performing dog (e.g., Control and Motivation). Exceptions to this were Independence and Distraction: previous work has indicated that some handlers view ideal levels of Independence as being a score of 3 or 4, rather than 5 (6); positive bias in Distraction would be evidenced by low scores, as high scores indicate a very distracted dog, which is not desirable. Handlers were also asked to give a clearly valenced score for Overall Ability out of 10, with one being the worst and 10 being the best performance possible.

Study 2

Subjects and Training Searches

A different cohort of nine trainee explosives search dog handlers to those in Study 1 (but at a similar point in training) were filmed, each performing an identical training exercise with their AES dog in which they searched an area for up to 15 min, aiming to locate an explosives training aid.

Video Observations

The same performance measures, briefing, and protocol were followed as in Study 1 (see above), with the exception that all nine searches were watched by all nine handlers (in groups of three). Thus, there were nine "own" search scores and 72 "other" for each characteristic of performance per search-team. Impartial expert ratings were obtained from three independent raters: one dog trainer and team instructor with extensive experience assessing performance, and two experimenters experienced in rating dog performance using the scales. Due to their impartiality and for simplicity, we refer to these as "experts" although some of the handlers also had high levels of experience. Expert raters were blind to the scores given by the handlers.

Statistical Methods

The data were analyzed using non-parametric methods in IBM SPSS statistics 19. Scores were categorized as "own" (the handler rating their own dog's performance) or "other" (rater was not the handler in the clip) or "expert" (Study 2 only).

Study 1

To assess whether handlers agreed in their rankings of search performances from best to worst (irrespective of absolute score) we used Spearman's rank order correlation coefficient (r_s), comparing all ratings for own and other scores (across all 120 searches) and within pairs (20 searches per pair). Values of r_s > 0.7 were taken to indicate a strong association; 0.6–0.7 a good association; 0.5–0.6 moderate and 0.3–0.5 a weak association.

Wilcoxon signed ranks statistic (z) was used to test whether the magnitude of scores from own and other handlers, for each dog, differed significantly.

Study 2

Mean "other" and "expert" scores per behavior per dog were produced. We used mean values as medians frequently masked variation between ratings; mean other handler scores for Confidence, for example, varied between 2 and 5, whereas median scores were 4 for all dogs, thus preventing any correlational analysis. Kendall's Coefficient of Concordance (w) was used to compare own, other and expert categories, and Spearman's correlation coefficients calculated for pairwise comparisons.

Friedman test (T_F) was used to check for overall differences in the magnitude of scores between the three categories of rater, using mean other and expert ratings. Pair-wise Wilcoxon signed rank tests were then used to determine which categories of rater differed significantly. We also tested within the expert-rater category whether raters differed from one another to assess the value of their scores as a "gold standard."

RESULTS

Study 1

When considering ratings for all pairs together, there was moderate agreement in scores for Control and Overall Ability, and weak agreement for Motivation, Distraction and Stamina and little agreement in scores for Confidence, Independence and Indication (**Table 1**). There was however, considerable variability in the level of agreement within pairs, with some pairs in closer agreement across traits than others: pair 1 for example showed good or strong agreement for three traits, whereas pair 4 only agreed (>0.6) for one trait. The likelihood of agreement not only differed between pairs, but also depending on the behavior being scored: for Distraction, for example, pairs 1, 3, and 5 showed good agreement, whereas the other three pairs showed little to no agreement. Rater pairs were more likely to show good agreement when scoring Control and Overall Ability (despite the latter being scored out of 10), with the poorest levels of agreement for Independence.

Handler's own scores for Overall Ability were significantly higher than were other handler's scores (**Table 2**). Handlers also generally scored their own dogs more highly for Control, Motivation, Stamina, Confidence, and Indication; as well as tending to score their own dogs lower (more favorably) for Distraction. As with agreement in rank ordering, whether scores differed significantly varied between pairs and behavioral measures (**Table 3**). For example, five handlers scored their own dog as having significantly higher levels of Motivation, whilst seven did not. There was general disagreement between own and other handlers in scores for Independence, but no clear pattern of favorable marking as three handlers scored their own dog significantly higher and three significantly lower than the other handlers.

Study 2

Considering all categories of rater (own, other, expert) there was moderate to strong agreement for most behavioral traits (**Table 4**) (>0.5), with weak agreement (<0.5) for Motivation, Stamina, and Overall Ability. Pairwise correlations between the categories of rater, indicate that for Distraction, Independence and Indication, agreement was only between other and expert raters; and in general, agreement between own and other, and own and expert, scores was poor (and lower than that between other and expert scores).

Handler's own scores were significantly higher than mean expert scores for Control and they tended to be higher for Indication ($p = 0.06$), but were lower for Distraction and Confidence (**Figure 1**). Other handler scores only differed significantly from experts for Confidence.

There was no significant difference between expert raters in their scores for any of the traits, with the exception of Indication, where one expert rater gave significantly higher scores than both of the other raters ($T_F = 8.12$, $p = 0.017$).

DISCUSSION

In Study 1, handlers generally rated their own dog more favorably than the other handlers, supporting the hypothesis that they exhibited leniency. This was true for all behaviors, except Independence (ability to search without direction), which some handlers rated as higher and others rated lower in their own dogs. This may be because handlers do not see very high levels of Independence as beneficial (6), whereas for all the other behaviors the higher the level of the trait (e.g., Control) the better (except Distraction where the opposite is true). Scores were significantly higher for Control (response to commands), Motivation

TABLE 1 | Study 1: Agreement between own handler and other handler scores (all ratings, $N = 120$).

	Control	Motivation	Stamina	Distraction	Confidence	Independence	Indication	Overall ability
Own/other (all ratings)	0.539	0.486	0.372	0.417	0.025	0.263[e]	0.282[f]	0.529[g]
Pair 1	**0.783**	0.480	0.257	**0.607**	0.312	0.525[a]	−0.379[a]	**0.617**
Pair 2	**0.683**	0.354	**0.607**	0.113	−0.334	0.300	0.414[d]	0.584
Pair 3	0.556	0.365	0.479	**0.622**	−0.061	−0.007	0.191[b]	**0.661[a]**
Pair 4	0.351	0.239	−0.026	0.279	−0.338	0.055	**0.601[a]**	0.272
Pair 5	0.577	0.212	0.067	**0.652**	−0.425	−0.112[a]	0.342[c]	0.518
Pair 6	0.354	**0.756**	0.385	0.099	0.362	0.118	0.494[b]	0.439

Correlation coefficients between handlers in each pair (Spearman's rho, 2-tailed, N = 20, unless stated otherwise) for each trait. Moderate agreement (>0.5) shaded and good agreement (>0.6) in bold.
Where N < 20 within pairs or < 120 for overall comparison.
[a]N = 19; [b]N = 17; [c]N = 16; [d]N = 15; [e]N = 118; [f]N = 103; [g]N = 119.

TABLE 2 | Study 1: Median scores given by handler for own dog's performance and scores given by other handler and Wilcoxon Signed Ranks statistic (z) comparing within dog, across all 12 handlers.

Difference	Control	Motivation	Stamina	Distraction	Confidence	Independence	Indication	Overall ability
Z	−2.658**	−3.251**	−3.390**	1.858 $^{p=0.063}$	−2.726**	−1.147	−2.853**	−3.236**
Median score given to own dog	3.5 (4)	3.9 (4)	4.1 (4)	2.0 (2)	4.3 (4)	3.9 (4)	3.9 (4)	7.1 (7)
Median score given to other dog	3.2 (3)	3.6 (4)	4.0 (4)	2.2 (2)	4.1 (4)	3.8 (4)	3.6 (4)	6.8 (7)

$^{**}p < 0.01$.

TABLE 3 | Study 1: Significant differences within pairs of handlers for each trait as shown by Wilcoxon Signed Rank tests.

Difference		Control	Motivation	Stamina	Distraction	Confidence	Independence	Indication	Overall ability
Pair 1	1			own*	other*			other*	
	2	own*						own**	own*
Pair 2	3								
	4				other*				
Pair 3	5						other*	other*	
	6		own*			own**	own**	own*	
Pair 4	7		own*	own*					own*
	8	own*	own*		other*		own**	own*	own*
Pair 5	9					other*	other*		
	10		own**	own*					
Pair 6	11	own**	own**				other**		
	12					own*	own*	own*	own*

Own denotes the dog's handler scored them significantly higher, other denotes the other handler rated the dog higher ($p < 0.05$; $p < 0.01$**).*

TABLE 4 | Study 2: Levels of agreement between scores given by own handler, other handlers, and experts.

Behavior	Agreement in rank score			
	Own/other/ expert	Own/ other	Own/ expert	Other/ expert
Control	**0.774**	0.587	**0.724**	**0.676**
Motivation	0.460	0.331	0.135	0.110
Stamina	0.443	0.191	0.187	0.129
Distraction	**0.678**	0.179	0.448	**0.906**
Confidence	**0.765**	0.470	**0.878**	**0.619**
Independence	**0.628**	0.184	0.370	**0.715**
Indication	0.587	0.393	0.217	0.519
Overall ability	0.403	−0.028	−0.113	0.421

Kendall's coefficient of concordance (W), for 3 -way comparison and Spearman's rho (r_s) for pairwise comparisons. Moderate agreement (>0.5) shaded and good agreement (>0.6) in bold.

to search, Stamina throughout the search, Confidence in the environment.

In Study 2 overall agreement in the scoring of behaviors was considerably better than that in Study 1, although pairwise correlations indicated that the improvement was most likely due to agreement between the experts and other handlers. Handlers in the second study also showed favorable marking toward their own dog, particularly for levels of Control, Distraction, and Indication. Interestingly, the expert raters scored Confidence in the environment significantly higher than both own and other raters.

Across both studies the level of agreement differed both between raters and between traits, but in general, where there was a difference between raters it seemed to be the result of more favorable scoring by the handler toward their own dog. Hence, handlers have a tendency to be lenient when assessing (or at least when scoring) their own dog's performance. As the same group of handlers show good agreement with experts when applying the rating scale to other handlers' dogs, yet poor agreement with experts when applying the same scales to their own dogs, this shows that they are not applying the same rating principles when assessing their own and other dogs.

Leniency Bias Did Not Affect All Behaviors Equally

Although we found considerable evidence for a leniency bias, the effect was not universal across all performance measures (nor all raters) and there could be several reasons for this.

Ability to Understand the Trait Being Measured

Some characteristics of performance are likely to be harder to rate accurately than others and we would expect greater agreement where behaviors are inherently easier to interpret, as there should be less variation between handlers and also less uncertainty within-raters in how to apply the 1–5 scale on repeated occasions. For example, Control (response to commands) is a relatively easily quantifiable trait and was the most universally comparable between raters. Independence (ability to search without guidance) on the other hand, showed little agreement. If handlers had a similar understanding of the concept and were marking their own dog's searches more leniently, we might still see agreement in ranking from best to worst, as well as more favorable scoring for their own dog's searches; which seemed to be the case for scores for Control.

Several behavior traits in Study 1 showed poor agreement whilst still being scored more favorably by own handlers (e.g., Motivation, Stamina, Confidence, and Indication). We deliberately chose raters with no experience of using the rating scales; however, they may have struggled to rate searches accurately because they didn't understand the traits, or the variation between the five levels of performance within each trait. A lack of understanding of the trait could lead to careless rating or resorting to particular response styles; for example, a net acquiescence response style (14), where a handler scores their own dog at an above average level (but not the highest level) for every search regardless. A lack of agreement between own

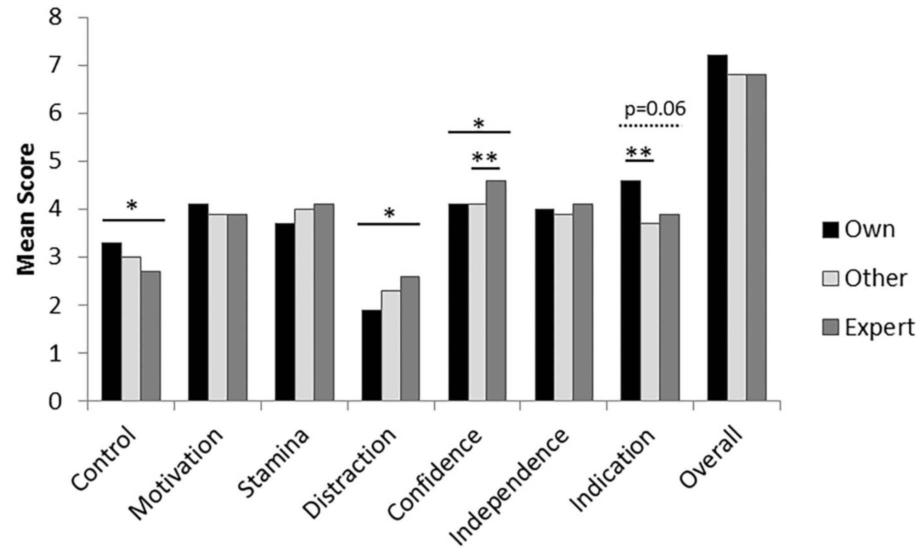


FIGURE 1 | Differences between own (nine raters), mean of other handler ($n = 8$) and expert ($n = 3$) ratings for the performance traits. Asterisks denote significant differences seen between specific raters, using pair-wise Wilcoxon Signed Ranks tests ($p < 0.05$*; $p < 0.01$**).

and others' ratings can be further exaggerated by ambiguity in the rating scales themselves (24) which can lead to raters scoring ambiguous traits in their own favor (19). This may be reduced by providing raters with more information on each of the traits.

While a lack of understanding of the traits may have been partly responsible for some bias, it seems unlikely to be the only reason. In Study 2, agreement between expert scores and the scores given to other handlers' searches was very good across most traits; however there was very poor agreement between the scores that the same group of handlers gave their own dog's searches and the expert scores. In this study, Distraction (from searching) and (strength of) Indication appeared to be readily quantifiable when handlers rated other dog's searches, as they correlated well with (and did not differ significantly from) expert scores, but were particularly vulnerable to bias when rating their own dogs. This effect has been documented in the field of human psychology, with ratings of own performance frequently overestimated, hence leading to greater agreement in work-place performance assessments between peers and supervisors compared to self-peer and self-supervisor ratings (19, 24). Whilst factors related to the scale design are important and can exacerbate this effect, the psychological processes involved in this optimism or over-estimation of own ability are complex and beyond the scope of this paper [see review (19) and "Bias did not affect raters equally" below]. Our findings do, however, suggest that ratings are biased in a similar way as would be expected if handlers rated themselves, potentially as a result of the closeness of the relationship between dog and handler.

It could be argued that the handlers are not lenient, just more familiar with the dog and hence better able to rate its performance. However, given that one of the experts had also trained all the dogs, and that all significant differences in the scores relative to experts and others were in the direction predicted by leniency, we consider this unlikely. Further studies on the effects of training handlers to provide accurate ratings, would now be valuable.

Desirability of Favorable Scores

The relative desirability for a high score within a given trait is also likely to influence how susceptible a measure is to bias (25). For example, we hypothesize that handlers would like their dogs to be maximally obedient and score 5 out of 5 for Control, whereas the ideal score for Independence may in fact be 3 or 4 out of 5. A combination of confusion between raters in what is meant by "Independence," as well as varying opinions on what constitutes the ideal level of the trait (26), may explain why some handlers (Study 1) rated their own dog significantly higher for this trait, and others significantly lower.

Overall Ability is the one measure where participants do not score how much of a trait is present, but how well they have subjectively assessed that the dog performed. Because Overall Ability scores are subjective and clearly valenced (higher scores are more desirable), we would have predicted this measure to show considerable rater bias. Yet, whilst handlers appeared to score their own dogs more highly in both studies, this was only significant in Study 1. Agreement was low to moderate, which may be a consequence of the greater number of scale options (one to 10 scale, as opposed to the 1–5 scales used for the other traits), or a result of the differing relative importance that handlers assign to the individual component characteristics of performance. It may be that because the scale is so clearly valenced, handlers were reluctant to use the whole scale (including the extremes of the scale) for rating either theirs or other handlers' searches. For example only 5% of Overall Ability scores in Study 1, and 7.4% in Study 2, were below five and while it may be that the searches were all of an above average standard, the existence of a net acquiescence bias (or an avoidance of scale extremes) cannot be discounted. One consequence of net acquiescence across all raters is "range restriction," whereby only part of a scale is utilized, which can undermine the validity and reliability of results (27). Net acquiescence may also be responsible for the experts' higher scores for Confidence in Study 2: experts were more likely to score dogs at 5 out of 5 (64% of searches), compared to handlers scoring either their own (22%) or other handlers' (22%) searches.

Bias Did Not Affect Raters Equally

The psychological processes underlying leniency bias are complex and raters may be naïve to their own bias. For example, raters are often able to see bias in the scoring of others, and yet insist that their own ratings are error-free (18). It was clear that not all handlers showed the same degree of bias in scoring. Relative competencies and knowledge are important in producing accurate self-ratings (19); thus, the differences we found may reflect disparity between raters in their understanding of the traits used, or a reliance on rating response styles. These could, in turn, be a result of differing levels of experience and understanding of what constitutes ideal performance (6). It would also interesting to investigate the effect of level of experience on the tendency to be lenient but within this study, although there was variation in handler experience, sample sizes were too small to investigate its effect on rating agreement.

The impact of ratee characteristics on rating ability is well-known in the social science literature [e.g., (28)]. Interest in completing rating tasks, the relevance to the rater, and the perceived importance or consequence of providing accurate ratings are all important motivating factors (12, 17, 29). The raters' personality type (30, 31), their affective state or mood (32), or, in this situation the ratees perception of the "team," such as the level of attachment between handler and dog and the extent to which they see the dog's performance as reflection of their own ability, may all influence the degree of positive bias. The relative impact of some of these factors on ratings provided by search dog handlers is still to be investigated.

Consequences for Performance Measurement

To ensure that the data collected is reliable, it is important to ensure that the performance monitoring process is as objective as possible and without bias, whilst also remaining practically feasible. Positive bias will impact on the validity of information collected using subjective scales, which has implications to any situation where data is reliant on subjective ratings, not just the measurement of working dog performance. Leniency may be particularly important when raters have a vested interest in the outcome, but even where this isn't the case, there may be issues with other biases, such as net acquiescence and a reluctance to use the whole scale. Hence, if rating scales are to be used effectively then efforts must be made to check for, and to overcome, biases.

Several measures can be undertaken to reduce the effect of bias. Scales should initially be validated to assess whether some components are more prone to bias. Improving scale design (14), for example providing scale benchmarks (33) may help to improve understanding of the dimensions being measured and the value of benchmarking has been investigated for these scales (5). Using statistical methods to adjust data (34) or partition error variance (35) could be considered, although caution should be exercised when manipulating data [see (36)], especially where bias is not universal across all measures or raters. Care must also be taken to ensure that any supposed bias is not, in fact, an accurate reflection of a skew in the population being measured (i.e., low natural variation in performance). Understanding differences between raters and the occurrence of response styles is also important and rater training may help to simultaneously reduce bias and increase motivation to provide accurate ratings (37).

CONCLUSIONS

Dog handlers showed favorable scoring, or leniency, for several traits of search performance. The degree of bias varied with the trait being scored and also between raters. Raters showed variation in agreement suggesting that they differed in their understanding of the meaning of the traits being measured, although rater bias may have been partly responsible as handlers agreed with expert ratings when assessing other handlers' dogs. Improvements therefore need to be made to ensure the reliability and validity of ratings if they are to be made by lone working handlers. We believe this will be achievable through effective methods of training handlers to rate dogs objectively, potentially both reducing bias and improving understanding and thereby consistent use of scales. This study, whilst using search dog handlers, has relevance in any situation where raters must assess the performance of others, particularly where they may have a vested interest in, or loyalty toward, the ratee.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because they contain information on military working dog performance and are thus sensitive. Requests to access the datasets should be directed to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. The animal study was reviewed and approved by University of Bristol, Animal Welfare & Ethical Review Board. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

NR and CC conception of idea and preparation of manuscript. CC, NS, and NR methods development. CC and NS data collection and analysis, with input from NR. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by UK Ministry of Defence (MOD). Quantification of Performance - VS and HASD dog roles Ref DFAS RK7283.

ACKNOWLEDGMENTS

We would like to thank the 1st Military Working Dogs Regiment (Royal Army Vet Corps), including personnel at the Defence Animal Centre for their invaluable input throughout this project. We would also like to thank Bill Browne for statistical advice, John Bradshaw for help with designing the study and Bill Browne and Jo Hockenhull for comments on the manuscript.

REFERENCES

1. DoHS. *Federal Protective Service Explosive Dog Detection Teams*. Department of Homeland Security. (2019). Available online at: www.dhs.gov/explosive-detection-dog-teams (accessed 17 March, 2019).
2. Harper RJ, Almirall JR, Furton KG. Identification of dominant odor chemicals emanating from explosives for use in developing optimal training aid combinations and mimics for canine detection. *Talanta*. (2005) 67:313–27. doi: 10.1016/j.talanta.2005.05.019
3. Rooney NJ, Morant S, Guest C. Investigation into the value of trained glycaemia alert dogs to clients with type I diabetes. *PLoS ONE*. (2013) 8: e69921. doi: 10.1371/journal.pone.0069921
4. Rooney N, Guest C, Swanson L, Morant S. How effective are trained dogs at alerting their owners to changes in blood glycaemic levels? Variations in performance of glycaemia alert dogs. *PLoS ONE*. (2019) 14:e0210092. doi: 10.1371/journal.pone.0210092
5. Rooney NJ, Clark CCA. Development of data capture tools and protocols for different classes of dogs (VS and HAS). (2013).
6. Rooney NJ, Bradshaw JWS, Almey H. Attributes of specialist search dogs—a questionnaire survey of, U. K. dog handlers and trainers. *J Forensic Sci*. (2004) 49:300–6. doi: 10.1371/journal.pone.0069921
7. Rooney NJ, Gaines SA, Bradshaw JWS, Penman S. Validation of a method for assessing the ability of trainee specialist search dogs. *Appl Anim Behav Sci*. (2007) 103:90–104. doi: 10.1016/j.applanim.2006.03.016
8. Holzbach RL. Rater bias in performance ratings: superior, self-, and peer ratings. *J Appl Psychol*. (1978) 63:579–88. doi: 10.1037/0021-9010.63.5.579
9. Michelson L, Mannarino A, Marchione K, Kazdin AE, Costello A. Expectancy bias in behavioral observations of therapeutic outcome: an experimental analysis of treatment and halo effects. *Behav Res Therapy*. (1985) 23:407–14. doi: 10.1016/0005-7967(85)90168-8
10. Smith DE. Training programs for performance appraisal: a review. *Acad Manag Rev*. (1986) 11:22–40. doi: 10.5465/amr.1986.4282615
11. Krosnick JA. Survey research. *Ann Rev Psychol*. (1999) 50:537–67. doi: 10.1146/annurev.psych.50.1.537
12. Bernardin HJ, Orban J. Leniency effect as a function of rating format, purpose for appraisal, and rater individual differences. *J Bus Psychol*. (1990) 5:197–211. doi: 10.1007/BF01014332
13. Ng KY, Koh C, Ang S, Kennedy JC, Chan KY. Rating leniency and halo in multisource feedback ratings: testing cultural assumptions of power distance and individualism-collectivism. *J Appl Psychol*. (2011) 96:1033–44. doi: 10.1037/a0023368
14. Van Vaerenbergh Y, Thomas TD. Response styles in survey research: a literature review of antecedents, consequences, and remedies. *Int J Public Opin Res*. (2013) 25:195–217. doi: 10.1093/ijpor/eds021
15. Weijters B, Geuens M, Schillewaert N. The stability of individual response styles. *Psychol Method*. (2010) 15:96–110. doi: 10.1037/a0018721
16. Lenzner T, Kaczmirek L, Lenzner A. Cognitive burden of survey questions and response times: a psycholinguistic experiment. *Appl Cognitive Psychol*. (2010) 24:1003–20. doi: 10.1002/acp.1602
17. Jawahar IM, Williams CR. Where all the children are above average: the performance appraisal purpose effect. *Personnel Psychol*. (1997) 50:905–25. doi: 10.1111/j.1744-6570.1997.tb01487.x
18. Pronin E, Lin DY, Ross L. The bias blind spot: perceptions of bias in self versus others. *Personal Soc Psychol Bulletin*. (2002) 28:369–81. doi: 10.1177/0146167202286008
19. Dunning D, Heath C, Suls JM. Flawed self-assessment: implications for health, education, and the workplace. *Psychol Sci Public Interest*. (2004) 5:69–106. doi: 10.1111/j.1529-1006.2004.00018.x
20. Hoffman CC, Nathan BR, Holden LM. A comparison of validation criteria: objective versus subjective performance measures and self-versus superior ratings. *Personnel Psychol*. (1991) 44:601–18. doi: 10.1111/j.1744-6570.1991.tb02405.x
21. Davis DA, Mazmanian PE, Fordis M, Van Harrison RR, Thorpe KE, et al. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. (2006) 296:1094–102. doi: 10.1001/jama.296.9.1094
22. Brewer MB. In-group bias in the minimal intergroup situation: a cognitive-motivational analysis. *Psychol Bulletin*. (1979) 86:307–24. doi: 10.1037/0033-2909.86.2.307
23. Dancer WT, Dancer J. Peer rating in higher education. *J Educ Bus*. (1992) 67:306. doi: 10.1080/08832323.1992.10117564
24. Harris MM, Schaubroeck J. A meta-analysis of self supervisor, self peer, and self supervisor ratings. *Personnel Psychol*. (1988) 41:43–62. doi: 10.1111/j.1744-6570.1988.tb00631.x
25. Gosling SD, John OP, Craik KH, Robins RW. Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *J Personal Soc Psychol*. (1998) 74:1337–49. doi: 10.1037/0022-3514.74.5.1337
26. Rooney NJ, Bradshaw JWS. Breed and sex differences in the behavioural attributes of specialist search dogs–a questionnaire survey of trainers and handlers. *Appl Anim Behav Sci*. (2004) 86:123–35. doi: 10.1016/j.applanim.2003.12.007
27. Murphy KR, Deshon R. Interrater correlations do not estimate the reliability of job performance ratings. *Personnel Psychol*. (2000) 53:873–900. doi: 10.1111/j.1744-6570.2000.tb02421.x
28. Landy FJ, Farr JL. Performance rating. *Psychol Bulletin*. (1980) 87:72–107. doi: 10.1037/0033-2909.87.1.72
29. Aichholzer J. Intra-individual variation of extreme response style in mixed-mode panel studies. *Soc Sci Res*. (2013) 42:957–70. doi: 10.1016/j.ssresearch.2013.01.002
30. Bernardin HJ, Cooke DK, Villanova P. Conscientiousness and agreeableness as predictors of rating leniency. *J Appl Psychol*. (2000) 85:232–6. doi: 10.1037/0021-9010.85.2.232
31. Bernardin HJ, Tyler CL, Villanova P. Rating level and accuracy as a function of rater personality. *Int J Select Assessment*. (2009) 17:300–10. doi: 10.1111/j.1468-2389.2009.00472.x
32. Antonioni D, Park H. The relationship between rater affect and three sources of 360-degree feedback ratings. *J Manag*. (2001) 27:479–95. doi: 10.1177/014920630102700405
33. Weijters B, Cabooter E, Schillewaert N. The effect of rating scale format on response styles: the number of response categories and response category labels. *Int J Res Marketing*. (2010) 27:236–47. doi: 10.1016/j.ijresmar.2010.02.004

34. Landy FJ, Barnesfarrell JL, Vance RJ, Steele JW. Statistical control of Halo error in performance ratings. *J Appl Psychol.* (1980) 65:501–6. doi: 10.1037/0021-9010.65.5.501
35. Diamantopoulos A, Reynolds NL, Simintiras AC. The impact of response styles on the stability of cross-national comparisons. *J Bus Res.* (2006) 59:925–35. doi: 10.1016/j.jbusres.2006.03.001
36. Landy FJ, Vance RJ, Barnesfarrell JL. Statistical control of Halo—a response. *J Appl Psychol.* (1982) 67:177–80. doi: 10.1037/0021-9010.67.2.177
37. Sulsky LM, Kline TJB. Understanding frame-of-reference training success: a social learning theory perspective. *Int J Training Dev.* (2007) 11:121–31. doi: 10.1111/j.1468-2419.2007.00273.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 16 October 2020
doi: 10.3389/fvets.2020.00624

TeamMate: A Longitudinal Study of New Zealand Working Farm Dogs. II. Occurrence of Musculoskeletal Abnormalities

Katja E. Isaksen[1], Lori Linney[2], Helen Williamson[2], Nick J. Cave[1], Elizabeth J. Norman[3] and Naomi Cogger[1]*

[1] School of Veterinary Science, Massey University, Palmerston North, New Zealand, [2] Vetlife, Timaru, New Zealand, [3] College of Sciences, Massey University, Palmerston North, New Zealand

Musculoskeletal injury and disease are common in dogs, and a major cause of retirement in working dogs. Many livestock farmers rely on dogs for the effective running of their farms. However, the incidence of musculoskeletal disease has not been explored in working farm dogs. Here we explore the occurrence of musculoskeletal abnormalities in 323 working farm dogs that were enrolled in TeamMate, a longitudinal study of working farm dogs in New Zealand. All dogs were free of musculoskeletal abnormalities on enrolment to the study and were present for at least one follow-up examination. During the follow-up period, 184 dogs (57%, 95% confidence interval (CI) = 52%−62%) developed at least one musculoskeletal abnormality during 4,508 dog-months at risk, corresponding to 4.1 dogs (95% CI = 3.5–4.7) with recorded abnormalities per 100 dog-months at risk. The most common abnormalities were reduced range of motion and swelling of the carpus or stifle, while the hip was the most common site of pain. No major differences in incidence rate (IR) between sexes or types of dogs were observed, though Huntaways had a slightly lower rate of carpal abnormalities than Heading dogs (IR ratio = 0.6, 95% CI = 0.3–1.0). Eighty-one of 119 dogs (68%, 95% CI = 60%−76%) that had a first musculoskeletal abormality developed a second abnormality. The most common type of abnormality that was seen in the same dog more than once was reduced range of motion in the carpus (14 of 119 dogs, 12%, 95% CI = 6%−18%). Although we do not provide data on diagnoses, the high incidence rate of recorded musculoskeletal abnormalities and dogs' high activity mean it is likely that working farm dogs are at a high risk of conditions that could impair their welfare and reduce the lengths of their working careers. Preventing and managing musculoskeletal injury and illness should be a priority for owners and veterinarians caring for working farm dogs.

Keywords: musculoskeletal, longitudinal, TeamMate, incidence, working dogs, herding dogs, working farm dogs

OPEN ACCESS

Edited by:
Cynthia M. Otto, University of Pennsylvania, United States

Reviewed by:
Mieghan Bruce, Murdoch University, Australia
Janice Lauren Baker, Veterinary Tactical Group, United States

****Correspondence:***
Katja E. Isaksen
katjaisaksen01@gmail.com

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences, a section of the journal Frontiers in Veterinary Science

Received: *15 February 2020*
Accepted: *03 August 2020*
Published: *16 October 2020*

Citation:
Isaksen KE, Linney L, Williamson H, Cave NJ, Norman EJ and Cogger N (2020) TeamMate: A Longitudinal Study of New Zealand Working Farm Dogs. II. Occurrence of Musculoskeletal Abnormalities. Front. Vet. Sci. 7:624. doi: 10.3389/fvets.2020.00624

INTRODUCTION

Musculoskeletal injury and disease is common in many populations of dogs, humans, and other species (1–4) and can be a serious problem that affects overall health, welfare, and working performance (5–7). In the United Kingdom, the second most commonly recorded cause of death of dogs attending clinical practice was musculoskeletal disorders (4). In New Zealand police dogs, and United Kingdom guide dogs, the most common cause for early retirement was an inability to continue working due to musculoskeletal disease or injury (8, 9). In United States military working dogs, the most commonly recorded cause of dogs dying was degenerative joint disease (10).

Working farm dogs in New Zealand have been found to have a high prevalence of musculoskeletal disease and injury, with over 40% having at least one musculoskeletal abnormality on physical examination (11). Additionally, during a 12-month period, 14% of working farm dogs had a non-traumatic musculoskeletal health event and 12% had a traumatic musculoskeletal health event, according to owners (12). Musculoskeletal disease can be a major cause of reduced quality of life due to its potential to cause pain and limit mobility (3, 13). High levels of activity such as those seen in working farm dogs (14) can contribute to increased levels of musculoskeletal disease, limiting the dogs' ability to work. Given the reliance of New Zealand farmers on their dogs for the efficient running of their farms (15), and the economic value stock-herding dogs bring to their owners (16), high incidences of musculoskeletal injury and disease may represent a major economic cost to owners of working farm dogs. Determining what types of musculoskeletal abnormalities are the most common and whether certain dogs are at increased risk of developing musculoskeletal disease could enable veterinarians and dog owners to target preventative measures more accurately. In turn, such targeting would improve dogs' health and welfare and ensure that they stay disease-free and able to work for as long as possible.

To date, the incidence of musculoskeletal injury and disease in working farm dogs has not been investigated. The aim of this study was to describe the incidence of different types of musculoskeletal abnormalities recorded in a population of working farm dogs. We anticipated that the incidence of musculoskeletal abnormalities would be associated with the sex and type of the dogs. The incidence of dogs developing musculoskeletal abnormalities is presented, stratified by the types and locations of the abnormalities seen.

METHODS

Study Design

TeamMate is a longitudinal study focusing on working farm dogs on the South Island of New Zealand. A companion research article describes the study design and data collection procedure in detail and presents data collected on the dogs' enrolment to the study (11). To summarize, 641 working farm dogs were convenience-sampled and enrolled in a four-year longitudinal study. All working farm dogs belonging to participating dog owners were enrolled, if they were least 18 months old and working with livestock regularly. In the current study, we included 323 dogs that did not have a recorded musculoskeletal abnormality on enrolment and that were present for at least one subsequent clinical examination.

Data collection was begun in May 2014. Data was collected approximately every eight to nine months subsequently, and data from five data collection rounds were included in the current study. The fifth data collection round was completed in November 2017. **Figure 1** is a flowchart showing the start dates for each data collection round and how many dogs, owners, and farming properties were enrolled at each round. At each data collection round, farm dog owners were visited on the farm where they worked, new dog owners and dogs were enrolled, and data was collected. New dog owners and dogs were enrolled up to and including the third data collection round. New dogs included dogs belonging to previously enrolled owners that had been acquired or had aged into the study between farm visits. Some new properties were registered subsequently to the third data collection round due to participating dog owners moving or changing jobs.

At each farm visit, including on enrolment, all enrolled dogs were physically examined by veterinarians and dog owners were interviewed to collect information about the dogs' husbandry, feeding, and work. Scribes were responsible for filling in the questionnaires and taking note of any clinical findings. The physical examination included manipulation of all the major joints and encouraging the dogs to trot for a short distance to check for lameness. All physical abnormalities were recorded, irrespective of their clinical significance. All questionnaires that were used as part of the study are available as supplementary materials to a companion research article (11).

All veterinarians and scribes were trained to ensure data collection was performed in a standardized way, with veterinarians asked to record specific clinical signs rather than make general diagnoses. Training included a run-through of all questionnaires and how they should be completed as well as practical sessions that involved filling in the questionnaires and examining, scoring, and measuring farm dogs. During training sessions, normal range of motion at each joint was demonstrated in healthy working farm dogs.

Abnormalities noted on clinical examination were systematically categorized using alphanumeric codes based on the examining veterinarian's notes. Each code consisted of a letter signifying the body system involved and up to five numbers signifying the location, symmetry, type, and cause of the abnormality. Abnormalities were not mutually exclusive, and dogs could have multiple recorded abnormalities, also in the same anatomical location. Coding was carried out by a single veterinarian (LL) and checked by another person with training in veterinary health. Codes that were unclear or incomplete were rechecked by a veterinarian (LL and/or NJC). The complete system used for alphanumeric coding of physical abnormalities is available as supplementary materials to a companion research article (11).

Dogs that were enrolled in TeamMate, were free of recorded musculoskeletal abnormalities at enrolment, and were present for at least one follow-up clinical examination were included in the current study. Data relating to the occurrence of musculoskeletal abnormalities in these dogs are presented.

Statistical Analysis

Abnormalities noted on physical examination were categorized according to type and location on the body. Anatomical locations and types of abnormalities were included in further data analysis if they were seen in 10% of dogs or more either on enrolment (11)

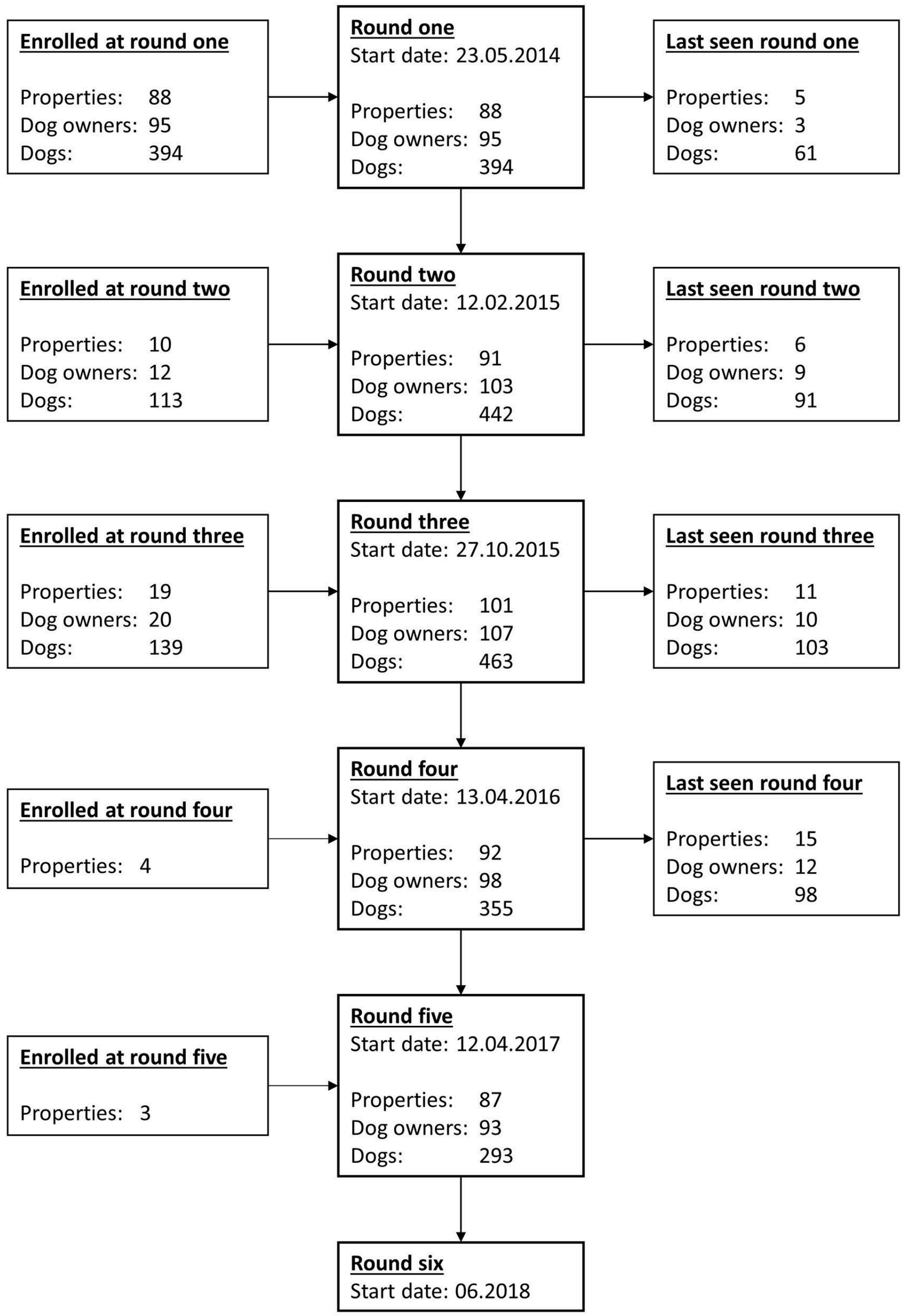


FIGURE 1 | Flow chart showing the start dates of each data collection round as well as the number of farms, dog owners and dogs enrolled in TeamMate up to and including the fifth round of farm visits. Additionally, 14 properties, 16 dog owners and 68 dogs missed at least one round of data collection. Note that data for the sixth data collection round was not yet available at the time of writing. This figure was previously published by the authors (11) and is licensed for re-use under the Creative Commons Attribution 4.0 International licence.

or as a first musculoskeletal abnormality following enrolment. The anatomical locations included the carpals, hips, digits, and stifles, and abnormalities were categorized as "abnormal range of motion," "hard swelling," "painful," "crepitus," or "other." Lameness on trot was recorded in 12% of dogs on enrolment (11). However, we did not include lameness in this study as we cannot know that the underlying cause of lameness is musculoskeletal. For example, dogs may be lame due to injuries to the footpads.

Dogs were classified by type based on the description provided by the owner. The most common types of dogs were Heading dogs and Huntaways, with other types of dogs combined and classified as "other." A more detailed discussion on the

TABLE 1 | Modes of work done by working farm dogs in New Zealand.

Mode of work	Description
Head	The dog circles around to the head of the herd and uses its positioning to gather, stop, and redirect animals.
Hunt	The dog uses its bark and position to apply pressure on the herd from behind in order to move the animals forward.
Yard work	Any work done in stockyards and runs.
Catch	Separating one or several specific animals from the herd.

classification of dog types and the differences between them can be found in a companion research paper (11). In short, Heading dogs and Huntaways are unique to New Zealand and have been bred to carry out different types of stock work. Heading dogs are similar to short-haired Border collies and weigh about 20 kg on average. They are usually trained to head and catch, and sometimes to do yard work (see **Table 1** for descriptions of work types). Huntaways are larger than Heading dogs at an average weight of around 30 kg. Huntaways are usually trained to hunt and to do yard work.

Time at risk to a first recorded musculoskeletal abnormality was calculated using an approximate calculation adapted from that described in Dohoo et al. (17), with dogs considered as having been withdrawn if they were lost to follow-up for any reason at an earlier date than their owner. The start time at risk was defined as the date on which an individual dog was enrolled in the study. Dogs were considered as no longer being at risk if they were recorded as having a musculoskeletal abnormality, if they or their owner were lost to follow-up for any reason, or if they reached the end of the study. Dogs that were recorded as having a musculoskeletal abnormality or were withdrawn were considered as being at risk until the halfway point between the date of their previous examination and the date on which the abnormality or the withdrawal was recorded. Dogs that were not recorded as having any musculoskeletal abnormalities or having been withdrawn were considered as being at risk until the date of their last recorded examination. Time at risk to a second recorded musculoskeletal abnormality was calculated in the same way as the first, except that the start time was considered as being the date on which dogs' first musculoskeletal abnormality was recorded.

Incidence rate was calculated as the number of dogs that had at least one musculoskeletal abnormality divided by the number of dog-months at risk. Note that this is not same as the number of injuries per time period. Dogs may have had more than one recorded abnormality on the same examination. Additionally, single cases of injury or disease were often coded more than once as a reflection of multiple clinical signs. For example, a dog may have swelling, reduced range of motion, and pain in the same joint. For these reasons, the number of dogs rather than the number of abnormalities were counted.

Incidence rates and 95% CIs were calculated for the first instance of any musculoskeletal abnormality in each dog as well as for the most common types and locations of abnormalities.

TABLE 2 | Population features of the 323 dogs enrolled in TeamMate that did not have a recorded abnormality on enrolment and were present for at least one follow-up examination.

Variables		Number of dogs	% (95% CI)
Sex	Female	151	47 (41–52)
	Male	172	53 (48–59)
Age on enrolment	1.5 to 2.9 years	187	58 (53–63)
	3 to 4.9 years	87	27 (22–32)
	5 to 6.9 years	27	8 (5–11)
	7 to 9.9 years	21	7 (4–9)
	10 years and above	1	0 (0–1)
Type of dog	Heading dog	165	51 (46–57)
	Huntaway	148	46 (40–51)
	Other	10	3 (1–5)

Specific incidence rates, stratified by sex and type of dog, were calculated for each of the most common joint locations, and incidence rate ratios for sex and dog types were calculated with 95% CIs.

Incidence rate was also reported for second occurrences of musculoskeletal abnormalities. The calculation of time at risk included dogs that were recorded as having a first musculoskeletal abnormality and that were present for at least one subsequent examination. The types of abnormalities that were most commonly observed more than once in the same dog are reported.

All data analysis was done using R version 3.6.x (18).

RESULTS

Three hundred twenty-three dogs, belonging to 113 dog owners, did not have a recorded musculoskeletal abnormality on enrolment to TeamMate and were present for at least one follow-up clinical examination. These 323 dogs contributed 4,508 dog-months at risk. **Table 2** shows the distribution of dogs by sex, age group at enrolment, type of dog. The median age at enrolment for both sexes was 3 years (IQR = 2–5 years). The median age at enrolment was 3 years (IQR = 2–4 years) for Heading dogs, 3 years (IQR = 2–5 years) for Huntaways, and 4 years (IQR = 3–8 years) for other types of dogs. For comparison, the median age on enrolment of all 641 dogs enrolled in TeamMate was 4 years (IQR = 2–6) (11).

Of 323 dogs, 184 (57%, 95% CI = 52–62%) developed at least one musculoskeletal abnormality during 4,508 dog-months at risk, corresponding to 4.1 dogs (95% CI = 3.5–4.7) per 100 dog-months at risk. **Table 3** describes the incidence rate of dogs' first recorded musculoskeletal abnormalities following enrolment, stratified by anatomical location and type of abnormality. **Tables 4**, **5** describe the distribution of incidence rates and rate ratios of the first occurrence of musculoskeletal abnormalities in the most commonly recorded anatomical locations, stratified by sex and type of dog, respectively.

TABLE 3 | Number of affected dogs, incidence rate, and incidence rate ratio (with 95% CI) of first recorded musculoskeletal abnormalities stratified by the location on the body and type of the first recorded abnormality.

Location	Type of abnormality	Number of dogs	IR/100 dog-months (95% CI)
Carpus	Abnormal range of motion*	44	1.0 (0.7–1.3)
	Painful	6	0.1 (0.1–0.3)
	Hard swelling	9	0.2 (0.1–0.4)
	Crepitus	4	0.1 (0.0–0.2)
	All carpus	53	1.2 (0.9–1.5)
Hip	Abnormal range of motion*	22	0.5 (0.3–0.7)
	Painful	18	0.4 (0.3–0.6)
	Crepitus	2	0.0 (0.0–0.2)
	Other	2	0.0 (0.0–0.2)
	All hip	39	0.9 (0.6–1.2)
Digits	Abnormal range of motion*	11	0.2 (0.1–0.4)
	Hard swelling	5	0.1 (0.0–0.3)
	Painful	24	0.5 (0.4–0.8)
	Crepitus	5	0.1 (0.0–0.3)
	All digits	36	0.8 (0.6–1.1)
Stifle	Abnormal range of motion*	7	0.2 (0.1–0.3)
	Hard swelling	4	0.1 (0.0–0.2)
	Painful	9	0.2 (0.1–0.4)
	Crepitus	9	0.2 (0.1–0.4)
	All stifle	25	0.6 (0.4–0.8)
Other	Abnormal range of motion*	41	0.9 (0.7–1.2)
	Hard swelling	30	0.7 (0.5–1.0)
	Painful	11	0.2 (0.1–0.4)
	Crepitus	6	0.1 (0.1–0.3)
	Other	8	0.2 (0.1–0.4)
	All other	86	1.9 (1.5–2.4)
All abnormalities	Abnormal range of motion*	102	2.3 (1.9–2.7)
	Hard swelling	56	1.2 (1.0–1.6)
	Painful	48	1.1 (0.8–1.4)
	Crepitus	21	0.5 (0.3–0.7)
	Other	17	0.4 (0.2–0.6)
	All abnormalities	184	4.1 (3.5–4.7)

Data from 323 dogs that contributed 4,508 dog-months at risk. Note that many dogs were recorded as having more than one abnormality on the same examination. Anatomical locations and types of abnormalities were classed as "Other" if they were recorded in fewer than 10% of dogs on enrolment, or as a first musculoskeletal abnormality following enrolment.

**Two dogs were found to have abnormally increased range of motion, one in the shoulder and the other in the tarsus. The remainder had reduced range of motion.*

Of 184 dogs that were recorded to have had a first musculoskeletal abnormality 119 dogs (65%, 95% CI = 65%−72%) were present for at least one subsequent follow-up physical examination and contributed 1,144 dog-months at risk. Eighty-one of 119 dogs (68%, 95% CI = 60%−76%) were found to have a second musculoskeletal abnormality of any type. This corresponds to 7.1 dogs (95% CI = 5.7–8.7) per 100 dog-months at risk. Thirty-one of 119 dogs (26%, 95% CI = 18%−34%) were found to have a musculoskeletal abnormality of both the same type and in the same location on a subsequent examination. The most common abnormalities that were seen in the same dog repeatedly were reduced range of motion in the carpus (14 of 119 dogs, 12%, 95% CI = 6%−18%) and hard swelling in one or more digits (4 of 119 dogs, 3%, 95% CI = 0%−7%). All other types of abnormalities were seen repeatedly in three dogs or fewer.

TABLE 4 | Number of affected dogs, incidence rate, and incidence rate ratio (with 95% CI) of first recorded musculoskeletal abnormalities in a range of anatomical locations, stratified by sex.

Location	Sex	Number of dogs	IR/100 dog-months (95% CI)	IR ratio (95% CI)
Carpus	Female	24	1.1 (0.9–1.3)	
	Male	29	1.3 (1.1–1.5)	1.2 (0.7–2.0)
Hip	Female	25	1.1 (1.0–1.3)	
	Male	14	0.6 (0.5–0.7)	0.6 (0.3–1.1)
Digits	Female	14	0.6 (0.5–0.7)	
	Male	22	1.0 (0.8–1.1)	1.5 (0.8–3.0)
Stifle	Female	11	0.5 (0.4–0.6)	
	Male	14	0.6 (0.5–0.7)	1.3 (0.6–2.8)
Other	Female	55	1.7 (1.5–2.0)	
	Male	65	2.1 (1.8–2.4)	1.2 (0.8–1.8)
All locations	Female	86	3.8 (3.3–4.5)	
	Male	98	4.3 (3.7–5.0)	1.1 (0.8–1.5)

One hundred fifty-one female dogs contributed 2,238 dog-months at risk and 172 male dogs contributed 2,270 dog-months at risk.

DISCUSSION

This study confirms that musculoskeletal abnormalities are common in working farm dogs, with almost six in 10 dogs developing at least one musculoskeletal abnormality during the course of the study, at a rate of more than 4 dogs per 100 dog-months at risk. To our knowledge, this is the first time incidence rate of musculoskeletal disease or injury has been reported in a population of working dogs. Musculoskeletal disease and injury cause discomfort, pain, and loss of mobility that can have implications for the welfare of the affected dogs and is likely to cause a reduction in working capacity. In the short term, this loss of working capacity might put extra strain on the remaining dogs on farm as they are required to fill the gap or cause productivity issues on farm as the dog owner is unable to move stock efficiently. Additionally, injured dogs may have incomplete recoveries or lowered fitness following rest, while the remaining healthy dogs are given increased workloads. In humans, previous injury, lowered fitness, and overuse are all linked to further injuries such as tendinopathy, stress fractures, and osteochondrosis (19), while a survey of sled racing dogs suggested that overuse may have been the cause of certain injuries (20). In the long term, overuse and repeated injuries are risk

TABLE 5 | Number of affected dogs, incidence rate and incidence rate ratio (with 95% CI) of first recorded musculoskeletal abnormalities in a range of anatomical locations, stratified by type of dogs.

Location	Type of dog	Number of dogs	IR / 100 dog-months (95% CI)	IR ratio (95% CI)
Carpus	Heading dog	33	1.4 (1.2–1.6)	
	Huntaway	15	0.8 (0.6–0.9)	0.6 (0.3–1.0)
	Other	5	3.2 (1.8–5.9)	2.3 (0.9–6.0)
Hip	Heading dog	20	0.8 (0.7–1.0)	
	Huntaway	18	0.9 (0.8–1.1)	1.1 (0.6–2.1)
	Other	1	0.6 (0.3–1.2)	0.8 (0.1–5.7)
Digits	Heading dog	20	0.8 (0.7–1.0)	
	Huntaway	15	0.8 (0.6–0.9)	0.9 (0.5–1.8)
	Other	1	0.6 (0.3–1.2)	0.8 (0.1–5.7)
Stifle	Heading dog	14	0.6 (0.5–0.7)	
	Huntaway	9	0.5 (0.4–0.5)	0.8 (0.3–1.8)
	Other	2	1.3 (0.7–2.4)	2.2 (0.5–9.6)
Other	Heading dog	45	1.9 (1.6–2.2)	
	Huntaway	38	1.9 (1.6–2.3)	1.0 (0.7–1.6)
	Other	3	1.9 (1.0–3.6)	1.0 (0.3–3.3)
All locations	Heading dog	92	3.9 (3.3–4.5)	
	Huntaway	85	4.3 (3.7–5.1)	1.1 (0.8–1.5)
	Other	7	4.5 (2.5–8.3)	1.2 (0.5–2.5)

One hundred sixty-five Heading dogs contributed 2,385 dog-months at risk, 148 Huntaways contributed 1,968 dog-months at risk, and 10 other types of dogs contributed 155 dog-months at risk.

factors for the development of chronic musculoskeletal disease such as osteoarthritis (21).

In this study, more than two thirds of dogs that had a musculoskeletal abnormality and were present for a subsequent examination were recorded to have a second musculoskeletal abnormality on a later examination, and more than a quarter were recorded as having the same abnormality a second time. The data recorded for this study focused on clinical signs rather than diagnosis, and there is no data available on whether repeated observations of abnormalities represent persistent musculoskeletal disease or new injuries in the same location. Either case, however, may be associated with the presence of chronic disease because repeated injuries may lead to chronic conditions such as osteoarthritis (21).

The carpal joint had the highest incidence rate of abnormalities in this study, and most of these involved reduced range of motion (**Table 3**). This type of abnormality was also, by far, the most common type to be recorded more than once in the same dogs, indicating that this type of abnormality may be more likely to persist over time than other types of abnormalities. However, more detailed data is needed to confirm or negate this assumption. Carpal injuries have been found to be common in racing Greyhounds (22, 23), while a survey of sled racing dogs suggested that carpal injuries may have been the result of overuse (20). Similarly, high activity levels may predispose working farm dogs to carpal injuries. This would explain the high incidence of carpal abnormalities seen in this study. Carpal abnormalities reported in this study rarely involved pain on manipulation, and it is likely that many were the results of minor injuries or changes caused by healing after injury. Dog owners may not consider these injuries serious enough to warrant a visit to a veterinary clinic. Given the effect of chronic musculoskeletal illness on other working dog populations, more research is warranted to quantify the effect of carpal injuries on the health and welfare of working farm dogs. Based on current data, it might be prudent for veterinarians and working dog owners to follow up dogs with carpal injuries and give them the necessary support to prevent and, if necessary, manage chronic musculoskeletal illness.

Except for a slightly higher rate of carpal abnormalities in Heading dogs than Huntaways, no major differences were seen in the rates of musculoskeletal abnormalities between the sexes or types of dogs (**Tables 4**, **5**). The 95% CIs of the incidence rate ratios were narrow, indicating that our results are probably quite close to the "true" values in the study population. If this is accurate, any differences in the rates of musculoskeletal illness or injuries between sexes or types of working farm dogs are so small that they can probably be disregarded in clinical settings. As the occurrence of musculoskeletal disease and injury is known to increase with age (3, 24), a possible source of confounding in our results would be if there were pronounced age differences between the sexes or types of dogs. However, age differences between groups were not observed in this population. The small difference seen in the rate of carpal abnormalities could be spurious, or it could be explained by several factors. Heading dogs and Huntaways are phenotypically distinct (Isaksen et al., unpublished data), with Huntaways being on average ∼10 kg heavier than Heading dogs (11). Health differences between breeds and phenotypes are commonly seen in dogs (2, 4, 25, 26). However, Heading dogs and Huntaways also do different types of stock work (11), which may put them at risk of different types of injuries. Cave et al. reported that along with automotive accidents, stock-related trauma was a major cause of injury in working farm dogs, and that Heading dogs were over-represented in comparison to Huntaways (27). Our data suggests that Heading dogs may be at slightly higher risk of carpal injuries than Huntaways, though further investigation of risk factors related to phenotypes and work in working farm dogs is needed. With carpal abnormalities being the most commonly reported in the population overall, these types of injuries should not be discounted in Huntaways based on the weak difference reported in this study.

No difference in the rate of hip abnormalities was seen between Heading dogs and Huntaways, and the overall incidence rate was around one per 100 dog-months. The majority of recorded hip abnormalities involved reduced range of motion and/or signs on pain, potentially impairing dogs' mobility, and overall welfare. A previous study by Hughes (28) suggested an 18% prevalence of hip dysplasia in working farm dogs, with Huntaways having a five times higher prevalence than Heading dogs. However, Hughes reports that the majority of dog owners had not noticed lameness in dogs that were scored as having hip dysplasia. Cave et al. (27) suggested that more Huntaways have hip dysplasia while more Heading dogs have hip luxation. However, the study recorded only 23 cases of hip dysplasia and 31 cases of traumatic injury to the hip in 2,214 clinic presentations. In TeamMate, prevalence of hip abnormalities on enrolment was 14% (11). The differences seen between these studies can probably be explained by differences in study design, with Hughes possibly recruiting dog owners that were concerned about hip disease in their teams, Cave et al. only recording dogs that were considered by their owners to be ill or injured enough to be taken to a veterinary clinic, and TeamMate recording all abnormalities irrespective of clinical significance. Based on the current data, signs of abnormalities in the hip joints may be quite common in working farm dogs. However, it is not clear whether these abnormalities are commonly associated with clinical disease. More detailed investigation is warranted into whether the recorded hip abnormalities are associated with conditions such as hind limb lameness and osteoarthritis that can impair dogs' welfare and ability to work.

A problem that occurs as a result of our data collection procedure is that we have no way of knowing whether similar abnormalities observed at different points in time are the results of the same or separate injuries or conditions. For this reason, we chose to carry out a descriptive study that focuses mainly on the first occurrence of musculoskeletal abnormalities. While we do report on second occurrence of musculoskeletal abnormalities, we did not calculate IR ratios using this data. As we did not analyze the data longitudinally, we were unable to investigate the effect of time-varying factors such as body weight, body condition, workload, and diet on the risk of dogs developing musculoskeletal abnormalities. These variables may have acted as confounders on the groups we chose to examine here. For example, differences in body weights between sexes and types of dogs may have had an impact on the incidence rates of certain types of abnormalities. Ideally, these variables should have been analyzed using a multivariable modeling approach. Future investigations should examine these risk factors, as they may be useful in determining appropriate husbandry practices necessary to minimize the risk of dogs developing musculoskeletal injury and illness. Future investigation should also examine the effect of musculoskeletal abnormalities on the lifespan and career length of working farm dogs. In combination with the work that has already been carried out, such an investigation will enable veterinarians and dog owners to make decisions about what types of musculoskeletal abnormalities are the most important to prevent and treat in order to ensure that farm dogs lead long and healthy lives.

Due to the fact that data was collected at intervals of several months, we do not have exact data on the time between enrolment and the occurrence of clinical abnormalities, and our calculation of time at risk is an approximation that assumes musculoskeletal abnormalities occurred at the halfway point between examinations. This implies that the recorded musculoskeletal abnormalities occurred evenly distributed between examinations and that they all persisted for long enough to be recorded. However, depending on the type and underlying cause of the abnormalities, they may have occurred at any time after the previous examination and persisted, or they could have occurred within days of the examination and be fully healed shortly after. Additionally, dogs may have sustained and recovered from one or more injuries in between examinations. These injuries would not have been recorded in our data at all. Assuming that recorded abnormalities in our dataset are evenly distributed could therefore be misleading, and we may also have missed a considerable number of less serious injuries. Injuries with a lower or shorter-term impact than those recorded here should not be discounted from a welfare perspective, especially if they are numerous and/or repetitive. Additionally, such injuries could have long-term consequences if they are repetitive and/or cause changes in tissues or joints. However, the abnormalities that we have reported on in this study, while possibly incomplete, still provide information about the types of injuries that occur and could be used to inform decisions around management and veterinary treatment of working farm dogs.

Another potential weakness of the TeamMate study is the reliance on veterinarians' examination notes to code clinical abnormalities. Several veterinarians participated in data collection, and different veterinarians sometimes examined the same dog at different points in time. This created a possibility that different individuals assessed and described similar or identical abnormalities in different ways. However, in order to minimize bias in the data, veterinarians were given training in how to carry out examinations in a standardized way and were asked to describe physical signs rather than to give overall diagnoses. While differences in data collected by different veterinarians are impossible to rule out, we have worked to minimize the risk of bias through our data collection, coding, and data entry procedures.

While there are several weaknesses that limit our ability to draw conclusions from the current study, this is the first time the incidence of musculoskeletal abnormalities has been investigated in working farm dogs. It is our hope that the study will form the basis for future investigation that can help improve the health and welfare of these hard-working dogs.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The animal study was reviewed and approved by Massey University Animal Ethics Committee (protocols 15/26 and 18/53). All dog owners have given oral consent to their dogs being included in the study. Written consent is not a requirement in New Zealand and there are many cases in which projects will be approved without written consent. In this survey verbal consent was considered both acceptable and appropriate: (1) the dog owners had to agree to allow the veterinarian to visit the property, (2) when the veterinarian arrived the dog owners had to consent to them being there, and (3) the owner had to provide the dog to the veterinarian for examination. Further, at each round of data collection dog owners were free to withdraw. Several did withdraw from the study and others did not return phone calls. In terms of the actual process of ethical approval, when the proposal was sent to the Massey University Animal Ethics Committee the method of gaining consent was not included and the Committee did not require the inclusion of this prior to approval.

AUTHOR CONTRIBUTIONS

NC and LL were major contributors to the design of the study. LL and HW organized and contributed to data collection. KEI analyzed the data and drafted the document. NJC, EJN, and NC provided advice on analysis and interpretation of results. All authors provided revisions to the text. All authors read and approved the final manuscript.

FUNDING

Funding was provided by Vetlife (https://vetlife.co.nz/) in the form of time, resources and staff needed to plan the project and carry out data collection. The Massey University Working Dog Centre, The Massey University Graduate Research School and the Massey University School of Veterinary Sciences provided funding in the form of postgraduate research grants. Additionally, The Massey University Working Dog Centre contributed NZ$ 18,000 toward help with data entry and Vetlife received 3-year sponsorships toward the running of TeamMate from Zoetis (NZ$ 10,000 per year), Masterpet (NZ$ 7,500 per year), Virbac (NZ$ 4,000 per year) and Gribbles Veterinary Pathology (15% discount on any laboratory tests for dogs enrolled in TeamMate). Funding bodies had no role in the study design, data collection, analysis, interpretation or writing of the manuscript.

REFERENCES

1. Freeman LM, Abood SK, Fascetti AJ, Fleeman LM, Michel KE, Laflamme DP, et al. Disease prevalence among dogs and cats in the United States and Australia and proportions of dogs and cats that receive therapeutic diets or dietary supplements. *J Am Vet Med Assoc.* (2006) 229:531–4. doi: 10.2460/javma.229.4.531
2. Anderson KL, O'Neill DG, Brodbelt DC, Church DB, Meeson RL, Sargan D, et al. Prevalence, duration and risk factors for appendicular osteoarthritis in a UK dog population under primary veterinary care. *Sci Rep.* (2018) 8:5641. doi: 10.1038/s41598-018-23940-z
3. Mele E. Epidemiology of osteoarthritis. *Vet Focus.* (2001) 17:4–10. Available online at: https://www.ivis.org/library/veterinaryfocus/joint-disease-veterinary-focus-vol-173-oct-2007/epidemiology-of
4. O'Neill DG, Church DB, McGreevy PD, Thomson PC, Brodbelt DC. Longevity and mortality of owned dogs in England. *Vet J.* (2013) 198:638–43. doi: 10.1016/j.tvjl.2013.09.020
5. Bevan S. Economic impact of musculoskeletal disorders (MSDs) on work in Europe. *Best Pract Res Clin Rheumatol.* (2015) 29:356–73. doi: 10.1016/j.berh.2015.08.002
6. Burton W, Morrison A, Maclean R, Ruderman E. Systematic review of studies of productivity loss due to rheumatoid arthritis. *Occup Med (Chic Ill).* (2006) 56:18–27. doi: 10.1093/occmed/kqi171
7. Cogger N, Perkins N, Hodgson DR, Reid SWJ, Evans DL. Risk factors for musculoskeletal injuries in 2-year-old Thoroughbred racehorses. *Prev Vet Med.* (2006) 74:36–43. doi: 10.1016/j.prevetmed.2006.01.005
8. Worth AJ, Sandford M, Gibson B, Stratton R, Erceg V, Bridges J, et al. Causes of loss or retirement from active duty for New Zealand police German shepherd dogs. *Anim Welf.* (2013) 22:166–73. doi: 10.7120/09627286.22.2.167
9. Caron-Lormier G, England GCW, Green MJ, Asher L. Using the incidence and impact of health conditions in guide dogs to investigate healthy ageing in working dogs. *Vet J.* (2016) 207:124–30. doi: 10.1016/j.tvjl.2015.10.046
10. Moore GE, Burkman KD, Carter MN, Peterson MR. Causes of death or reasons for euthanasia in military working dogs: 927 cases (1993-1996). *J Am Vet Med Assoc.* (2001) 219:209–14. doi: 10.2460/javma.2001.219.209
11. Isaksen KE, Linney L, Williamson H, Cave NJ, Beausoleil NJ, Norman EJ, et al. TeamMate: a longitudinal study of New Zealand working farm dogs. I. Methods, population characteristics and health on enrolment. *BMC Vet Res.* (2020) 16:59. doi: 10.1186/s12917-020-2273-2
12. Sheard H. Demographics and health of New Zealand working farm dogs: A survey of dogs on sheep and beef farms in New Zealand in 2009. (2014).
13. Rychel JK. Diagnosis and Treatment of Osteoarthritis. *Top Companion Anim Med.* (2010) 25:20–25. doi: 10.1053/j.tcam.2009.10.005
14. Early JB, Arnott ER, Wade CM, McGreevy PD. Interval dogs: Results and evaluation of Global Positioning System units in measuring athletic performance in stock-herding dogs. *J Vet Behav Clin Appl Res.* (2016) 14:1–4. doi: 10.1016/j.jveb.2016.06.009
15. Cogger N, Sheard H. Working farm dogs. In: *Livestock Production in New Zealand*, Stafford (Auckland, New Zealand: Massey University Press), 280–297.
16. Arnott ER, Early JB, Wade CM, McGreevy PD. Estimating the economic value of Australian stock herding dogs. *Anim Welf.* (2014) 23:189–197. doi: 10.7120/09627286.23.2.189
17. Dohoo I, Martin W, Stryhn H. Measures of disease frequency. In: *Veterinary Epidemiologic Research*. 1st ed. Prince Edward Island: AVC Inc. (2003). p. 65–84.
18. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing (2020). Available online at: http://www.r-project.org/
19. Aicale R, Tarantino D, Maffulli N. Overuse injuries in sport: a comprehensive overview. *J Orthop Surg Res.* (2018) 13:1–11. doi: 10.1186/s13018-018-1017-5
20. von Pfeil DJ., Liska WD, Nelson S, Mann S, Wakshlag JJ. A survey on orthopedic injuries during a marathon sled dog race. *Vet Med Res Reports.* (2015) 2015:329. doi: 10.2147/vmrr.s88276
21. Johnston SA. Osteoarthritis. Joint anatomy, physiology, and pathobiology. *Vet Clin North Am Small Anim Pract.* (1997) 27:699–723. doi: 10.1016/S0195-5616(97)50076-3
22. Sicard GK, Short K, Manley PA. A survey of injuries at five greyhound racing tracks. *J Small Anim Pract.* (1999) 40:428–32. doi: 10.1111/j.1748-5827.1999.tb03117.x
23. Prole JHB. A survey of racing injuries in the Greyhound. *J Small Anim Pract.* (1976) 17:207–18. doi: 10.1111/j.1748-5827.1976.tb06951.x
24. Mey W, Schuh-Renner A, Anderson MK, Stevenson-LaMartina H, Grier T. Risk factors for injury among military working dogs deployed to Iraq. *Prev Vet Med.* (2020) 176:104911. doi: 10.1016/j.prevetmed.2020.104911

25. Bonnett B, Egenvall A, Hedhammar Å, Olson P. Mortality in over 350,000 insured Swedish dogs from 1995-2000: I. Breed-, gender-, age- and cause-specific rates. *Acta Vet Scand.* (2005) 46:105. doi: 10.1186/1751-0147-46-105
26. Asher L, Diesel G, Summers JF, McGreevy PD, Collins LM. Inherited defects in pedigree dogs. Part 1: Disorders related to breed standards. *Vet J.* (2009) 182:402–11. doi: 10.1016/j.tvjl.2009.08.033
27. Cave NJ, Bridges JP, Cogger N, Farman RS. A survey of diseases of working farm dogs in New Zealand. *N Z Vet J.* (2009) 57:305–12. doi: 10.1080/00480169.2009.60926
28. Hughes PL. Hip dysplasia in the New Zealand huntaway and heading dog. *N Z Vet J.* (2001) 49:138–41. doi: 10.1080/00480169.2001.36221

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 19 October 2020
doi: 10.3389/fvets.2020.565346

Fur Color and Nutritional Status Predict Hair Cortisol Concentrations of Dogs in Nicaragua

Grace B. Bowland[1*], *Robin M. Bernstein*[1,2], *Jeremy Koster*[3,4], *Christine Fiorello*[5], *Maris Brenn-White*[5,6], *James Liu*[5], *Laura Schwartz*[5], *Amanda Campbell*[5], *Devin von Stade*[5,7], *Janet Beagley*[8], *Julie Pomerantz*[9], *Alejandro González*[10], *Mackenzie Quick*[5], *Kailyn McKinnon*[5], *Andrea Aghaian*[5], *Corey Sparks*[11] and *Joshua B. Gross*[12]

[1] Anthropology, University of Colorado Boulder, Boulder, CO, United States, [2] Institute of Behavioral Science, University of Colorado Boulder, Boulder, CO, United States, [3] Department of Anthropology, University of Cincinnati, Cincinnati, OH, United States, [4] Department of Human Behavior, Ecology, and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, [5] University of California, Davis, Davis, CA, United States, [6] Saint Louis Zoo, St. Louis, MO, United States, [7] Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO, United States, [8] Independent Researcher, Seattle, WA, United States, [9] Independent Researcher, Castalia, NC, United States, [10] Independent Researcher, Veracruz, Mexico, [11] Department of Demography, University of Texas at San Antonio, San Antonio, TX, United States, [12] Department of Biological Sciences, University of Cincinnati, Cincinnati, OH, United States

OPEN ACCESS

Edited by:
*Cynthia M. Otto,
University of Pennsylvania,
United States*

Reviewed by:
*Nancy Dreschel,
Pennsylvania State University (PSU),
United States
Joseph Wakshlag,
Cornell University, United States*

***Correspondence:**
*Grace B. Bowland
grbo2889@colorado.edu*

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences, a section of the journal Frontiers in Veterinary Science

Received: *25 May 2020*
Accepted: *24 August 2020*
Published: *19 October 2020*

Citation:
Bowland GB, Bernstein RM, Koster J, Fiorello C, Brenn-White M, Liu J, Schwartz L, Campbell A, von Stade D, Beagley J, Pomerantz J, González A, Quick M, McKinnon K, Aghaian A, Sparks C and Gross JB (2020) Fur Color and Nutritional Status Predict Hair Cortisol Concentrations of Dogs in Nicaragua. Front. Vet. Sci. 7:565346. doi: 10.3389/fvets.2020.565346

This study examined the relationships between hair cortisol concentrations (HCC) and sex, age, nutritional status (as determined by body condition scores, or BCS), and body mass (geometric mean calculated from morphometric measurements), as well as the potential influence of hair pigmentation (light, dark, or agouti/mixed) on HCC in dogs of the Bosawas Biosphere Reserve, Nicaragua. The dogs examined in this study live in a marginal environment where disease, malnutrition, and mortality rates are high. For fur color, HCC was significantly higher in light fur than in than dark and mixed fur ($p < 0.001$). In addition, BCS scores were found to have a negative effect on HCC ($p < 0.001$). Measures of sex and body size exhibited inconclusive effects on HCC, and when compared to adult dogs, juvenile dogs did not exhibit significantly different HCC. Repeated measures of dogs over time reveal a moderate intra-class correlation, suggesting that there are unmeasured sources of individual-level heterogeneity. These findings imply a need to account for fur color in studies of HCC in dogs, and the study suggests an overlooked relationship between cortisol and body condition scores in undernourished dogs in diverse settings.

Keywords: hair cortisol concentrations, canine health, canine nutrition, cortisol and metabolism, stress energetics in working dogs, hair pigmentation

INTRODUCTION

Cortisol is a glucocorticoid hormone produced by the neuroendocrine pathways of the hypothalamic-pituitary-adrenal (HPA) axis in response to changes in basal homeostasis. In many mammals, cortisol is responsible for maintaining many of the body's day-to-day and long-term functions, including metabolism, immune system function, growth and reproduction, and sleep/wake cycles (1–4). Cortisol plays a critical role in the body's stress response and increases in cortisol are an adaptive response to aid the body in handling physical or perceived environmental stressors. In the short-term, this response is beneficial. However, long-term cortisol elevation is maladaptive and has been implicated in immune suppression and stunting of growth and development (2–4). Due to the integral role of cortisol to overall health and behavior, analysis of cortisol is often used when assessing the effects of potential stressors (which may range in degree of severity) such as malnutrition, illness, and physical or psychological trauma (5). For both working and companion dogs, cortisol analysis can provide vital insight into physiological health and behavior and allow greater insight into how our canine counterparts adapt to their environment and human-directed tasks.

Studies on the impact of environmental factors such as home life, seasonality, weather, and human-dog interactions on dogs have been conducted through cortisol analysis (6–14). Research has also been conducted in dogs to assess the effect of certain activities such as playing and socializing on cortisol levels, and further research has explored the link between cortisol and lifestyles in working and companion dogs (10, 12, 15).

Associations between cortisol and phenotypic factors, such as body mass, metabolic rate, sex, and age have been explored in domestic canines as well as other mammals. For instance, in a study of 48 German Shepherd and Labrador Retrievers, Sandri et al. (12) found that salivary cortisol concentrations in dogs were inversely related to size. Gillooly et al. (5) confirmed that this trend appears to hold across cortisol-dominant mammal species. Size, defined as body mass, and cortisol have both been related to metabolic rate in mammals. Cortisol plays a regulatory role in metabolic rate, which also decreases with mass (5). In addition, sex-based differences in cortisol levels have been noted in several mammalian species, though these findings have been less consistent. In humans, males tend to have higher cortisol levels on average than females (16) whereas in vervet monkeys this pattern was observed to differ, with males exhibiting lower cortisol levels than females (17). In some cases, research on dogs has failed to find any significant difference in cortisol by sex (11). However, whereas Sandri et al. (12) noted no difference between intact male and female dogs, castrated males and spayed females both showed significantly lower cortisol levels than their intact counterparts. Overall, these studies reveal substantial heterogeneity by sex—both across and within species—which suggests that further research on sex-based differences in cortisol is merited.

A study of salivary cortisol in canines revealed age-related differences in cortisol wherein younger dogs (< 6 months) have significantly lower cortisol levels than their older counterparts (18). Interestingly, these findings conflict with the results of another salivary cortisol study in Dingos, which found puppies had significantly higher cortisol levels than adults (19). The incongruent findings from these studies suggests the potential for developmental differences in cortisol levels in canines, but further research is necessary to elucidate the effects of age.

In addition to size and metabolic rate, nutritional health has also been closely related to cortisol levels in humans and some wild mammals (20, 21). A widely practiced method among veterinarians of assessing body composition in dogs is the use of body condition scoring (22). While often used in industrial settings to assess the presence of excess fat, and the potential risk for a dog to become obese, it can also be used to evaluate the risk or presence of underweight status. Ranging from a minimum value of 1 for malnourished dogs to a maximum value of 9 for obese dogs, veterinarians consider a body condition score (BCS) of 4–5 to be ideal, while values outside this range are potentially worrisome (23). At present, we are unaware of any studies which have explored the relationship between cortisol and BCS scores among dogs.

Hair provides both a non-invasive and reliable method for measuring time-averaged cortisol levels, as the bioavailable (or unbound) fraction of the hormone is incorporated into the hair shaft during hair growth (24). Hair cortisol analysis has been conducted in a wide range of species, including a number of both wild and domesticated animals such as dogs, to assess cortisol variation in relation to stress, growth, and development (6, 11, 12, 24–28). By providing a time-averaged measure, hair samples are distinguished from other sample types, including saliva, blood, and feces, which largely reflect short-term fluctuations in cortisol (29).

The measurement of hair cortisol in domesticated dogs has been validated by numerous studies (9, 24, 30). This body of research suggests that hair pigmentation is a potential confounding factor that should be considered when investigating other predictors. For example, sable fur has been shown to retain significantly lower concentrations of cortisol when compared to yellow, white, or red fur, while agouti fur did not significantly differ from either the dark or light coat coloring (30). The reasons for differences in cortisol levels of differently pigmented hair is not clearly understood but may relate to the physical size of different melanin granules (30).

This paper presents results from hair cortisol analysis of hunting dogs from the indigenous Miskito and Mayangna communities living in the Bosawas Biosphere Reserve of Nicaragua. This population provides a unique opportunity to inform and expand current literature on hair cortisol analysis in dogs and stress energetics. Current literature on the subject of stress in dogs is based on studies with small sample sizes in highly regulated clinical settings in North America or Europe. Furthermore, the dogs included in these studies are primarily purebred. In contrast to these populations, many dogs around the world live in environments with inadequate nutrition and veterinary care (31–33). Little is known about the energetics of stress in these contexts, which are arguably more relevant for understanding the settings in which domestic dogs first evolved. Previous studies in the Bosawas Reserve and analogous settings show that high rates of disease, malnutrition, and injury contribute to high mortality among dogs (34–36). The contrast with purebred dogs in high-income countries is noteworthy (37).

The objectives of this study were to test for the effects of fur color, sex, age, body size, and nutritional status on hair cortisol to assess the robustness of predictions in this population of village dogs from Nicaragua.

METHODS

Ethics Statement

The study was reviewed and approved by the Institutional Animal Care and Use Committee of University of California, Davis. Informed consent was verbally obtained from all of the human participants, as many of the dog owners were functionally illiterate.

Study Site

This study was conducted among the indigenous Mayangna and Miskito residents of the Bosawas Biosphere Reserve in Nicaragua (**Figure 1**). The reserve is characterized by a tropical rain forest biome and harbors a diverse flora and fauna (38–40). The Mayangna and Miskito, who compose the country's most predominant indigenous groups, rely on subsistence practices such as swidden horticulture for most of their nutritional needs (41, 42). Cultivated crops include rice, beans, manioc, and bananas. The Mayangna and Miskito also maintain domestic animals, such as chickens, cattle, and pigs, and they supplement their diets with hunting and fishing, which are the primary sources of dietary protein (43).

Dogs in the Bosawas Reserve are owned by families, who assign them names and provision them with food. The dogs are fed portions of the same foods prepared for their households, and stable isotope analysis suggests that the composition of their diets is largely comparable to the human diet (44). At night, dogs typically sleep in the house, often in the cooking area. Puppies receive affectionate attention, but owners rarely pet their adult dogs, and human-dog relationships differ from the norms in Western contexts (45, 46). Dogs are typically unrestrained and can walk around the community during the day. However, dogs sometimes arouse antagonistic treatment from others in the community, and dogs are seemingly wary of physical abuse when they tread into unfamiliar parts of the community. Notably, owners rarely bathe their dogs, which mitigates concerns that their hair cortisol concentrations are impacted by shampoos or hot water (47, 48).

Dogs are commonly used as hunting companions, and ∼85% of harvested mammals are captured with the assistance of dogs (49). However, there is high heterogeneity in the hunting performance of dogs, and many of the dogs provide little value as hunting companions (50). Veterinary evaluations of the health status of dogs at Bosawas have indicated that malnutrition and dehydration are common in this population (35). Serosurveys also suggest high rates of canine distemper virus, canine parvovirus, *Rickettsia rickettsii*, and *Leptospira* spp (35). The elevated prevalence of disease and injury during hunting contribute to a high mortality rate, and few dogs live beyond 6 years old (35, 43). Dogs within these communities are not spayed or neutered, and routine veterinary care is virtually non-existent. Owners may treat their dogs with antiparasitic medications (e.g., Ivermectin) or antibiotics (e.g., Oxytetracycline) when they suspect infections, but the use of such medications tends to be sporadic. Local residents report that government teams periodically circulate throughout the reserve to administer vaccinations for rabies. See **Supplementary Figures 1–3** for photographs of three dogs in the sample, which provide examples of the categorization of fur color used in this study.

Data Collection

A group of veterinarians examined each dog, documenting variables such as general physical and behavioral attributes, vital measurements, sex, and appearance of health, including any medical problems present. To determine the size of the dogs, morphometric measurements were obtained for height, body length, chest-width, chest-girth, and head-width (51). In assessing the nutritional status of each dog, a numerical value was assigned using the Body Condition System (BCS) scale of 1–9 (**Supplementary Figure 4**). The owner reports of the dogs' ages were subject to measurement error, so as a preliminary measure of age, the data collection team checked dental eruption patterns and noted which dogs appeared to be juveniles (36). Employing a threshold of ∼6 months, we use a binary variable to distinguish the mature dogs from the juveniles in the sample.

Hair samples were also obtained during examination by cutting a small chunk of fur from the coat using scissors or shears. Hair samples were taken from the same location on each dog, specifically the dorsal base of the tail.

Annual exams were conducted during the months of July and August in three field seasons from 2014 through 2016. During this time, ∼750 hair samples were collected, from a total of 580 unique dogs from the communities of Amak, Pulu Was, Ahsa Was, and Wina. Due to occasional errors in labeling of hair samples, insufficient sample weights, and sample loss during processing, seventy-nine samples (∼10%) were unusable (final study $N = 672$, 454 unique individuals). Of the total population of sampled dogs in this study, 213 were female and 240 were male (the sex of one individual was not recorded). When possible, the team attempted to examine and collect hair samples from the same dogs during each of their visits, thus providing a set of longitudinal data on these subjects. However, some dogs were not present for multiple examinations because of mortality or relocations outside the study communities. Of the 454 unique dogs in the sample, 118 dogs contributed two measurements to the compiled sample and 50 dogs contributed three measurements to the compiled sample.

Dog Hair Cortisol Processing

Upon collection, each hair sample was placed in a paper envelope labeled with an identifying number and the date of exam. The samples were then shipped to the Growth and Development Laboratory at the University of Colorado, Boulder for processing and analysis. Samples ranging from 10 to 30 mg of hair were weighed and placed into 2 mL Eppendorf tubes. Weights were recorded along with the corresponding ID number and date, as well as the fur color. Fur color was denoted as either dark (D), mixed (M), or light (L), and categories were defined according to the parameters used by Bennett and Hayssen (30). Each hair sample was then washed a minimum of three times with 1.5 mL isopropanol to ensure all dirt, dead skin, and sebum were removed. Samples were left to air dry under a fume hood for ∼2 days to ensure complete evaporation of the isopropanol. The hair was ground by placing a stainless steel or tungsten carbide ball

FIGURE 1 | Map of the Bosawas Biosphere Reserve. The study communities of Amak, Pulu Was, Ahsa Was, and Wina are circled in red.

in each tube and using a Ball Mill at 25 Hz for 10 min. Samples where hair was not sufficiently ground were placed back on the machine and ground for an additional 5 min. After grinding, 1 mL of methanol was added, and samples were vortexed before being placed on a shaker plate (~100 rpm) at room temperature to extract overnight. The following day, samples were centrifuged for 12 min at 2,500 rpm. Subsequently, 850 uL of supernatant was extracted and transferred to a clean Eppendorf tube. The supernatant was then evaporated using a Microvap nitrogen dry-down evaporator (Organomation, Berlin MA) with tubes in a heated (63°C) block for ∼16 min. If at the end of the drying time some liquid was still present, samples were left for a few more minutes. Samples were then reconstituted using 0.2 mL buffer solution (Salimetrics, Carlsbad CA) and stored at −20°C until processing.

Hair cortisol concentrations were quantified using enzyme linked immunosorbent assay (ELISA- Salimetrics, Carlsbad CA). The use of the Salimetrics kit for canine hair cortisol analysis has been validated by previous studies (8, 30). Results were converted from ng/mL to pg/mg for statistical analysis.

Analysis

HCC was log10 transformed because of its positively skewed distribution. Plots are displayed on the log-transformed scale, though the supplemental material displays plots on the untransformed scale (**Supplementary Figures 5–7**).

As a measure of the overall size of individual dogs, the geometric mean of the combined height, body length, chest width, chest girth, and head width measurement data was calculated[1].

Our analytical strategy involved the specification of multiple regression models, which provide insights into the variance components of the data and the bivariate effects of the covariates. These covariates include fur color, sex, age, body condition scores, and body size (i.e., the aforementioned geometric mean).

Owing to the repeated measure of dogs across field seasons, all models included a random effect (a varying intercept) for the individual dogs. The variance estimates of these random effects can be compared to the residual variance to calculate the intra-class correlation (52, 53). The intra-class correlation, which is the ratio of the random effect variance to the total variance, is interpretable as the expected correlation between successive measurements of the same unique dogs. In addition to the other regression models, we include an "empty" model that has no fixed effect covariates other than the intercept in order to obtain an estimate of the intra-class correlation that is not adjusted for the predictors.

The final model includes all of the covariates, which are expected to exhibit additive effects on HCC.

All analyses were performed using JMP Pro 14 software.

RESULTS

Descriptive statistics are reported in **Table 1**. Results of the statistical analyses are reported in **Table 2**.

Model 0 is the empty model, which reveals an intra-class correlation of 0.27. This correlation is the expected correlation between successive measurements of the same dogs over time (with samples collected ∼1 year apart). Although somewhat modest, the correlation remains largely consistent across all models, which suggests that the correlation stems from unmeasured sources of individual-level heterogeneity.

Model 1 suggests a significant relationship between fur color and HCC, with dark and mixed fur exhibiting significantly lower cortisol levels than light fur (**Figure 2**). When back-transformed, the respective predicted means are 18.53 pg/mg (light fur), 15.78 pg/mg (mixed fur), and 15.60 pg/mg (dark fur).

Model 2 indicates a significant effect of sex, with females exhibiting moderately higher HCC than males ($p < 0.05$). The predicted mean for males is 16.22 pg/mg, and the predicted mean for females is 17.26 pg/mg (**Figure 3**).

Model 3 does not reveal a significant effect for the binary measurement of age ($\beta = -0.034$). This result suggests that juvenile and adult dogs in this population have roughly comparable hair cortisol concentrations.

Model 4 suggests that Body Condition Score has a significant negative effect on HCC ($\beta = -0.03$). For a BCS value of 1, the model predicts HCC of 18.20 mg/pg, as compared to a predicted value of 13.80 for a BCS value of 5 (**Figure 4**). Similarly, as seen in Model 5, body size has a negative effect ($\beta = -0.005$) on HCC (**Figure 5**).

In Model 6, which includes all of the predictors, only fur color and BCS continued to exhibit significant effects. Because

TABLE 1 | Descriptive statistics.

Variable	Description	*N*	Mean	Std. Dev.	Min	Max
Cortisol (pg/mg)		672	17.87	9.62	6.1	121.2
Body Condition Score (BCS)	On a scale of 1–9	663	2.3	0.88	1	6
Size (Geometric Mean) (cm)	Composite of height, body length, head width, chest width, chest girth	579	57.3	7.3	25.1	78.1

Variable	Description	*N*	Proportion
Fur Color		672	
L	Light–yellow, red, or white		0.36
M	Mixed or agouti		0.27
D	Dark or Sable		0.36
Sex		672	
Female	Female		0.47
Male	Male		0.53
Age		663	
Juvenile	>6 months		0.08
Adult	<6 months		0.92

Of the dogs in the sample, 118 dogs contributed two measurements of hair cortisol to the compiled sample, and 50 separate dogs contributed three measurements to the sample. Categorical variables are reported as proportions of the total (owing to rounding error, the proportions for fur color do not sum to 1).

[1] As an alternative measure of body size, we conducted a factor analysis of the morphometric measurements and regressed HCC on the first factor scores. The inferences were largely comparable to the present use of the geometric mean, though there is evidence that the morphometric measurements do not lend themselves to a single latent factor.

TABLE 2 | Multilevel regression model results where hair cortisol [log10(pg/mg)] is the dependent variable, and fur color, sex, body condition score, geometric mean, and a binary measure of age are the independent variables.

Parameter	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Fur (Mixed)		−0.070 (0.015)***					−0.068 (0.016)***
Fur (Dark)		−0.075 (0.014)***					−0.071 (0.016)***
Female			0.027 (0.013)*				0.014 (0.014)
Age (Adult)				−0.034 (0.021)			−0.014 (0.032)
BCS					−0.03 (0.007)***		−0.03 (0.007)***
Geometric Mean						−0.005 (0.002)**	−0.002 (0.002)
Constant	1.222 (0.006)***	1.268 (0.01)***	1.21 (0.009)***	1.253 (0.02)***	1.29 (0.017)***	1.35 (0.042)***	1.397 (0.052)***
Dog ID Variance	0.006 (0.002)**	0.005 (0.002)**	0.006 (0.002)**	0.006 (0.002)**	0.006 (0.002)**	0.006 (0.002)**	0.004 (0.002)*
Residual Variance	0.017 (0.002)	0.017 (0.002)	0.017 (0.002)	0.017 (0.002)	0.017 (0.002)	0.017 (0.002)	0.018 (0.002)
Intra-class correlation	0.27	0.22	0.26	0.27	0.26	0.26	0.20
Observations	672	672	671	663	663	579	566

Standard errors are reported in parentheses.
**, **, ***indicates significance at $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.*
All models included a varying intercept for individual dogs, some of which were sampled multiple times throughout the study.

male dogs are larger than females, it may be that sex and body size exhibit collinearity that reduces the predicted effects of the respective coefficients.

DISCUSSION

In this study of hair cortisol concentrations of hunting dogs living in Nicaragua's Bosawas Biosphere Reserve, the effect of hair pigmentation was found to be statistically significant on reported cortisol levels. Light fur color was associated with significantly higher cortisol concentrations than both dark and mixed fur. In addition, body condition exhibited a negative correlation with HCC. While a bivariate model initially indicated that female dogs show higher cortisol levels than males, sex was not found to be a significant factor in the model including the full set of covariates. Similarly, geometric mean as an indicator of body size shows a negative correlation with cortisol concentrations that is not evident in the model that includes other predictors.

The findings for fur color expand on the previous findings of Bennett and Hayssen (30), which suggested that different hair colors sequester cortisol heterogeneously. We did not attempt to replicate their test of different hair colors from the same dog. Our results, however, further imply that fur color should be a control variable in analyses that use hair samples to examine cortisol concentrations in dogs.

Given the inconsistent findings from previous research on mammals, we did not have strong a priori hypotheses about sex-related differences in HCC among our sample of dogs. Our bivariate analysis indicates that females show higher cortisol than males, but this effect is weaker and inconclusive in models with other predictors. Given the broader literature on the effects of body size on HCC levels, we suspect that the sexual dimorphism of dogs explains this result. As an area for future research, though, the literature would benefit from studies that examine the mechanisms behind sex-related variation in HCC. One possible factor that was not considered in our models is the reproductive status of the female dogs (29). Energetic stress on females during pregnancy is high and continues to be elevated during lactation due to the continued nutritional demand from offspring. To mitigate a potential omitted variable bias, future studies will ideally control for reproductive status.

As with sex, previous findings on cortisol levels and age in canines have proven inconsistent and we did not have a strong prediction as to the link between cortisol and age in our sample population of dogs. Previous findings on cortisol analysis in canine saliva have yielded results indicating juvenile dogs to have significantly differing cortisol levels from adults (18, 19). However, our study indicated no significant difference between juvenile and adult age classes in the Bosawas dog population. These contrasting findings to the prior studies cited may be due to the difference in cortisol analysis methods, as saliva is useful for short-term cortisol analysis whereas hair is beneficial when analyzing cortisol over a period of months.

Partly owing to high levels of malnutrition at Bosawas, we hypothesized that BCS would be negatively associated with the dogs' body condition. The sampled dogs tend to be malnourished, and HCC increases were evident among the dogs in particularly poor condition. This effect remains prominent in a model that includes fur color and other predictors. These findings support our hypothesis that BCS would be negatively correlated with HCC and are congruent with those of prior studies and publications on the energetics of stress. For future research, it would be valuable to determine if this effect remains robust using other methods to operationalize nutritional status.

Prior literature has noted a negative correlation within multiple species between size and cortisol levels where cortisol levels decrease as size increases (5). Therefore, we hypothesized that the geometric mean of morphometric measurements taken from the sampled dogs in Nicaragua would exhibit a negative relationship with HCC. As predicted, our bivariate model indicated that HCC was negatively associated with body size. This effect was weaker and inconclusive in the full model, which again potentially relates to the collinearity between sex, and body size.

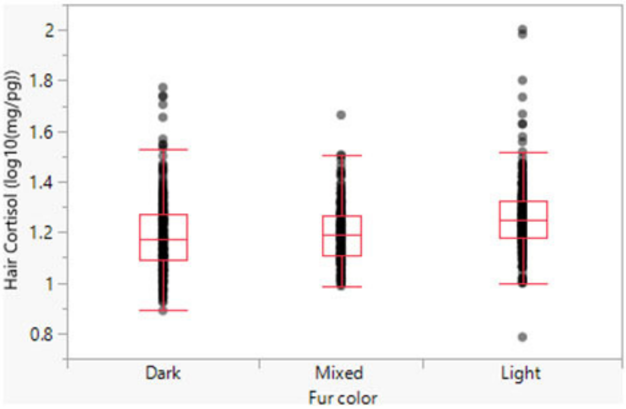


FIGURE 2 | Hair cortisol concentrations as a function of fur color.

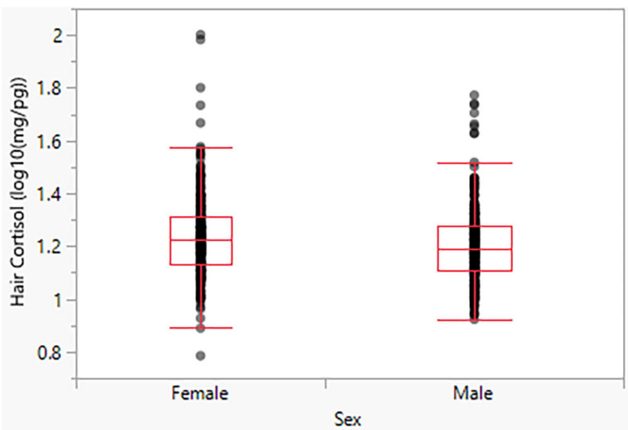


FIGURE 3 | Hair cortisol concentrations as a function of sex.

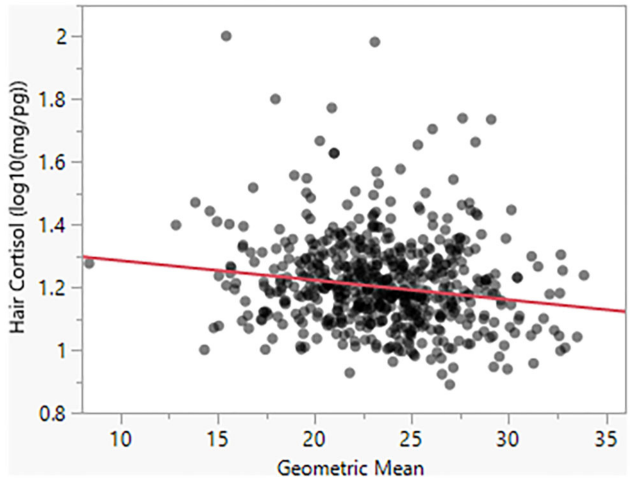


FIGURE 4 | Hair cortisol concentrations as a function of body condition scores.

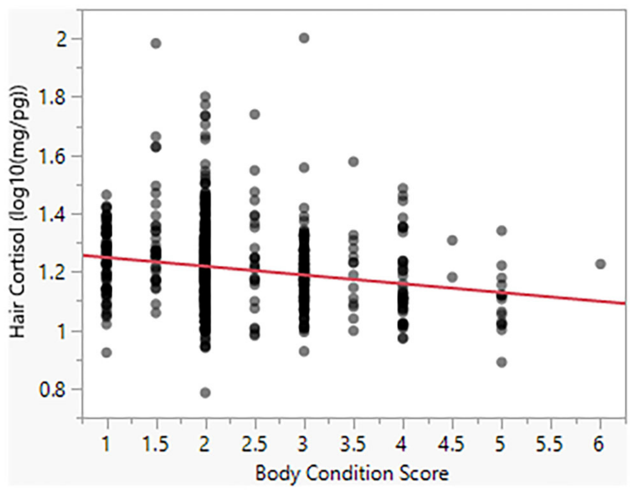


FIGURE 5 | Hair cortisol concentrations as a function of body size, as reflected in the geometric mean of morphometric measurements.

TABLE 3 | Comparison of HCC results from the dogs from this study of the dog population in the Bosawas Reserve and prior studies.

Study	*N*	Range (pg/mg)	Std Dev. (pg/mg)	Mean (pg/mg)
Bennett and Hayssen (30)	47	4.56–27.09	5.45	12.63
Roth et al. (10)	94	~6–54	(N/A)	~15
Bosawas Population	672	6.1–121.2	9.63	17.88

More generally, as a preliminary exploration of how cortisol levels may relate to the environmental differences that distinguish our sample, the results from the dogs in Bosawas were compared to those from previous studies on hair cortisol in dogs. **Table 3** reports the summary statistics from two prior studies in Switzerland and Italy (10, 30) and compares them with those from our study.

The means of the three studies did not differ substantially, and all remained between 10 and 20 pg/mg of hair cortisol. However, the range of HCC in our study was much greater than that of the studies conducted above—over four times as high as the maximum HCC reported by Bennett and Hayssen (30), and twice as high as the maximum reported by Roth et al. (10). In addition to the larger sample size in our study, the differences in reported cortisol concentrations between the Roth et al. (10) study and our study may be due to differences in assay methods (radioimmunoassay vs. enzyme immunoassay), but we note that our methodology very closely followed that of Bennett and Hayssen (30). Therefore, a possible explanation is that the wider range of cortisol concentrations reported in the dogs in Bosawas is due to greater prevalence of physiological or environmental stress experienced by the population in comparison to those of the previous studies on healthy dogs.

The similarity in HCC results between the dogs at Bosawas and those from previous studies might initially seem contradictory but could be due to chronic stress experienced by the Bosawas population. Under extended stress, adrenal fatigue occurs, and the body cannot continue to produce high levels of cortisol (2–4). This makes it difficult to use basal-cortisol levels as an indicator when assessing differences between healthy and chronically stressed populations. In addition, during periods of chronic hypocortisolism, the body's response to a stressor is greatly reduced. Accordingly, it may be more valuable to study the short-term cortisol changes in response to a stressor when assessing overall differences between populations (e.g., through sampling feces or urine). In such studies, populations under higher stress on average would likely show less reactivity to a stressor than a population which does not experience chronic stress.

The findings of this study provide further evidence supporting the link between average cortisol levels within the body and variables related to size and nutritional health. Although the dogs at Bosawas were in considerably poorer health relative to those from previous studies on healthy dogs, relationships between cortisol and metabolic rate still appeared in line with previously observed patterns. The inverse relationship between BCS and HCC levels further supports current understandings regarding the interaction of cortisol with nutritional factors in mammals while expanding it to include domestic canids. Among other considerations, this finding may have implications for further research regarding human domestication of dogs and selective breeding (54). Understanding the possible impacts of human selection for size in dogs may enable a better understanding of the physiological processes at work and the ways in which these changes increase the fitness of dogs.

In conclusion, this work contributes to a growing literature on cortisol levels in animal populations, as assessed via hair samples (29). Our results for fur color and nutritional status are consistent with previous findings. This study also suggests important considerations for future research. Previous work suggests that hair cortisol may vary as a function of variables that were not included in our analysis, such as reproductive status, and the season of year (10, 29). In this tropical setting, there are comparatively modest changes in daylight length and temperature. However, there are pronounced differences in rainfall between the rainy season, when these samples were obtained, and the dry season from January to May (55). It would be worthwhile to investigate how this seasonality impacts activity levels and nutritional status, which could in turn impact HCC. More generally, cortisol levels in dogs have been shown to be responsive to social stressors and related aspects of their day-to-day lives (9, 56). In general, dogs in the Bosawas Reserve are typically allowed to roam freely around the community, and they often accompany their owners on excursions to hunt or work in the horticultural plots. However, the Mayangna are rarely observed to pet their adult dogs or interact with them much outside of rebukes for poor behavior. Investigating variation in cortisol as a function of heterogeneity in activities or contacts with others would be a valuable complement to analogous research in settings that bear little resemblance to the Bosawas Reserve. Despite the logistical challenges of conducting this research, cross-cultural studies have considerable potential to elucidate overlooked aspects of human-dog relationships.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: All data are available as supplemental files on the Open Science Framework: https://osf.io/z7wh9/.

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Care and Use Committee of University of California, Davis. Written informed consent for participation was not obtained from the owners because Informed consent was obtained verbally as many of the dog owners were functionally illiterate.

AUTHOR CONTRIBUTIONS

JK, CF, CS, JG, and RB designed the study. CF, MB-W, JL, LS, AC, DS, JB, JP, AG, MQ, KM, and AA participated in data collection. GB and RB performed sample analysis. MB-W reviewed and organized data. GB and JK performed data analysis. GB, JK, and RB participated in manuscript preparation. All authors contributed to the article and approved the submitted version.

FUNDING

This project was supported by funding from The National Geographic Society, The University of Cincinnati and the Charles P. Taft Research Center.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.565346/full#supplementary-material

REFERENCES

1. Fink G. Stress: definition and history. In: *Encyclopedia of Neuroscience*. Oxford: Academic Press (2009). p. 549–55. doi: 10.1016/B978-008045046-9.00076-0
2. Chrousos GP. Stress and disorders of the stress system. *Nat Rev Endocrinol.* (2009) 5:374–81. doi: 10.1038/nrendo.2009.106
3. Selye H. Physiology of stress. In: Quinn NL, Flagg AL, editors. *Jacqueline Mark-Geraci, Julie Champagne Bolduc, Advanced Clinical Insights*. 5th Ed. Sudbury: Jones and Bartlett Publishing (2002). p. 35–48.
4. Charmandari E, Tsigos C, Chrousos G. Endocrinology of the stress response. *Annu Rev Physiol.* (2005) 67:259–84. doi: 10.1146/annurev.physiol.67.040403.120816

5. Gillooly JF, Haase CG, Long AK. Energetics of stress: linking plasma cortisol levels to metabolic rate in mammals. *Biol Lett*. (2016) 12:20150867. doi: 10.1098/rsbl.2015.0867
6. Beerda B, Schilder MB, Bernadina W, van Hooff JA, De Vries HW, Mol JA, et al. Chronic stress in dogs subjected to social and spatial restriction. II. Hormonal and immunological responses. *Physiol Behav.* (1999) 66:243–54. doi: 10.1016/S0031-9384(98)00290-X
7. Coppola CL, Grandin T, Enns RM. Human interaction and cortisol: can human contact reduce stress for shelter dogs? *Physiol Behav.* (2006) 87:537–41. doi: 10.1016/j.physbeh.2005.12.001
8. Dreschel NA, Granger DA. Physiological and behavioral reactivity to stress in thunderstorm-phobic dogs and their caregivers. *Appl Anim Behav Sci.* (2005) 95:153–68. doi: 10.1016/j.applanim.2005.04.009
9. Gunter LM, Feuerbacher EN, Gilchrist RJ, Wynne CD. Evaluating the effects of a temporary fostering program on shelter dog welfare. *PeerJ.* (2019) 7:e6620. doi: 10.7717/peerj.6620
10. Roth LS V, Faresjö Å, Theodorsson E. Hair cortisol varies with season and lifestyle and relates to human interactions in German shepherd dogs. *Sci Rep.* (2015) 6:19631. doi: 10.1038/srep19631
11. Stephen JM, Ledger RA. A longitudinal evaluation of urinary cortisol in kennelled dogs, Canis familiaris. *Physiol Behav*. (2006) 87:911–6. doi: 10.1016/j.physbeh.2006.02.015
12. Sandri M, Colussi A, Perrotta MG, Stefanon B. Salivary cortisol concentration in healthy dogs is affected by size, sex, and housing context. *J Vet Behav.* (2015) 10:302–6. doi: 10.1016/j.jveb.2015.03.011
13. Ouschan C, Kuchar A, Möstl E. Measurement of cortisol in dog hair: a noninvasive tool for the diagnosis of hypercortisolism. *Vet Dermatol.* (2013) 24:428–31, e93–4. doi: 10.1111/vde.12043
14. Mesarcova L, Kottferova J, Skurkova L, Leskova L, Kmecova N. Analysis of cortisol in dog hair - a potential biomarker of chronic stress: a review. *Vet Med.* (2017) 62:363–76. doi: 10.17221/19/2017-VETMED
15. Slotta-Bachmayr L, Schwarzenberger F. Faecal cortisol metabolites as indicators of stress during training and search missions in avalanche dogs. *Wien Tierärztl Mschr.* (2007) 64:110–17.
16. Binz TM, Rietschel L, Streit F, Hofmann M, Gehrke J, Herdener M, et al. Endogenous cortisol in keratinized matrices: systematic determination of baseline cortisol levels in hair and the influence of sex, age and hair color. *Forensic Sci Int.* (2018) 284:33–8. doi: 10.1016/j.forsciint.2017.12.032
17. Laudenslager ML, Jorgensen MJ, Fairbanks LA. Developmental patterns of hair cortisol in male and female nonhuman primates: lower hair cortisol levels in vervet males emerge at puberty. *Psychoneuroendocrinology.* (2012) 37:1736–9. doi: 10.1016/j.psyneuen.2012.03.015
18. Cobb M, Iskandarani K, Chinchilli V, Dreschel N. A systematic review and meta-analysis of salivary cortisol measurement in domestic canines. *Domestic Anim Endocrinol.* (2016) 57:31–42. doi: 10.1016/j.domaniend.2016.04.003
19. Smith B, Flavel M, Simpson B. Quantification of salivary cortisol from captive dingoes (Canis dingo) in relation to age, sex, and breeding season: implications for captive management. *Austr Mammal.* (2016) 38:21–8. doi: 10.1071/AM15017
20. Tomiyama AJ, Mann T, Vinas D, Hunger JM, Dejager J, Taylor SE. Low calorie dieting increases cortisol. *Psychosom Med.* (2010) 72:357–64. doi: 10.1097/PSY.0b013e3181d9523c
21. Macbeth BJ, Cattet MRL, Obbard ME, Middel K, Janz DM. Evaluation of hair cortisol concentration as a biomarker of long-term stress in free-ranging polar bears. *Wildl Soc Bull*. (2012) 36:747–58. doi: 10.1002/wsb.219
22. Santarossa A, Parr JM, Verbrugghe A. Assessment of canine and feline body composition by veterinary health care teams in Ontario, Canada. *Can Vet J.* (2018) 59:1280–6.
23. Karr-Lilienthal L. *Body Condition Scoring Your Dog - eXtension* (2013).
24. Bryan HM, Adams AG, Invik RM, Wynne-Edwards KE, Smits JEG. Hair as a meaningful measure of baseline cortisol levels over time in dogs. *J Am Assoc Lab Anim Sci.* (2013) 52:189–96.
25. Vincent IC, Michell AR. Comparison of cortisol concentrations in saliva and plasma of dogs. *Res Vet Sci.* (1992) 53:342–5. doi: 10.1016/0034-5288(92)90137-Q
26. Macbeth BJ, Cattet MRL, Stenhouse GB, Gibeau ML, Janz DM. Hair cortisol concentration as a noninvasive measure of long-term stress in free-ranging grizzly bears (*Ursus arctos*): considerations with implications for other wildlife. *Can J Zool.* (2010) 88:935–49. doi: 10.1139/Z10-057
27. Bechshøft T, Sonne C, Rigét FF, Letcher RJ, Novak MA, Henchey E, et al. Polar bear stress hormone cortisol fluctuates with the North atlantic oscillation climate index. *Polar Biol.* (2013) 36:1525–9. doi: 10.1007/s00300-013-1364-y
28. Terwissen CV, Mastromonaco GF, Murray DL. Influence of adrenocorticotrophin hormone challenge and external factors (age, sex, and body region) on hair cortisol concentration in Canada lynx (*Lynx canadensis*). *Gen Comp Endocrinol.* (2013) 194:162–7. doi: 10.1016/j.ygcen.2013.09.010
29. Heimbürge S, Kanitz E, Otten W. The use of hair cortisol for the assessment of stress in animals. *Gen Comp Endocrinol.* (2019) 270:10–17. doi: 10.1016/j.ygcen.2018.09.016
30. Bennett A, Hayssen V. Measuring cortisol in hair and saliva from dogs: coat color and pigment differences. *Domest Anim Endocrinol.* (2010) 39:17180. doi: 10.1016/j.domaniend.2010.04.003
31. Gompper ME. The dog–human–wildlife interface. In: Gompper M, editor. *Free-Ranging Dogs and Wildlife Conservation.* Oxford University Press (2013). p. 9–54. doi: 10.1093/acprof:osobl/9780199663217.003.0001
32. Koster J, Noss A. Hunting dogs and the extraction of wildlife as a resource. In: Gompper M, editor. *Free-Ranging Dogs and Wildlife Conservation.* Oxford University Press (2013). p. 85. doi: 10.1093/acprof:osobl/9780199663217.003.0011
33. Mitchell P. Did disease constrain the spread of domestic dogs (*Canis familiaris*) into sub-saharan Africa? *Azania: Archaeol Res Africa.* (2015) 50:92–135. doi: 10.1080/0067270X.2015.1006441
34. Fiorello CV, Noss AJ, Deem SL. Demography, hunting ecology, and pathogen exposure of domestic dogs in the isoso of bolivia. *Conserv Biol.* (2006) 20:762–71. doi: 10.1111/j.1523-1739.2006.00466.x
35. Fiorello CV, Straub MH, Schwartz LM, Liu J, Campbell A, Kownacki AK, et al. Multiple-host pathogens in domestic hunting dogs in nicaragua's bosawás biosphere reserve. *Acta Trop.* (2017) 167:183–90. doi: 10.1016/j.actatropica.2016.12.020
36. Czupryna AM, Brown JS, Bigambo MA, Whelan CJ, Mehta SD, Santymire RM, et al. Ecology and demography of free-roaming domestic dogs in rural villages near serengeti national park in tanzania. *PLoS ONE.* (2016) 11:e0167092. doi: 10.1371/journal.pone.0167092
37. O'Neill D, Church D, Mcgreevy P, Thomson P, Brodbelt D. Longevity and mortality of owned dogs in England. *Vet J.* (2013) 198:638–43. doi: 10.1016/j.tvjl.2013.09.020
38. Stocks A. Mapping dreams in nicaragua's bosawas reserve. *Hum Organ.* (2003) 62:344–56. doi: 10.17730/humo.62.4.exekjumfqej3vpx4
39. Koster JM. Assessing the sustainability of baird's tapir hunting in the bosawas reserve, nicaragua. *Tapir Conserv.* (2006) 15:23–8.
40. Griffith DM, Coronado I, Asa C, Polisar J, Carmona I. El impacto de los principales usos indígenas del suelo en la estructura, la diversidad y la composición de especies arbóreas en la Reserva de la Biosfera de Bosawás, Nicaragua. *Revis Ecosist.* (2012) 21:101–17.
41. House PR. *Ethnobotany of the Tawahka: agricultural practice and forest management in lowland Central America* (PhD Dissertation). The University of Reading (1997).
42. Cooper CG, Burns JL, Koster JM, Perri AR, Richards MP. Economic and demographic predictors of dietary variation and nutritional indicators in Nicaragua. *Am J Hum Biol.* (2018) 30:e23125. doi: 10.1002/ajhb.23125
43. Koster JM. *Hunting and subsistence among the Mayangna and Miskito of Nicaragua's Bosawas Biosphere Reserve* (Anthropology Thesis). Pennsylvania State University (2007).
44. Perri AR, Koster JM, Otárola-Castillo E, Burns JL, Cooper CG. Dietary variation among indigenous Nicaraguan horticulturalists and their dogs: An ethnoarchaeological application of the canine surrogacy approach. *J Anthropol Archaeol.* (2019) 55:101066. doi: 10.1016/j.jaa.2019.05.002
45. Koster J. Hunting Dogs in the Lowland Neotropics. *J Anthropol Res.* (2009) 65:575–610. doi: 10.3998/jar.0521004.0065.403
46. Gray PB, Young SM. Human–pet dynamics in cross-cultural perspective. *Anthrozoös.* (2011) 24:17–30. doi: 10.2752/175303711X12923300467285
47. Hamel AF, Meyer JS, Henchey E, Dettmer AM, Suomi SJ, Novak MA. Effects of shampoo and water washing on hair cortisol concentrations. *Clinica Chimica Acta.* (2011) 412:382–5. doi: 10.1016/j.cca.2010.10.019

48. Li J, Xie Q, Gao W, Xu Y, Wang S, Deng H, et al. Time course of cortisol loss in hair segments under immersion in hot water. *Clinica Chimica Acta*. (2012) 413:434–40. doi: 10.1016/j.cca.2011.10.024
49. Koster J. The impact of hunting with dogs on wildlife harvests in the bosawas reserve, nicaragua. *Environ Conserv*. (2008) 35:211–20. doi: 10.1017/S0376892908005055
50. Koster JM, Tankersley KB. Heterogeneity of hunting ability and nutritional status among domestic dogs in lowland nicaragua. *Proc Natl Acad Sci USA*. (2012) 109:E463–70. doi: 10.1073/pnas.1112515109
51. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, et al. A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol*. (2010) 8:e1000451. doi: 10.1371/journal.pbio.1000451
52. Hruschka DJ, Kohrt BA, Worthman CM. Estimating between- and within-individual variation in cortisol levels using multilevel models. *Psychoneuroendocrinology*. (2005) 30:698–714. doi: 10.1016/j.psyneuen.2005.03.002
53. Snijders TAB, Bosker RJ. *Multilevel Analysis an Introduction to Basic and Advanced Multilevel Modeling*. Los Angeles, CA: Sage (2012).
54. Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature*. (2013) 495:360–4. doi: 10.1038/nature11837
55. Koster, J, Bruno S, Burns JL. Wisdom of the elders? Ethnobiological knowledge across the lifespan. *Curr Anthropol*. (2016) 57:113–21. doi: 10.1086/684645
56. Packer RMA, Davies AM, Volk HA, Puckett HL, Hobbs SL, Fowkes RC. What can we learn from the hair of the dog? Complex effects of endogenous exogenous stressors on canine hair cortisol. *PLoS ONE*. (2019) 14:e0216000. doi: 10.1371/journal.pone.0216000

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

REVIEW
published: 20 October 2020
doi: 10.3389/fvets.2020.559055

Working Dog Structure: Evaluation and Relationship to Function

Chris Zink[1]* *and Marcia R. Schlehr*[2]

[1] *Zink Integrative Sports Medicine, Ellicott City, MD, United States,* [2] *Independent Researcher, Clinton, MI, United States*

Working dogs help to keep society and individuals safe, secure, and healthy. To perform their varied functions, it is critical to select dogs that are structurally sound and capable of demonstrating power, coordination and agility. Characteristics such as size and substance, head and axial skeletal structure, chest size and conformation, and thoracic and pelvic limb angulation should be evaluated to select the optimal combination of characteristics to suit the tasks to which each dog will be assigned. This review provides guidance on how to evaluate each of these structural components and discusses the contributions of those body parts to a working dog's function.

Keywords: working dog, structure, function, evaluation, assessment, power, coordination, agility

OPEN ACCESS

Edited by:
Nathaniel James Hall,
Texas Tech University, United States

Reviewed by:
George E. Moore,
Purdue University, United States
Janice Lauren Baker,
Veterinary Tactical Group,
United States

***Correspondence:**
Chris Zink
cz@caninesports.com

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences,
a section of the journal Frontiers in Veterinary Science

Received: *05 May 2020*
Accepted: *03 September 2020*
Published: *20 October 2020*

Citation:
Zink C and Schlehr MR (2020) Working Dog Structure: Evaluation and Relationship to Function. Front. Vet. Sci. 7:559055. doi: 10.3389/fvets.2020.559055

INTRODUCTION

There are many different types of working dogs – dogs with jobs that help to keep society and individuals safe, secure, and healthy. Some of these dogs work as military, police, search and rescue, detection (bombs, drugs, cash, agricultural products, termites, mold, cancer, etc.) dogs. Others have jobs as dog guides for the blind, hearing assistance dogs, assistance dogs for the disabled, and work in many other capacities to help their human partners. In this review, discussion will be limited to working dogs that help communities, as opposed to assisting individuals. The majority of these dogs work for government institutions, such as the military, police forces, the Transportation Security Administration, Customs and Border Protection, and agriculture defense dogs. These dogs will be referred to using the upper case designation Working Dogs.

Centuries ago, most selective breeding strategies had the goal of producing dogs to assist with specific tasks that helped humans survive and thrive, such as hunting, herding, or capturing vermin. However, in the last 150 years, this tight relationship between structure and function has, in many cases, dissolved as people began to breed specifically for success in the conformation ring, where dogs are judged predominantly on appearance. At the same time, some individuals chose to breed those same breeds strictly for performance competitions, often leading to distinct differences in the structure of performance and conformation lines of the same breed. This has progressed almost to the point where the performance and conformation lines of many breeds have few structural similarities. These differences in structure between different lines/functions within a breed are perhaps most noticeable in the German Shepherd Dog, the Labrador Retriever, the Golden Retriever, and the Border Collie, breeds that are often recruited for use as Working Dogs.

The detailed anatomy of all dogs, including the bones, muscles, tendons, ligaments, innervation, and vasculature is the same (1). However, the ways in which those components vary and are combined in each breed, resulting in their size and shape, constitute structure. Dogs have the greatest morphological diversity of all mammals (2). Further, the cranial and limb morphology of *Canis familiaris* are more variable than in all of the other canid species combined (3, 4). Those differences arise from the functions for which each breed was originally developed, combined with features selected for by the dog fancy throughout the 20th and 21st centuries. How that structure relates to function in Working Dogs is the subject of this review.

The structural requirements of today's Working Dogs are quite varied because these dogs perform such a wide variety of functions. Working Dogs might need the strength to undergo sudden acceleration to their maximal speed or to leap over a tall barrier, but they might also require the physical stamina to stand or walk all day long. Working Dogs might need to search over rubble or in difficult environmental circumstances such as blistering heat or icy, freezing conditions, often wearing heavy body armor. They also might spend the day detecting specific scents amongst thousands of others, requiring intense mental concentration, which can be physically exhausting. Indeed, several differing functions might be required in the same dog. Each working task requires specialized training and activities that place different and often extreme physical demands on the dogs.

STRUCTURE-FUNCTION RELATIONSHIPS

Given the many and varied tasks of Working Dogs and the wide variety of structures of different dog breeds, it is important to develop a deeper understanding of structure-function relationships in these dogs. An intensive examination of the peer-reviewed literature on canine structure-function relationships reveals that there are a few specific areas of intense focus, such as examination of the relationships between tibial plateau angle and cranial cruciate ligament insufficiency (5) and of femoral trochlear groove structure and patellar luxation (6).

However, there is a dearth of peer-reviewed publications that discuss overall canine structure and its relationship to function. This problem perhaps arises exactly because dogs vary so greatly in their structure. When planning an experiment that will correlate a canine structural component with its function, where does one start? Many publications use the Labrador Retriever as an example of the "standard dog," yet the Labrador Retriever's structure (pelvic limb angulation, e.g.,) varies significantly from that of the German Shepherd Dog. Nonetheless, both of these breeds feature prominently as Working Dogs. And within those breeds, individual structural variation can be dramatic.

Our understanding of the biomechanics of movement, of bone leverage and of muscle/tendon/ligament dynamics, while currently incomplete, is constantly evolving with the use of new technology such as body-worn accelerometers, video and animation technology, and 3D printing based on CT data (7, 8). Still, many existing studies using these technologies to examine structure-function relationships use samples of only 3 to 4 dogs. For example, an outstanding study that examined 3D kinematics of just the canine pelvic limb in only 4 breeds of dogs selected for their functional differences (speed vs. strength) produced a huge amount of data (8). It is hard to imagine that the same study could have included a larger number of breeds. A number of studies of racing greyhounds have provided us with new information on the relative importance of thoracic and pelvic limb musculature for speed in this breed (9, 10), and additional studies have compared structure-function relationships in speed vs. strength breeds (11, 12). There are two outstanding studies that examine the relationships between structure (length vs. cross-sectional area) and function of perivertebral and neck musculature in dogs, and their findings are likely applicable to most dog breeds because to the best of our knowledge all dogs have the same muscles (13, 14). By far, the most comprehensive and scientifically based treatise on the subject of structure-function relationships in dogs is the text *Dogs in Motion* by Martin S. Fischer and Karin E Lilje (15). These authors studied kinematics and kinetics using high frequency videography, marker-based movement analysis, force plates, and biplanar X-ray videography in 327 dogs of 32 different breeds, an outstanding feat of biology and engineering. Increasing numbers of publications can be expected in the future as new technology combined with the ability to analyze extremely large data sets improves.

In general, specific breeds and cross-breeds of Working Dogs are selected because of their trainability, instincts, and temperament for the desired tasks, as well as their size and threat potential, providing a visible deterrent to crime. However, there is minimal evidence-based information regarding what specific structures are most desirable for a Working Dog to have superior abilities and a long and healthy career. For example, what pelvic limb angulation (a term for the combination of angles at which the pelvis, femur, tibia/fibula and metatarsal bones naturally meet in the standing dog) is ideal for superior performance as well as health and longevity in a police dog that needs to perform optimally during a full day of both apprehension and detection? What combination of body size, length, height and muscularity is ideal for a military dog that will be transported by helicopter to hot, dry environments to detect explosives for several hours a day?

This review discusses knowledge derived from peer-reviewed publications where that information exists. However, to fill in the significant gaps in our scientific knowledge, this review also depends on the observations of experienced dog breeders and judges regarding structure-function relationships. These are often based on decades of personal experience and observations of the effects of selective breeding over centuries. Many of these structure-function relationships are described in the breed standards, which are written and sometimes also illustrated descriptions of the ideal dog of each breed. The breed standards are established by individuals with decades of experience in the breed who are considered guardians of those breeds. Breed standards are often considered sacrosanct and are not modified without significant consideration and input from individuals experienced with the breed's structure and original functions. Table 1 provides quotes from the breed standards of the German Shepherd Dog, the Belgian Malinois and the Labrador Retriever that describe the breeds' overall structure as they relate to function. This review discusses structure-function relationships in these three breeds because they are the most common breeds used as Working Dogs. However, it is also important to recognize that there are other Working Dog breeds that have been selected for specific functions, such as the Beagles used at airports and shipping ports to detect illegally imported agriculture products or pests, and thus have different size and structure.

In general, there are two ages at which dogs are selected for careers as a Working Dogs. Puppies are often selected for future careers as Working Dogs when they are ready to leave the breeder, usually at around 8 weeks of age. Breeders and judges of canine structure have long observed that structural evaluation of puppies at 8 weeks of age most accurately predicts adult structure. One all-breed judge who has evaluated thousands of dogs as puppies and again as adults has described her procedure for the structural evaluation of puppies (16).

A second age at which Working Dogs are selected is late adolescence or young adulthood. Government agencies frequently purchase young adult, partly trained Working Dogs because at this age, dogs are already demonstrating their working temperament and many of their adult structural features.

SIZE AND SUBSTANCE

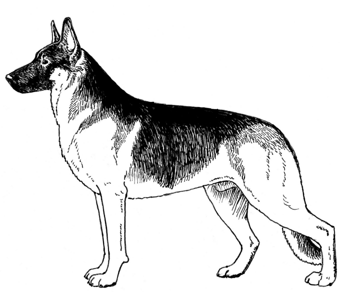
When evaluating canine structure, it is important to have the dogs positioned in a standardized stance that allows comparison between individuals. In this review we will use the position in which dogs are stood (or *stacked*) to be structurally evaluated in conformation shows. In that stance, the radius and ulna of both thoracic limbs are placed perpendicular to the ground, the metatarsals are placed perpendicular to the ground, and the head is held up with face looking forward (**Table 1**).

It is essential for Working Dogs to have sufficient size and substance to be able to carry out their various functions. For example, during acceleration, the greatest amount of power in the pelvic limb occurs at the coxofemoral joint (10). These forces require not only stable coxofemoral joint conformation, but also optimal development of the muscles that power hip movement. Sufficient size and substance are necessary to produce this power.

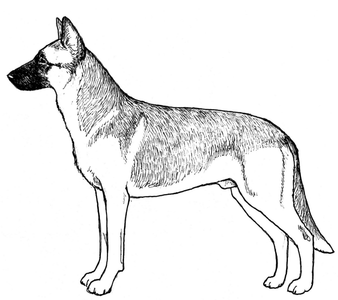
When discussing size and substance, the following components are considered: height, generally measured from the ground to the top of the scapula (the *withers*); body length, usually measured from the cranial aspect of the manubrium (the *prosternum*) to the caudal aspect of the ischiatic tuberosity

TABLE 1 | Structure-Function Components for Breed Standards for Three Working Dog Breeds.

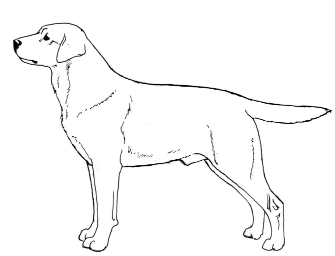
German shepherd dog (effective 1978)	"The first impression of a good German Shepherd Dog is that of a strong, agile, well muscled animal, alert and full of life. It is well-balanced, with harmonious development of the forequarter and hindquarter. The dog is longer than tall, deep-bodied, and presents an outline of smooth curves rather than angles. It looks substantial and not spindly, giving the impression, both at rest and in motion, of muscular fitness and nimbleness without any look of clumsiness or soft living." "The breed has a distinct personality marked by direct and fearless, but not hostile, expression, self-confidence and a certain aloofness that does not lend itself to immediate and indiscriminate friendships. The dog must be approachable, quietly standing its ground and showing confidence and willingness to meet overtures without itself making them. It is poised, but when the occasion demands, eager and alert; both fit and willing to serve in its capacity as companion, watchdog, blind leader, herding dog, or guardian, whichever the circumstances may demand." "The ideal dog is a working animal with an incorruptible character combined with body and gait suitable for the arduous work that constitutes its primary purpose."
Belgian malinois (effective 1990)	"The Belgian Malinois is a well-balanced, square dog, elegant in appearance with an exceedingly proud carriage of the head and neck. The dog is strong, agile, well-muscled, alert, and full of life. He stands squarely on all fours and viewed from the side, the topline, forelegs, and hind legs closely approximate a square. The whole conformation gives the impression of depth and solidity without bulkiness."
Labrador retriever* (effective 1994)	"The Labrador Retriever is a strongly built, medium-sized, short-coupled, dog possessing a sound, athletic, well-balanced conformation that enables it to function as a retrieving gun dog; the substance and soundness to hunt waterfowl or upland game for long hours under difficult conditions; the character and quality to win in the show ring; and the temperament to be a family companion. Physical features and mental characteristics should denote a dog bred to perform as an efficient Retriever of game with a stable temperament suitable for a variety of pursuits beyond the hunting environment. The most distinguishing characteristics of the Labrador Retriever are its short, dense, weather resistant coat … a clean-cut head with broad back skull and moderate stop; powerful jaws… The typical Labrador possesses style and quality without over refinement, and substance without lumber or cloddiness. The Labrador is bred primarily as a working gun dog; structure and soundness are of great importance."



German Shepherd Dog



Belgian Malinois



Labrador Retriever

**Quotes are from the American Kennel Club breed standards. The breed standards of these three breeds can be found at https://images.akc.org/pdf/breeds/standards/GermanShepherdDog.pdf, https://images.akc.org/pdf/breeds/standards/BelgianMalinois.pdf, and https://images.akc.org/pdf/breeds/standards/LabradorRetriever.pdf, respectively. Breed standards of the Canadian Kennel Club, The Kennel Club of the United Kingdom, and other registries may differ slightly and can be obtained online.*

(**Figure 1**); the relative proportions of the thoracic, lumbar and pelvic components of the body; the dog's weight. Most Working Dogs range in height from 21.5" (53.75 cm) to 26" (65 cm) at the withers, with females generally ∼2" (5 cm) shorter than males. The German Shepherd Dog and Labrador Retriever standards both state that the body length should be slightly longer than the height. In contrast, the Belgian Malinois states that these two lengths should be equal. With respect to weight, most Working Dogs lie in the range of 50–80 lb (23–36 kg).

A Working Dog needs to be tall enough to be able to walk at a speed consistent with that of its handler, to run at speeds necessary for chase and apprehension, and have sufficient substance to be able to constitute a substantial threat and stop a fleeing person if necessary. However, moderation in size and substance are also important. All other things being equal, a heavier dog is unlikely to run as fast or have the same endurance as a lighter dog of the same height [**Figure 2**; (17)]. By the same token, a dog that is lacking in substance might not have the muscular strength to apprehend a large man or to carry the weight of equipment and/or an armored vest throughout an active day. Most individuals of the German Shepherd Dog, Belgian Malinois and Labrador Retriever breeds have balanced combinations of size and substance sufficient to carry out their functions as Working Dogs, although particularly heavy-set or small, weedy individuals should be avoided during the selection process.

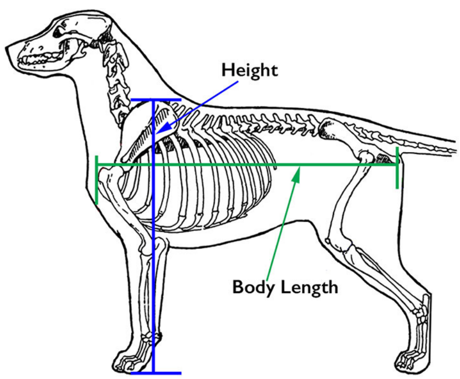


FIGURE 1 | Customary sites for measurement of body length and height. Illustration by M. Schlehr.

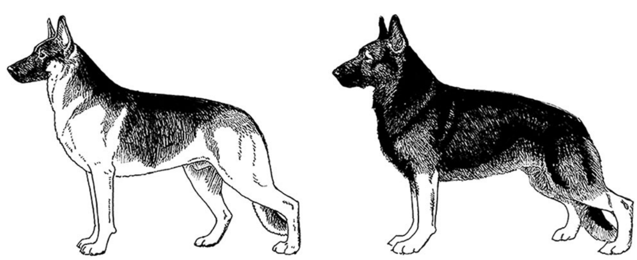


FIGURE 2 | Two German Shepherd Dogs of the same height and overall structure, but with different size and substance. Because of its heavier substance, the dog on the right would be physically less suitable for tasks involving speed and endurance. Illustration by M. Schlehr.

Working Dogs should have a large chest for sufficient lung capacity, but the chest should not be so wide as to interfere with gait, as happens with bulldogs, for example (8). Therefore, it is desirable to have the rib cage occupying a large part of the body length, certainly more than half and probably closer to two-thirds of the distance between the manubrium and the ischiatic tuberosity, and to have sufficient depth of chest as well. The lumbar component of the spine provides for a great deal of spinal flexibility, both dorsoventral and lateral, but should also be well-muscled to prevent hyperflexion, particularly during sudden or unexpected movements as can happen during apprehension. Paraspinal and abdominal muscles should be firm to the touch on the standing dog.

Neck and Topline

The *topline* consists of the upper profile of the dog from the top of the head to the base of the tail. The neck and topline reflect the positioning of the axial skeleton, which supports the ribcage and pelvis and forms a structure for the attachment of the thoracic and pelvic limbs. The axial skeleton is wrapped in the core musculature, which is critical for all components of movement (14). Most students of canine structure and function believe that the neck should be of medium length (18). A long, thin neck lacks the strength to carry heavy objects or support and stabilize the dog during apprehension. A short neck will prevent the full use of the head as a counterbalance and can inhibit thoracic limb movement. The neck should merge with the shoulders gradually; an abrupt junction between neck and shoulders is believed by those who study canine structure to indicate less than ideal shoulder structure.

The *backline* is the part of the topline from the withers caudally. This should be strong and level in Labrador Retrievers and Belgian Malinois and slightly sloping from cranial to caudal in German Shepherd Dogs from working lines; this slope can be very extreme in German Shepherd dogs from conformation lines. The effects of this extremely sloped topline on the dog's strength and mobility have not been objectively studied. A topline that sags in the middle (lordosis) usually indicates weak core (paraspinal and abdominal) musculature but can also be indicative of abnormal vertebral structure. A kyphotic (roached) back is often an indicator of pain, although many German Shepherd Dogs that are bred for conformation have this structure. The effects of this altered axial skeletal conformation on function also have not been objectively studied. Note that all dogs have a normal, small dip in their topline at T11. The dorsal spinous processes of the cervical vertebrae and first 10 thoracic vertebrae are pointed dorsocaudally, while those of the vertebrae caudal to T11 are pointed dorsocranially, and the dorsal spinous process of T11 (the anticlinal vertebra) is very short to accommodate this change in direction of the spinous processes, creating a slight depression.

THORACIC LIMB STRUCTURE

There are few peer-reviewed publications on the relationships between canine thoracic or pelvic limb structure and function. As a result, we are dependent on the observations of individuals who have spent decades observing dogs and correlating structure with efficiency of movement and performance longevity. Three of these individuals have published their observations in excellently illustrated books (19–21). The discussions of structure presented here represent an amalgam of their observations, with the results of peer-reviewed publications inserted where such exist.

Thoracic Limb Angulation – Side View

When evaluating structure of the thoracic and pelvic limbs, it is important to be able to observe and/or palpate the bones as they exist under the skin and soft tissues. The term *thoracic limb angulation* is used by those who study canine structure to describe the angle at which the scapula lies off of vertical and the angles at which the scapula, humerus, and radius and ulna meet at the shoulder and elbow joints, respectively, when the dog is standing in the standard, stacked position. Together, these angles help to determine the ability of the thoracic limb to carry out all of its functions in both moving and stabilizing the body. In most breeds the thoracic limbs bear approximately 60% of the dog's weight when standing, walking and trotting, and they bear the entire weight of the dog in addition to the effects of gravity when the dog is landing from a jump and when the thoracic limbs are bearing the dog's weight during the gallop. The thoracic limbs also function in providing lift at the initiation of jumping. Although in the past, the thoracic limbs were thought to function more for stabilization than forward drive in the moving dog, recent studies suggest that they also play an important role in providing propulsion for forward motion (12).

When the dog is moving forward, abundant thoracic limb angulation, along with optimal musculature, allows the limb to unfold and reach well forward ahead of the dog, to pull the dog's body forward while supporting its weight. Correct angulation and strength also permit the thoracic limbs to extend far caudally, allowing for a long stride length, and to provide lift before beginning the swing phase of the stride in which the dog again reaches forward. Since taking short or long strides require approximately the same amount of energy, it is an advantage to take fewer strides when moving from A to B. At the same time, taking excessively long strides can reduce stability, since stability decreases the further the paw is from the center of gravity. As is so often the case, a balance between stability and forward motion is required.

Thoracic limb angulation is most readily evaluated by assessing two specific features: the angle at which the scapula lies off of vertical and the relative length of humerus, which also determines the angles at the shoulder and elbow joints (19, 20). Scapular angle and length of humerus appear to have different inheritance. Together, they significantly affect the efficiency of thoracic limb function.

Angle of Scapula

Movement of the shoulder blade along the rib cage makes up at least 65% of the stride length in dogs (15). A lack of bony attachment of the scapula to the axial skeleton provides for increased range of motion of the thoracic limb, and the angle at which the scapula lies against the ribs is an important factor in allowing extension of the shoulder joint and thus free movement of the entire thoracic limb. To evaluate thoracic limb angulation, the dog should be positioned in the stacked position with the radius and ulna perpendicular to the ground, the metatarsals perpendicular to the ground, and the head held up and muzzle approximately parallel to the ground. This standardized position allows angulation of the thoracic limb to be evaluated in a consistent manner and permits dog-to-dog comparisons.

The angle at which the scapula lies off of vertical is also referred to as *shoulder layback*. It is determined by imagining a line perpendicular to the ground that passes through the cranial-most aspect of the greater tubercle of the humerus, then imagining another line that starts at the cranial aspect of the greater tubercle of the humerus and extends to the dorsal-most aspect of the dorsal rim of the scapula (**Figure 3**). This angle at which these two lines meet ideally is 30° based on cineradiographic imaging studies (20). Many books and breed standards describe the correct angle of scapula as 45°, but without objective substantiation (22). This angle can be assessed, with some difficulty, using a goniometer with one arm laid along a straight edge set perpendicular to the ground and abutting the cranial aspect of the greater tubercle of the humerus, and the other arm extending from the cranial aspect of the greater tubercle of the humerus to the dorsal-most part of the dorsal rim of the scapula. More often, the scapular angle is assessed subjectively by placing the thumb and index finger of one hand on the dorsal-most aspect of the two scapulae and comparing how far caudally they are positioned relative to other individuals of the same or other breeds. The more caudally the dorsal-most aspect of the rim of the scapula is positioned, the greater the angle of the scapula. Sufficient angle of scapula is desirable because it allows greater shoulder joint extension and thus more forward reach of the thoracic limb.

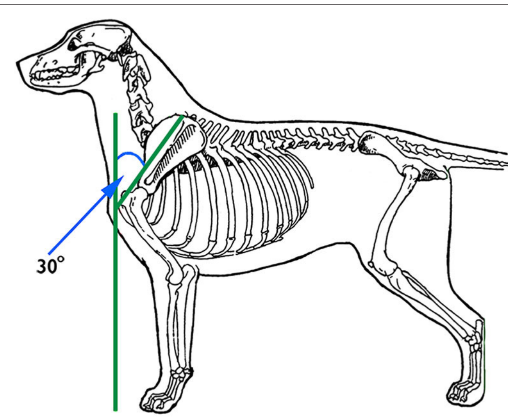


FIGURE 3 | The angle of the scapula is determined by imagining a line perpendicular to the ground that passes through the cranial-most aspect of the greater tubercle of the humerus, then imagining another line that starts at the cranial aspect of the greater tubercle of the humerus and extends to the dorsal-most aspect of the dorsal rim of the scapula. Illustration by M. Schlehr.

Dogs with greater angle of scapula tend to have more developed shoulder muscles, particularly the supraspinatus, infraspinatus and triceps muscles. This might be because these three muscles support the shoulder joint in its angled state in the standing dog. If the scapula lies in a more vertical position, the bones might play a larger role in support. Dogs with a greater angle of scapula are thought to experience less concussion on the shoulder joint particularly when landing with the limb in extension, such as when landing from a jump or when the dog is in a gallop. This is because the well-angled shoulder with greater shoulder muscle strength and greater length of the muscle/tendon units can better flex to absorb the shock of landing and elongate to withstand eccentric contraction of the supraspinatus and biceps muscles as the dog's body falls forward. Resistance to injury from eccentric contraction of these muscles is important given that tendinopathies of these two muscles are amongst the most common injuries in active dogs (23).

Length of Humerus

A second structural variable of the canine thoracic limb is the length of the humerus, which largely determines the angles of the shoulder and elbow joints. Ideally, the humerus should be long enough to place the dog's radius and ulna in a caudal position, where it can help to support the weight of the chest, when the dog is standing with the radius and ulna perpendicular to the ground. It has been observed by those who study canine structure, that in a dog with an optimal length humerus, a line from the dorsal rim of the scapula to the cranial aspect of the greater tubercle of the humerus is equal in length to a line drawn from the cranial aspect of the greater tubercle of the humerus to the olecranon process in the standing dog (**Figure 4**). These lengths are easily measured using a flexible tape measure.

Another way that dog breeders and judges evaluate humeral length is to imagine a line drawn perpendicular to the ground through the center of the radius and ulna on a stacked dog. This line should intersect with the dog's topline at the junction of the neck and the back (the *withers*). In a dog with a short humerus, the distal thoracic limb is positioned more cranially, resulting in a line that intersects the topline further cranially along the neck (**Figure 5**).

Dogs with a short humerus have less acute angles at the shoulder and elbow. This might be the reason why it has been observed that these dogs tend to have less well-developed thoracic limb musculature, since they do not have to support these joints in a more angled position. Logically, this would produce more concussion on the bones of these two joints during movement, and more stress on the extensor muscles for these joints during eccentric contraction. To the extent that scapular angle and/or humeral length deviate from ideal, thoracic limb function will be compromised.

Limb angulation does not remain static throughout a dog's life; it changes in response to injury and level of fitness. Dogs with injuries to the thoracic or pelvic limb often experience disuse muscle atrophy. As a result, they frequently stand with less acute angles of the joints, letting the bones stacked one above the other take over more of the function of supporting the limbs. In addition, since it takes muscular effort to support a well-angulated limb, if a dog is not optimally fit, it will have less than optimal thoracic and/or pelvic limb angulation. Measuring the degree of limb angulation when the dog is standing naturally is one way to monitor progress during rehabilitation.

Thoracic Limb - Front View

For the thoracic limbs to function optimally in movement, they have to be able to grip a stable substrate (usually the ground) and then use muscular strength to transfer power along the length of the limb in a sagittal plane to propel the body (9). The most effective way to transfer power is in a straight line. As a result, the standing dog's thoracic limbs, when viewed from the front, should form a straight line perpendicular to the ground from the foot to the body, with minimal bend at carpus or elbow, as demonstrated by the dog on the left in **Figure 6.** When the

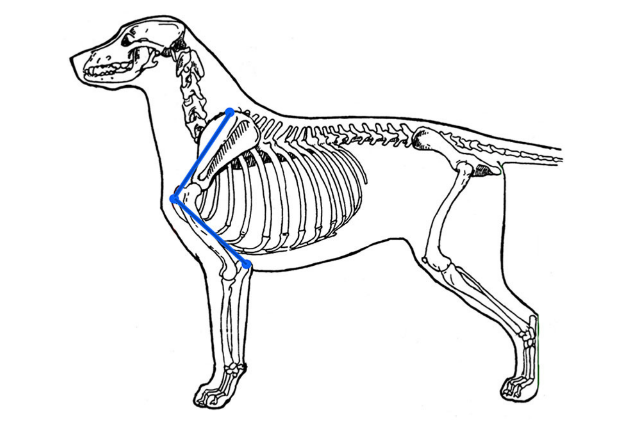


FIGURE 4 | For ideal humeral length, a line drawn from the olecranon process to the cranial aspect of the greater tubercle of the humerus should be the same length as one drawn from the cranial aspect of the greater tubercle of the humerus to the dorsal-most aspect of the scapula. Illustration by M. Schlehr.

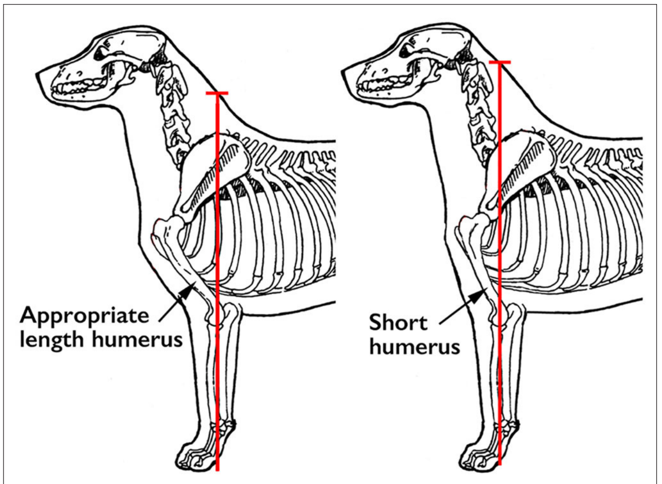


FIGURE 5 | Ideally, a line drawn perpendicular to the ground through the center of the radius and ulna should intersect with the topline at the junction of the neck and back (the withers). Illustration by M. Schlehr.

thoracic limbs are not straight when viewed from the front, as in the dog on the right in **Figure 6**, which demonstrates bilateral carpal valgus deformities, the same amount of muscular effort results in dissipation of the power, reducing the effect of the power output on movement. In addition, this can increase stress on the lateral and medical collateral ligaments and tendons that support the joint.

In the relaxed standing dog, it is normal for the thoracic limbs to be slightly externally rotated resulting in the feet being positioned with the toes pointing slightly laterally (**Figure 7**). It is believed that this rotation provides stability for the standing dog, just as a lateral position of the feet do in the standing human, and should not be confused with the valgus deformity seen in the right panel of **Figure 6**. When the dog in **Figure 7** gaits, the thoracic limbs rotate on their axes, thanks to the rotational movements of the radius and ulna, and the feet strike the ground with the toes pointing cranially and no bend at the carpus, providing the most efficient grip on the ground and transfer of power to the body. In contrast, when a dog with valgus deformity gaits, the foot remains externally rotated and the carpal deformity persists.

Feet and Dew Claws

The shape of dogs' feet varies depending on the dog's original function. Dogs that were bred to move over rocky or uneven ground, tend to have compact feet (termed *cat feet*). Cat feet have toes that are all of equal length, forming a half-circle around the central pad (**Figure 8**). These feet are often considered analogous to the knobby tires of an ATV, that are designed for improved grip on uneven surfaces. Indeed, many breeds with cat feet were bred to be agile moving in all directions or over rough ground. A typical instance is the Afghan hound, which was bred to hunt agile prey over rocky ground.

In contrast, dogs that were bred to run fast in relatively straight lines, such as the Greyhound, tend to have a more elongated foot shape (termed *hare feet*). In these feet, the first and second phalanges of the third and fourth digits are longer than those of the second and the fifth digits, so those toes are longer. An elongated foot is thought to provide an advantage when running straight ahead and is somewhat analogous to the slick tires of a race car, which provide additional grip for forward motion.

The superficial digital flexor tendon inserts on the distal second phalanx of each toe, so the dog's toes are spring-like, allowing for improved impact absorption. Repetitive strain to the superficial digital flexor tendon of one or more toes can cause permanent lengthening of these tendons. This results in an increase in the angle of extension at the carpus, and flattening of the phalanges, reducing the ability of the carpus and feet to absorb impact. The breed standards for all three breeds under consideration in this review call for compact feet, and the Belgian Malinois standard specifically states, "The feet are round (cat footed) and well-padded with the toes curved close together" while the rear feet "may be slightly elongated." Practically speaking, however, most German Shepherd Dogs today tend to have excessive angle of extension at the carpus, and elongated, rather than round, feet. In those German Shepherd Dogs that have more thoracic and pelvic limb angulation, the toes are often, but not always, splayed (**Figure 9**), although an individual of any breed can have splayed feet. This is thought to reduce the ability of the toes to work as a unit and to increase the risk of toe injuries, since a single toe can be more easily separated from the others, resulting in medial or lateral collateral ligament sprain.

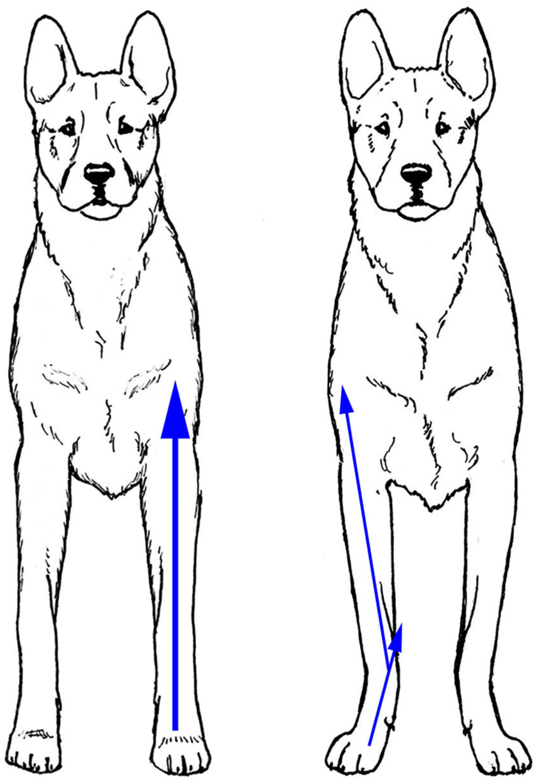


FIGURE 6 | When viewed from the front, the thoracic limbs should form a straight line perpendicular to the ground **(Left)**. Angled limbs **(Right)** dissipate the power output during movement (arrows). Illustration by M. Schlehr.

FIGURE 7 | When a dog is standing relaxed, it is normal for the thoracic limbs to be externally rotated, resulting in the feet being positioned with the toes pointing away from each other. This provides stability in the standing position.

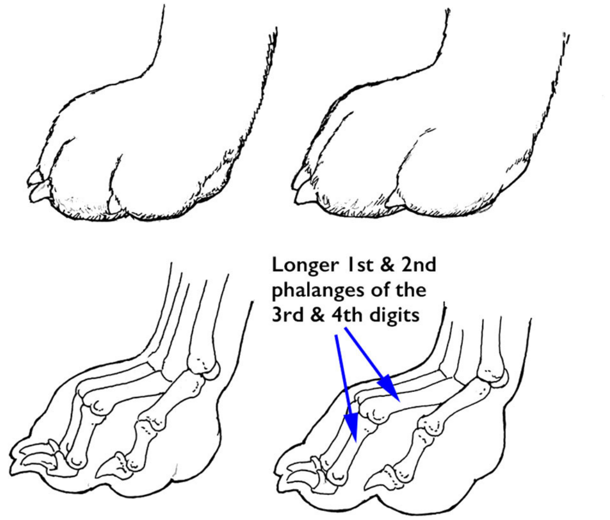


FIGURE 8 | Cat **(Left)** and hare **(Right)** feet, demonstrating the longer 1st and 2nd phalanges of the 3rd and 4th digits in the hare foot (arrows). Illustration by M. Schlehr.

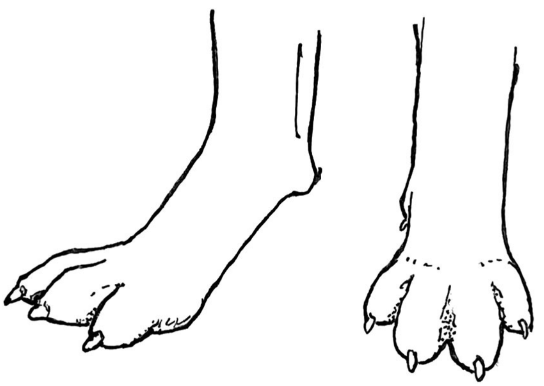


FIGURE 9 | Many German Shepherd Dogs bred for conformation have splayed toes, which might be a reflection of the generalized increased laxity of ligaments and tendons in this breed. Illustration by M. Schlehr.

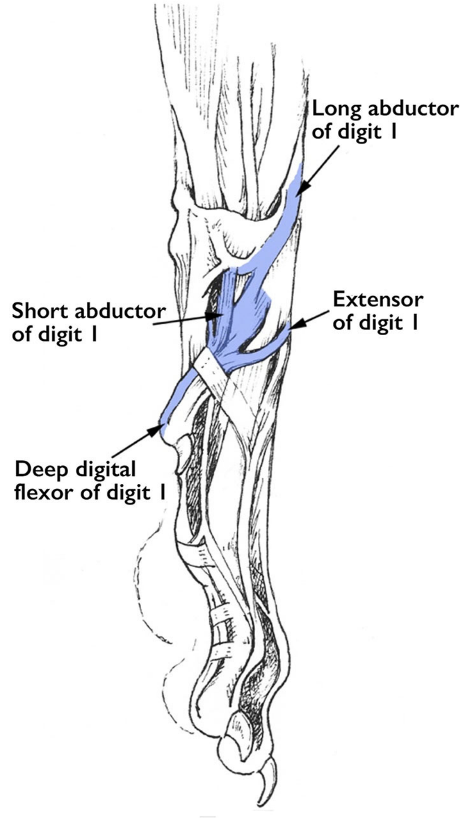


FIGURE 10 | Tendons that attach to the thoracic limb dew claw. Illustration by M. Schlehr (from Miller and Evan's Guide to the Anatomy of the Dog).

All dogs are born with a first digit on the thoracic limb, also known as the dew claw. Many dogs have their front dew claws removed as 3-day-old puppies because their breeders wish to reduce the risk of dew claw injuries. Breeders of conformation dogs also believe that the absence of a dew claw makes the legs appear straighter when viewed from the front. None of the breed standards of the German Shepherd Dog, the Belgian Malinois or the Labrador Retriever require the dew claws to be removed, and in fact less than a handful of the ∼200 breed standards do.

Examination of the muscles and tendons attached to the front dewclaws confirm that these digits are functional (1). Four tendons that connect the dewclaw to muscles of the distal thoracic limb (**Figure 10**) demonstrate that this digit does have the ability to move individually. To the best of our knowledge, all wild carnivores with the exception of African wild dogs have front dewclaws, providing evolutionary proof that they are functional digits.

The front dew claws appear to be non-functional when the dog is in a standing position because they are not in contact with the ground. However, when dogs are cantering, galloping or jumping and thus bearing the majority of their weight on the thoracic limbs, the dew claw does contact the ground (**Figure 11**). It is then available to dig into the ground to help stabilize the thoracic limb and reduce torque to the carpus and proximal limb when the dog is turning. An unexpected function of dew claws is to help dogs climb out on ice when the dog accidentally slips through the ice of a pond (or intentionally goes swimming in freezing water). In their position on the medial aspect of the thoracic limbs, they can act as little ice picks to help the dog grip the ice and lift itself out of the water. As a result, many individuals who train performance and working dogs recommend that dew claws not be amputated.

Dew claws on the pelvic limb are almost always vestigial and lack the tendinous attachments of the thoracic limb dew claws. They generally are removed within a few days of birth, except in those breeds such as the Beauceron, Briard, Great Pyrenees, Icelandic Sheepdog, and some others for which the breed standard specifies the presence of rear dew claws.

FIGURE 11 | A Corgi herding a sheep demonstrating the use of its left thoracic limb dew claw (arrow) in turning.

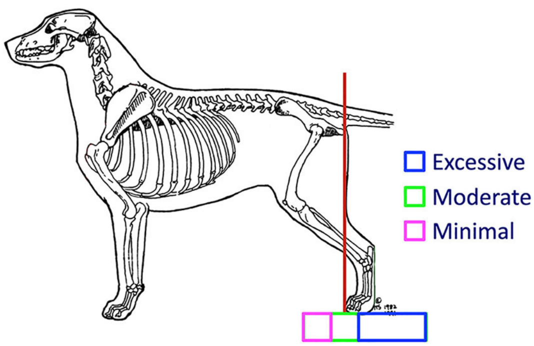


FIGURE 12 | A dog has ideal (moderate) pelvic limb angulation when a line drawn perpendicular to the ground touching the caudal aspect of the ischiatic tuberosity meets the ground at the cranial aspect of the toes (red line). Illustration by M. Schlehr.

PELVIC LIMB STRUCTURE

Pelvic Limb Angulation – Side View

Pelvic limb angulation, the angles at which the pelvis and long bones meet one another when the dog is standing, varies widely between different breeds and also between individuals within those breeds. Those who study and evaluate canine structure often refer to this as *rear angulation* by (19–21). As with other structural evaluations, pelvic limb angulation is best assessed by having the dog stand in the stacked position, with the metatarsals oriented perpendicular to the ground. A rule of thumb used by those who study canine structure to evaluate rear angulation is to draw an imaginary line perpendicular to the ground along the caudal aspect of the ischiatic tuberosity (**Figure 12**). Ideally that line should pass through the cranial aspect of the toes, or within a half of the dog's foot length cranial or caudal to that point.

There are advantages and disadvantages to having either minimal or excessive pelvic limb angulation. Dogs with abundant pelvic limb angulation are able to unfold their limbs to reach farther forward with each stride, powering the body further forward as they extend their pelvic limbs far caudally before lifting the foot for the swing phase of the stride. Excessive pelvic limb angulation, however, is often associated with instability. Since the majority of the pelvic limb musculature is in the proximal part of the limb, there is minimal musculature to stabilize the distal pelvic limb, particularly the tarsus, against lateral or rotational movement. In addition, as with the thoracic limb, stability decreases the further the foot is from a position directly under the dog's trunk.

Williams et al. demonstrated that the greatest increases in power during acceleration of Greyhounds occurred at the coxofemoral and tarsal joints (10). There cannot be power driving movement without stability. The pelvic limb needs to drive acceleration in the sagittal plane. Any lateral movement dissipates this power. Biomechanically, there is an inverse relationship between rear angulation and stability. In the moving dog there is a need for balance between sufficient pelvic limb angulation to provide for power for acceleration and continued movement, but also sufficient stability to apply that power effectively against the ground. This balance is thought to be achieved through moderate pelvic limb angulation as demonstrated in **Figure 12**.

There is strong evidence of functional trade-offs in comparing the limb muscles of dogs that have been selectively bred for running vs. fighting (11). Dogs such as Greyhounds that were bred for running have substantially less musculature in the distal limbs so that there is less weight distally and thus reduced rotational inertia of their oscillating limbs. In addition, they tend to have weaker musculature in the thoracic limbs than the pelvic limbs. The pelvic limbs are thought to have a greater role in acceleration while the thoracic limbs are more important for deceleration (24, 25).

In contrast, dogs bred for fighting, such as Pit Bulls, tend to have well-muscled distal limbs that can produce more power and sustain improved agility as well as balance and opponent manipulation (11). They also have more equal musculature in their thoracic and pelvic limbs. In these breeds thoracic limb strength is believed to be essential for rapid turning and agility. It is interesting to ponder which of these structural differences are ideal for Working Dogs, which have functions that require both acceleration and agility. As with so many other structural features, a balance between the two extremes is likely ideal.

Some breeds have been selectively bred to have extreme pelvic limb angulation. One of these is the German Shepherd Dog, particularly those bred for conformation dog shows, which has shown marked increases in rear limb angulation from moderate to extremely angulated over the last several decades (**Figure 13A**). Many individuals of this breed have such extreme angulation that they are unable to stand in the typical stacked position but instead must stand with one pelvic limb's metatarsals perpendicular. to the ground, and with the other pelvic limb's foot placed under the body to improve stability. The result of this extreme pelvic limb angulation is that the pelvis is positioned closer to the ground, and the dog's spine is extremely sloped from cranial to caudal. This extreme pelvic limb angulation often cannot be compensated for by muscular strength, and these dogs' tarsi swing medially each time the feet are planted, thus reducing the power transmitted to the body. Often these dogs experience such instability on the standing leg that they are unable to lift the contralateral foot fully on the swing phase of the stride (**Figure 13B**).

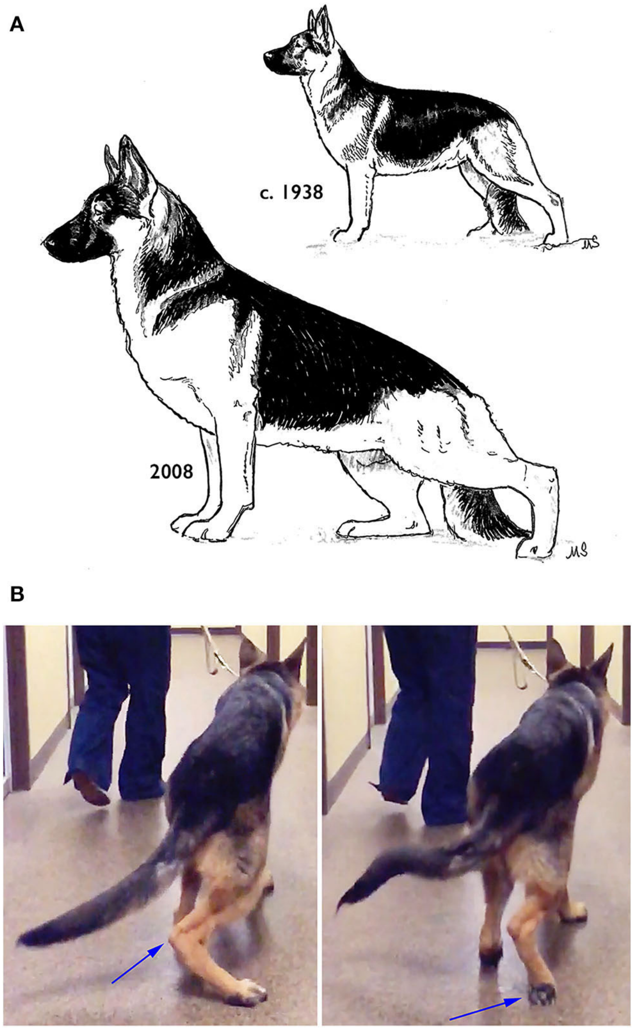


FIGURE 13 | (A) Changes in the pelvic limb structure of the German Shepherd Dog over the last several decades. **(B)** A healthy, year-old German Shepherd Dog with extreme pelvic limb angulation showing medial displacement of the right tarsus at the end of the stance phase (left), and a knuckling of the right pelvic limb due to the inability of the left pelvic limb to support the contralateral limb during the swing phase. Illustration by M. Schlehr.

To the best of our knowledge, these structural changes in the German Shepherd Dog provide no functional advantages. Any potential advantage in function appears to be offset by instability. As observed by Fischer and Lilje, "whenever selection starts, whether it is the skull or locomotion, it will affect other parts of the body" (15). The German Shepherd Dog tends to have laxity in many joints throughout the body, not just in the pelvic limb. These dogs also frequently have an increased angle of carpal extension when standing, splayed toes, etc. It is possible that this reflects an unintended selection in these dogs toward increased extensibility of all tendons and ligaments while selecting for extreme pelvic limb angulation. It is therefore perhaps not surprising that German Shepherd Dogs have a very high prevalence of hip dysplasia as compared to other large breeds with more moderate pelvic limb angulation such as Golden Retrievers, Labrador Retrievers, and Rottweilers (26). This might be one reason why many organizations are moving away from using German Shepherd Dogs as Working Dogs, or cross-breeding them with Belgian Malinois.

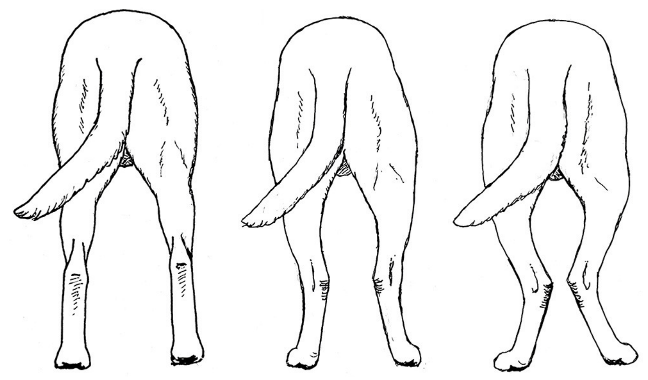


FIGURE 14 | Correct pelvic limb conformation, when viewed from the rear, in the Labrador Retriever **(Left)**. The pelvic limbs of herding breeds often exhibit mild external rotation **(Middle)**, but excessive external rotation **(Right)** should be avoided. Illustration by M. Schlehr.

At the other extreme of pelvic limb angulation are breeds with very straight pelvic limb angulation. Although minimal pelvic limb angulation is more typical of breeds originally developed for guarding, some individuals of the usual Working Dog breeds can also have relatively limited pelvic limb angulation. Biomechanically, minimal pelvic limb angulation tends to increase the potential for torque along the axis of the limb and may result in increased stress on the ligaments of the stifle and tarsus. Both extremes of pelvic limb angulation should be avoided when selecting Working Dogs.

Pelvic Limb - Rear View

In many breeds, when viewed from the rear, the pelvic limbs should extend distally from the greater trochanter parallel to each other and perpendicular to the ground (**Figure 14**, left). Breeds such as herding dogs, whose functions require the dog to make quick turns, frequently stand with the pelvic limbs externally rotated, such that the tarsi are positioned medially relative to the stifles and feet (**Figure 14**, middle). This pelvic limb structure provides greater stability when the dog is required to frequently crouch, lie down and stand up. Further, it allows the toes to push off with more power when turning. This pelvic limb conformation is almost universal in German Shepherd Dogs and is very common in Belgian Malinois, both of which are herding breeds. It is less common in Labrador Retrievers, which were bred to run in straight lines to retrieve game. If this external rotation of the pelvic limbs is extreme (**Figure 14**, right), however, it can interfere with forward movement and should be avoided when selecting Working Dogs.

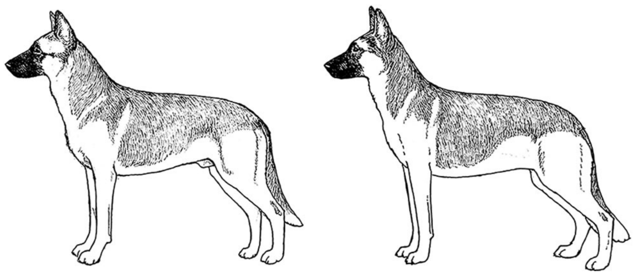


FIGURE 15 | Balanced angulation in the Belgian Malinois (left). The dog on the right is imbalanced, with less thoracic limb angulation than pelvic limb angulation. Illustration by M. Schlehr.

BALANCED THORACIC AND PELVIC LIMB ANGULATION

The thoracic and pelvic limbs in a given dog should have approximately equal, or balanced, angulation. This is important for coordination of movement, particularly at the trot, when diagonally opposite thoracic and pelvic limbs strike the ground at the same time. If the thoracic limbs, for example, have less angulation than the pelvic limbs, they will have a shorter stride length and therefore a shorter cycle time than the pelvic limbs, making it difficult for diagonally opposite limbs to strike the ground at the same time. In addition, the less angulated limbs are generally less muscular than the more angulated ones. The most common form of lack of balance is when dogs have less thoracic than pelvic limb angulation (**Figure 15**).

For a dog to achieve the optimal thoracic or pelvic limb angulation that is consistent with its genetics, the dog's musculature must be fully developed. Further, strong muscles are required for the limbs to provide optimal power for movement. Dogs with straighter thoracic or pelvic limb angulation tend to have weaker limb musculature. This might be in part because in the standing dog, supporting the weight with bones at a more acute angle requires active muscle contraction. In a dog with less angulation a greater percentage of the dog's weight can be supported by the bones. All Working Dogs should be engaged in routine fitness programs to optimize their musculature and thus their angulation and function.

THE HEAD

Skull morphology is a major factor in bite force (27). Working Dogs should have large heads to provide sufficiently powerful bite muscles (predominantly the masseter and the temporalis muscles), strong jaw bones, and well-muscled necks. They also should have full dentition; a good scissors bite provides the strongest grip. Mesocephalic skulls provide the best combination of a moderate length muzzle and good teeth (28). Working Dogs also should have large, open nostrils to facilitate the passage of air when scenting.

THE TAIL

The tail provides an important counterbalance for dogs when they need to turn quickly, either on land or when swimming. The tail also helps elevate the dog's rear assembly after the apex of trajectory of a jump, helping the dog land on its front feet. A Working Dog's tail should be strong and of enough length to provide sufficient counterbalance, especially for jobs that require jumping or sharp turns.

THE COAT

Working dogs need a weather-resistant coat that dries easily when wet, sheds dirt, and is easy to care for. Most Working Dog breeds have a double coat, characterized by large guard hairs that stand up from the skin supported by the undercoat, which consists of more numerous, finer hairs. Most organizations prefer to have dogs of a color that blends with the environment, so white dogs or extensive white markings are not advisable.

CONCLUSION

There are many components of structure that can affect the ability of a Working Dog to achieve its optimum abilities and to have a long, injury-free career. These components are important to consider when selecting an adult dog for a career as a Working Dog. Breeders of future Working Dogs should give strong consideration to selecting for the characteristics that will allow these dogs to excel in their careers and live long and productive lives.

AUTHOR CONTRIBUTIONS

CZ conceived of and wrote the manuscript. MS drew all illustrations to demonstrate the specific components described in the manuscript and contributed numerous concepts and ideas to the structure and content of the manuscript. Both authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We thank Ainslie Mills for hours of fruitful discussion about canine structure based on her years as an all-breed dog judge.

REFERENCES

1. Hermanson JW, de LaHunta A, Evans HE. *Miller and Evan's Guide to the Anatomy of the Dog*. Philadelphia: WB Saunders (2019). p. 1004.
2. Sutter NB, Mosher DS, Gray MM, Ostrander EA. Morphometrics within dog breeds are highly reproducible and dispute Rensch's rule. *Mamm Genome.* (2008) 19:713–23. doi: 10.1007/s00335-008-9153-6

3. Wayne RK. Cranial morphology of domestic and wild canids: the influence of development on morphological change. *Evolution*. (1986a) 40:243–61. doi: 10.1111/j.1558-5646.1986.tb00467.x
4. Wayne RK. Limb morphology of domestic and wild canids: the influence of development on morphological change. *J Morphol*. (1986b) 187:301–19. doi: 10.1002/jmor.1051870304
5. Wang YL, Yang T, Zeng C, Wei J, Xie D-X, Yang Y-H, et al. Association between tibial plateau slopes and anterior cruciate ligament injury: a meta-analysis. *Arthroscopy*. (2017) 33:1248–59.e4. doi: 10.1016/j.arthro.2017.01.015
6. Prakash J, Seon JK, Ahn HW, Cho KJ, Im CJ, Song EK. Factors affecting tibial tuberosity-trochlear groove distance in recurrent patellar dislocation. *Clin Orthop Surg*. (2018) 10:420–6. doi: 10.4055/cios.2018.10.4.420
7. Meija S, Steward N, Miller A, Savidky R, Monarski C, Moor GE, et al. Accuracy of external measurements of 3-dimensional (3D) printed biomodels of the canine radius used in an in-hospital setting. *Can J Vet Res*. (2019) 83:181–6.
8. Fischer MS, Lehmann SV, Andrada E. Three-dimensional kinematics of canine hind limbs: in vivo, biplanar, high-frequency fluoroscopic analysis of four breeds during walking and trotting. *Sci Rep*. (2018) 8:16982. doi: 10.1038/s41598-018-34310-0
9. Williams SB, Wilson AM, Rhodes L, Andrews J, Payne RC. Functional anatomy and muscle moment arms of the thoracic limb of an elite sprinting athlete: the racing greyhound (*Canis familiaris*). *J Anat*. (2008) 213:373–82. doi: 10.1111/j.1469-7580.2008.00962.x
10. Williams SB, Usherwood JR, Jespers K, Channon AJ, Wilson AM. Exploring the mechanical basis for acceleration: pelvic limb locomotor function during accelerations in racing greyhounds (*Canis familiaris*). *J Exp Biol*. (2009) 212:550–65. doi: 10.1242/jeb.018093
11. Pasi BM, Carrier DR. Functional trade-offs in the limb muscles of dogs selected for running vs. fighting. *J Evol Biol*. (2003) 16:324–32. doi: 10.1046/j.1420-9101.2003.00512.x
12. Webster EL, Hudson PE, Channon SB. Comparative functional anatomy of the epaxial musculature of dogs (Canis familiaris) bred for springing vs. fighting. *J Anat*. (2014) 225:317–27. doi: 10.1111/joa.12208
13. Schilling N. Evolution of the axial system in craniates: morphology and function of the perivertebral musculature. *Front Zool*. (2011) 8:4 doi: 10.1186/1742-9994-8-4
14. Sharir A, Milgram J, Shahar R. Structural and functional anatomy of the neck musculature of the dog (*Canis familiaris*). *J Anat*. (2006) 208:331–51. doi: 10.1111/j.1469-7580.2006.00533.x
15. Fischer MS, Lilje KE. *Dogs in Motion*. Dortmund: VDH Service GmbH. (2011). p. 207.
16. Hastings P. *Puppy Puzzle – Evaluating Structural Quality (DVD)*. Beaverton, OR: DogFolk Enterprises (2006).
17. Coppinger R, Coppinger L. *Dogs. A New Understanding of Canine Origin, Behavior, And Evolution*. Chicago, IL: University Of Chicago Press (2001). p. 352.
18. Schlehr MR. *Judging the Golden Retriever. A Discussion of the Breed Standard*. Golden Retriever Club of America. (2019).
19. Brown CM. *Dog Locomotion and Gait Analysis*. Wheat Ridge, CO: Hoflin Publishing (1986). p. 160.
20. Elliott RP. *Dogsteps: A New Look*. 3rd Edn. Irvine, CA: Fancy Publications (2009). p. 133.
21. Cole RW. *An Eye for a Dog. Illustrated Guide to Judging Purebred Dogs*. Wenatchee, WA: Dogwise Publishing (2004). p. 180.
22. Hastings P. S*tructure in Action: The Makings of a Durable Dog*. Beaverton, OR: DogFolk Enterprises. (2011). p. 168.
23. Canapp SO, Canapp DA, Ibrahim V, Carr BJ, Cox C, Barrett JG. The use of adipose-derived progenitor cells and platelet-rich plasma combination for the treatment of supraspinatus tendinopathy in 55 dogs: a retrospective study. *Front Vet Sci*. (2016) 3:61. doi: 10.3389/fvets.2016.00061
24. Cavagna GA, Heghund NC, Taylor CR. Mechanical work in terrestrial locomotion: two basic mechanisms for minimizing energy expenditure. *Am J Physiol*. (1977) 233:R243–61. doi: 10.1152/ajpregu.1977.233.5.R243
25. Jayes AS, Alexander RMcN. Mechanics of locomotion of dogs (*Canis familiaris*) and sheep (*Ovis aries*). *J Zool Lond*. (1978) 185:289–308. doi: 10.1111/j.1469-7998.1978.tb03334.x
26. Smith GK, Mayhew PD, Kapatkin AS, McKelvie PJ, Shofer FS, Gregor TP. Evaluation of risk factors for degenerative joint disease associated with hip dysplasia in German Shepherd Dogs, Golden Retrievers, Labrador Retrievers, and Rottweilers. *J Am Vet Med Assoc*. (2001) 219:1719–24. doi: 10.2460/javma.2001.219.1719
27. Kim SE, Arzi B, Garcia TC, Verstraete JM. Bite forces and their measurement in dogs and cats. *Front Vet Sci*. (2018) 5:76. doi: 10.3389/fvets.2018.00076
28. Ellis JL, Thomason JJ, Kebreab E, France J. Calibration of estimated biting forces in domestic canids: comparison of post-mortem and in vivo measurements. *J Anat*. (2008) 212:769–80. doi: 10.1111/j.1469-7580.2008.00911.x

Conflict of Interest: CZ was employed by Zink Integrative Sports Medicine.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 30 October 2020
doi: 10.3389/fvets.2020.590834

Testing the Ability of Dogs to Detect Different Odor Concentrations of the Carolina Anole (*Anolis carolinensis*) in Japan

Megumi Fukuzawa* and Koki Shibata

Department of Animal Science and Resources, College of Bioresource Sciences, Nihon University, Fujisawa, Japan

OPEN ACCESS

Edited by:
Cynthia M. Otto,
University of Pennsylvania,
United States

Reviewed by:
Bart Rogers,
Auburn University, United States
Peter Coppolillo,
Working Dogs for Conservation,
United States

****Correspondence:***
Megumi Fukuzawa
fukuzawa.megumi@nihon-u.ac.jp

Specialty section:
This article was submitted to
Veterinary Humanities and Social
Sciences,
a section of the journal
Frontiers in Veterinary Science

Received: *03 August 2020*
Accepted: *30 September 2020*
Published: *30 October 2020*

Citation:
Fukuzawa M and Shibata K (2020)
Testing the Ability of Dogs to Detect
Different Odor Concentrations of the
Carolina Anole (Anolis carolinensis) in
Japan. Front. Vet. Sci. 7:590834.
doi: 10.3389/fvets.2020.590834

The Carolina anole (*Anolis carolinensis*) is regarded as a problem in the Ogasawara Islands. The decision to use eradication measures depends on the limit of detection at low densities. We tested the ability of two dogs to discriminate the odor of anole to assess the possibility of using dogs to detect anoles at low densities. The two dogs were trained to discriminate the basic target odor concentration (512 anoles/ha) on 10-g coconut peat sachets. When they reached 100% accuracy, they were tested at different odor concentrations (densities of 385, 256, 128, 26, and 3 anoles/ha). During training, both dogs achieved 100% accuracy after 2 daily sessions in only 2 days. They were able to select the positive odor concentration sachet, and their accuracy was from 75 to 100%. We believe that testing using soil from sites of high anole high density and at the limit of detection in the Ogasawara Islands will be useful.

Keywords: alien species, detection, dog, judging eradication, training

INTRODUCTION

The Ogasawara Islands consist of more than 30 archipelagos in 5 groups, with a total land area of 7,939 hectares. These islands' ecosystems clearly reflect the evolutionary processes of many endemic species, and the islands were registered as a World Heritage Site in 2011 (1). On the otherhand, the native insect community in Ogasawara has declined greatly in all areas owing to predation by the anole. The Japanese Ministry of the Environment, therefore, listed the Carolina anole as an invasive alien species in Japan in June 2005, barring its spread to new islands and setting natural regeneration zones in remaining habitats of native insects. The anole is controlled by setting polypropylene adhesive traps around tree trunks to capture them, and by the erection of barrier fences and electric fences. Adhesive traps are effective except where numbers are low, and therefore do not allow us to judge whether a population has been eliminated or not. Since the Carolina anole (*Anolis carolinensis*) was found on Anijima island in 2013, intensive measures to eradicate them have been in progress (2). Although the anole population density on Anijima is at most only 1/10 of that on Chichijima, it is difficult to judge the success of eradication when the density is so low.

The keen olfactory ability of dogs is used in the conservation of various species [e.g., (3–5)], and is considered effective in judging eradication in areas of low density. For example, Kretser et al. (6) concluded that scat-detection dogs were effective at locating moose (*Alces alces*) scat in areas of low population density in New York state, USA. Statham et al. (7) also reported that dog-handler teams are a promising survey tool to detect the presence of blunt-nosed leopard lizard

(*Gambelia sila*) and to help increase sample sizes of scats necessary for subsequent DNA analysis for research and conservation purposes. Since the Ogasawara Islands are a World Heritage Sites, access to uninhabited islands is restricted to both people and detection dogs. Not only are detection dogs required to survey huge areas daily, but they are needed with certain considerations (e.g., cultivate cost, training period). Even if the dog has high olfactory detection ability, we cannot use as a detection dog unless it has a physical body that can withstand fieldwork. Ironically, gaining experience will strengthen (stabilize) the dog's detection ability, however, the environment to be detected is often severe for dogs, and their physical level also decreases with age. Therefore, it may be more practical to carry out discrimination tests on soil samples rather than *in situ*. If we are able to confirm the detection ability of dogs in indoor, then it may reduce various risks to dogs when these detecting performed under adverse conditions.

We tested the ability of two dogs to discriminate the odor of anoles to assess the possibility of using dogs to detect anoles at low densities.

MATERIALS AND METHODS

Dogs

We trained two healthy adult female dogs (German Shepherd, 43 months, 25.8 kg; Labrador Retriever, 52 months, 23.6 kg). They had previously been trained to detect odors of Carolina anoles' bodies and excrement/urine (unpublished data), first learning to recognize body odor and then excrement/urine odor. The dogs were kept in a familiar kennel at the university during the day, where they were allowed contact with the outdoors and with humans at all times. Each dog was trained by a different handler. The German Shepherd's handler was a woman who had been a handler for nearly 2 years with GS, and the Labrador's handler was a familiar male person with both dogs, but this was his first time to handle LR.

Target Odor

Twelve male anoles captured with mealworm bait on Chichijima in July 2018 were housed in the Specified Alien Livestock Allowance Area at Nihon University in individual plastic breeding cages (W 300 mm × D 195 mm × H 205 mm) in accordance with the protocol for *Anolis* lizards (8). The lizards were unable to see each other. Each cage held a perch. The floor of each cage was covered with 37 g (5 mm deep) of home-garden coconut peat (15–30% moisture; ≤0.5% total N; ≤0.1% total phosphoric acid; ≤1.0% total K, pH 5.5–6.5; 70 to 90% organic matter; cation exchange capacity = 80–110 meq/100 g; maximum water capacity = 800–1000%) (**Figure 1**). After 72 h, the peat was recovered, sealed in a press-seal plastic bag, and stored in a freezer at −27°C. Other cages containing only peat were left in the same environment, and the peat was used to adjust concentrations and also to create the controls. The control peat also frozen. The anole population density (/ha) was calculated from the case floor area (390 cm^2) at 256 410/ha. As the population density in the Ogasawara Islands was 1270/ha (9), we set 0.2% (512 anoles/ha) as the target odor concentration for the basic training.

FIGURE 1 | An anole in its cage.

To prepare a test odor sachet (10 g peat), the day before the test we homogenized the 37 g from one cage and weighed out 0.02 g on an electronic balance, in addition to 9.98 g of control peat, and placed the total 10.00 g in a nonwoven cloth sachet (Marusan Industry Co., Ltd., Ehime, Japan; **Figure 2**). We also made up a 10 g sachet from the control peat (i.e., control sachet). These sachets were sealed in press-seal plastic bags stored at −27°C until the test. Just before the test, each sachet was sealed in a perforated plastic container (V-5, Sanoya Industry Co., Ltd. Aichi, Japan; **Figure 3**). The dogs were able to touch the container but not the sachet. For the test, we used 5 odor levels, setting anole densities of 385/ha (0.15%: 0.015 + 9.985 g), 256/ha (0.10%: 0.01 + 9.99 g), 128/ha (0.05%: 0.005 + 9.995 g), 26/ha (0.01%: 0.001 + 9.999 g), and 3/ha (0.001%: 0.0001 + 9.9999 g), which were prepared by the same procedure.

Procedures

All training trials and tests were carried out in the same room (200 × 345 cm). The room held a low stainless steel table (88 cm W × 16 cm D × 15 cm H) on which the odor sachets were set. To record the dog's search behavior, we set a video camera (DCR-SR87, Sony, Tokyo, Japan) before the table and another behind it. The room temperature was controlled at around 20°C (mean, 19.63 ± 0.83°C; RH, 60.31 ± 8.64%). For both the training trials and the discrimination tests, each dog was brought from her kennel to the room.

Training used the 512/ha sachet (positive stimulus) and a control sachet. Two plastic containers containing each sachet were placed in advance with a gloved hand at 50 cm intervals on the steel table. The positive and control sachet was presented on the left and right sides equally, randomly. At the beginning of each trial, the gloved handler presented the dog with a plastic container containing the positive odor sachet to smell for up to 60 s. When the handler judged that the dog recognized the odor, she told the dog to "search" and released her. When the dog recognized the odor in the container on the table, she lay (or sat) down in front of it. The dog was given 60 s to choose. When she selected the correct container, she received a food reward. When she selected the incorrect container, the

FIGURE 2 | A prepared odor sachet.

FIGURE 3 | A plastic container holding an odor sachet.

trainer ignored her for 3 min, then started next trial. Each session consisted of 20 such trials. Twenty different samples were used and changed between each session. The dog was allowed short rests both between and during sessions. One or two sessions were performed each day for 4 or 5 days per week. The number of session per a day depend on the dog's motivation. The trainer collected data from the videotapes, including the time the dog took to select the correct scent (to a precision of 0.01 s). When a dog reached a 100% success rate in one session, the two-way alternative discrimination test (e.g., 512/ha vs. control) was started. Each trial was blind. A person other than the handler set the arrangement of the container in which the odor sachet was set in advance, and the handler placed it on the table during the test. That is, the handler did not know the odor level in the container. This test starts with a population density of 512 anoles/ha. If the result of one session is >80% correct answer rate (i.e., 16/20 trials), the odor level is reduced; if it is <80%, the level is increased. That is, as the test progressed, the odor level of positive stimulus was decreased. The procedure was the same as in basic training except for odor samples. There were no situations in which no odor was presented (a blank test).

Statistical Analysis

We compared the numbers of sessions required to reach the basic training criterion between dogs and the average search times between correct and incorrect choices in each session by Wilcoxon's signed-rank test. We examined the effect of session on the average search time of a dog's correct and incorrect choices in training by the same test. In the two-way alternative discriminative tests, we compared differences in average search times of correct or incorrect choices both between odor levels and within each odor level by Kruskal–Wallis test. When significant effects were found, we used a *post hoc* Steel–Dwass tests for pairwise comparison of the means of the search times.

RESULTS

Training

The correct answer rate for the first session was different for the two dogs; with 85% (17/20) for Labrador Retriever and 95% (19/20) for German Shepherd dog. However, the total number of sessions (of 20 trials each) to reach the criterion was only 2 for both dogs.

There was no significant difference between the mean search time for two dogs when they selected the positive stimulus (sachet with odor; 15.47 ± 16.12 s), and the incorrect stimulus (control sachet; 27.47 ± 18.88 s) ($U = 89$, $P = 0.168$). On the other hand, the mean search time when a dog selected the positive stimulus (sachet with odor: LR 25.54 ± 17.80 s vs. GS 5.91 ±

TABLE 1 | Numbers and rates of correct choice for each odor level.

Odor level (anoles/ha)	Mean correct % of trial	LR		GS	
		Correct number of trial	%	Correct number of trial	%
512	100	20/20	100	20/20	100
385	95	18/20	90	20/20	100
256	92.5	17/20	85	20/20	100
128	90	16/20	80	20/20	100
26	85	15/20	75	19/20	95
3	75	10/20	50	20/20	100

TABLE 2 | Average search time ± SD at each odor level.

Odor level (anoles/ha)	Correct search time	Incorrect search time	LR Correct search time	LR Incorrect search time	GS Correct search time	GS Incorrect search time
512	9.37 ± 5.44	–	12.28 ± 5.45[b]	–	6.45 ± 3.30[ab]	-
385	8.07 ± 4.95	16.65 ± 7.83	10.72 ± 5.24[b]	16.65 ± 7.83	5.68 ± 2.93[b]	-
256	9.93 ± 7.28	17.31 ± 8.56	14.65 ± 7.80[ab]	17.31 ± 8.56	5.68 ± 2.93[b]	-
128	13.84 ± 13.89	24.13 ± 13.01	23.87 ± 15.26[a]	24.13 ± 13.01	5.81 ± 2.57[b]	-
26	8.28 ± 5.70[x]	12.38 ± 8.03[y]	11.23 ± 6.30[b]	13.45 ± 7.59	5.94 ± 3.58[b]	7.02
3	14.16 ± 12.47	17.92 ± 12.14	23.44 ± 17.06[ab]	17.92 ± 11.52	9.51 ± 3.91[a]	-
Ave.	10.48 ± 9.06	17.41 ± 11.19	14.40 ± 9.86	17.79 ± 10.57	5.96 ± 3.09	7.02

x-y: Mann-Whitney, $p < 0.05$.
a-b: Steel-Dwass, $p < 0.05$ within dog.

3.47 s; $w = 6.68$, $P \leq 0.01$) or the incorrect stimulus (control sachet: 35.1 ± 11.13 s vs. 4.59 s; $w = 4.50$, $P < 0.05$) differed significantly between dogs. The LR's mean correct search time in training session 2 (17.82 ± 12.86 s) was significantly shorter than that in session 1 (34.60 ± 18.50 s; $w = 2.74$, $P < 0.05$).

Different Odor Concentrations Test

The correct rate when they selected the positive odor concentration sachet in the different odor concentration tests was 75 to 100% (Median 91.25%), and the rate of correct identification of each odor level differed between the dogs (**Table 1**). Dogs ranged from 100% at the highest concentration to 50 and 100 % at the lowest.

The search time range when the dogs made the correct choice at any odor level was 4.72 to 60.0 s for LR and 3.19 to 20.19 s for GS. There was no effect of each odor level for two dogs' mean correct ($\chi^2 = 10.48$, $P= 0.06$)/incorrect search times ($\chi^2 = 3.44$, $P = 0.487$). When the correct/incorrect search time of each odor level was compared, the incorrect search time at 26 anoles/ha was significantly longer than the correct search time ($U = 48$, $P < 0.05$) (**Table 2**). Each dog's correct search time differed significantly among odor levels (LR $\chi^2 = 17.18$, $P < 0.0001$; GS $\chi^2 = 18.54$, $P < 0.0001$), but not her incorrect search time (LR $\chi^2 = 2.47$, $P = 0.65$). There was no significant difference between the correct and incorrect search times at the same odor level by either dog.

DISCUSSION

Both dogs reached 100% correct discrimination in two training sessions. The average correct selection rate of two dogs for the first session was 90%. It was thought that their previous experience was advantageous. In previous training with anole excrement/ urine odor, they achieved 100% correct recognition of body odor in 27 sessions (1 session = 20 trials), and of excrement/urine odor in 2 (LR) and 4 sessions (GS). The mean correct search time for excrement/urine odor was 31.23 ± 1.38 s (LR) and 8.52 ± 0.55 s (GS). In that training, the dogs were presented with cloth that had been stored with excrement/urine mass, unlike the conditions in the new training. The accuracy of odor detection is affected by the quality and intensity of the target odor; as the anole may secrete pheromones, probably from the cloacal glands, onto the surface of feces or scats (10), discrimination between exposed peat and control peat was a relatively easy task for the two dogs.

The accuracy when they selected the positive odor concentration sachet in the different odor concentration tests was 75 to 100% (Median 91.25%). Their accuracy of 512 to 128 anoles/ha showed 90% or more, which suggested the usefulness of olfactory ability in dogs. It was reported that dogs are able to sample a variety of deer species scats 0.21 scats/km (11), to sample moose scats ~ 1.4 samples/km (6). Our test did not sample for scats, but the focus was on being able to respond to the excrement/urine odor contained in the soil. In any case, it is clear that dogs can detect a slight odor from a large area. Cristescu et al. (12) examined the use of dogs in detecting koalas (*Phascolarctos cinereus*) at low densities. They suggested that the detection ability was influenced by whether dogs were on or off the leash, but not the age of the feces (fresh or old scat). In 150 field trials, the average time to find koala scats per 100 m^2 was 5.2 min and the accuracy was 97%. In contrast, in our controlled environment the search time was <60 s. It seems that such training can both reduce the physical burden on dogs and greatly shorten the detection effort and time. Furthermore, if it is difficult for dogs to search in the field, indoor tests are still effective. Because if done indoors, it is possible to control not only the temperature, humidity and wind direction but also environmental noise and temptation odor etc. Thus, the dog is able to focus on the odor that has to be detected.

However, differences between dogs should also be taken into consideration. Svartberg (13) found a general relationship between a bold personality and an ability to learn and perform well in tasks requiring varied training. Of course, our two dogs and handlers must have experience in order to keep high motivation and accuracy of detection. While dogs have many potentials, it is also necessary to consider how we humans can use their abilities efficiently. If we are able to confirm the detection ability of dogs indoors, then we will need to test their abilities in the field as a next step. For example, it could be a test of the effect of variable weather-ground conditions on dog abilities, an alternative discrimination test of male and female anole excretion odor, or no correct choice test. We believe that testing using soil from sites of high anole high density and at the limit of detection in the Ogasawara Islands will be useful.

CONCLUSION

Dogs were able to recognize different population densities created from coconut peat used to house anoles, and they can keep high motivation during detection task, even if the target odor was low density.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by The Nihon University Ethics committee (EXC17B002). Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR'S NOTE

The use of dogs for detecting invasive animal species has increased in Japan. Damage to endemic species by the Carolina anole (*Anolis carolinensis*) is expanding in the Ogasawara Islands. We hypothesized that dogs trained to recognize the anole's odor would be useful in detection work. The dog was trained in a three-way test to discriminate among native Ogasawara reptiles and the anole. The dog achieved 100% accuracy in training for body odor in 27 sessions over 14 days and for excrement/urine odor in 4 sessions over 2 days. The mean correct search time differed significantly between odors in training ($P < 0.05$), but not in the discrimination test ($P = 0.71$). The rates of the correct response in the two-way discrimination test were high (body odor, 90%; excrement/urine odor, 96%). Correction rate in the three-way test was >90%. Discrimination between transferred odor and the odorless sample in training was relatively easy. However, discrimination between odors of the same anole and between odors of the native reptiles and the Carolina anole was harder.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

1. UNESCO. *World Heritage list; Ogasawara Islands*. (2011). Available online at: https://whc.unesco.org/en/list/1362 [Accessed Jun 15, 2019].
2. OINIC. *Road Map for the Carolina Anole Control Measures for Ecosystem Conservation in the Ogasawara Islands.* (2015). Available online at: http://ogasawara-info.jp/pdf/isan/guideline_invasive_4.pdf (Accessed June 15, 2019).
3. Duggan JM, Heske EJ, Schooley RL, Hurt A, Whitelaw A. A comparing detection dog and livetrapping surveys for a cryptic rodent. *J Wildl Manage*. (2011) 75:1209–17. doi: 10.1002/jwmg.150
4. Paula J, Leal MC, Silva MJ, Mascarenhas R, Costa H, Mascarenhas M. Dogs as a tool to improve bird-strike mortality estimates at wind farms. *J Nat Conserv*. (2011) 19:202–8. doi: 10.1016/j.jnc.2011.01.002
5. Browne CM, Stafford KJ, Fordham RA. The detection and identification of tuatara and gecko scents by dogs. *J Vet Behav*. (2015) 10:496–503. doi: 10.1016/j.jveb.2015.08.002
6. Kretser H, Glennon M, Whitelaw A, Hurt A, Pilgrim K, Schwartz M. Scat-detection dogs survey low density moose in New York. *ALCES*. (2016) 52:55–66.
7. Statham MJ, Woollett DA, Fresquez S, Pfeiffer J, Richmond J, Whitelaw A, et al. Noninvasive identification of herpetofauna: pairing conservation dogs and genetic analysis. *J Wildlife Manage*. (2019) 84:66–74. doi: 10.1002/jwmg.21772
8. Sanger TJ, Hime PM, Diani J, Losos JB. Laboratory protocols for husbandry and embryo collection of Anolis lizards. *Herpetol Rev*. (2008) 39:58–63.
9. Okochi I, Yoshimura M, Abe T, Suzuki T. High population densities of an exotic lizard, Anolis carolinensis and its possible role as a pollinator in the Ogasawara Islands. *Bull Forest Forest*. (2006) 5:265–9.
10. Martin J, López P. Chapter 3. Pheromones and chemical communication in lizards. In: Rheubert JL, Siegel DS, Trauth SE, editors. *The Reproductive Biology and Phylogeny of Lizards and Tuatara*. Florida: CRC Press. (2014). p. 43–77.
11. De Oliveira ML, Norris D, Ramires JFM, De perez PH, Galett M, et al. Dogs can detect scat samples more efficiently than humans: an experiment in a continuous Atlantic Forest remmant. *Zoologica*. (2012) 29:183–6. doi: 10.1590/S1984-46702012000200012
12. Cristescu RH, Foley E, Markula A, Jackson G, Jones D, Frère C. Accuracy and efficiency of detection dogs: a powerful new tool for koala conservation and management. *Sci Rep*. (2015) 5:8349. doi: 10.1038/srep08349
13. Svartberg K. Shyness-boldness predicts performance in working dogs. *Appl Anim Behav Sci*. (2002) 79:157–74. doi: 10.1016/S0168-1591(02)00120-X

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 02 February 2021
doi: 10.3389/fvets.2021.545398

Does Benchmarking of Rating Scales Improve Ratings of Search Performance Given by Specialist Search Dog Handlers?

Corinna C. A. Clark[1] *and Nicola J. Rooney*[2*]

[1] Department of Life Sciences, University of Warwick, Coventry, United Kingdom, [2] Animal Welfare and Behavior Group, Bristol Veterinary School, University of Bristol, Bristol, United Kingdom

Rating scales are widely used to rate working dog behavior and performance. Whilst behaviour scales have been extensively validated, instruments used to rate ability have usually been designed by training and practitioner organizations, and often little consideration has been given to how seemingly insignificant aspects of the scale design might alter the validity of the results obtained. Here we illustrate how manipulating one aspect of rating scale design, the provision of verbal benchmarks or labels (as opposed to just a numerical scale), can affect the ability of observers to distinguish between differing levels of search dog performance in an operational environment. Previous studies have found evidence for range restriction (using only part of the scale) in raters' use of the scales and variability between raters in their understanding of the traits used to measures performance. As provision of verbal benchmarks has been shown to help raters in a variety of disciplines to select appropriate scale categories (or scores), it may be predicted that inclusion of verbal benchmarks will bring raters' conceptualization of the traits closer together, increasing agreement between raters, as well as improving the ability of observers to distinguish between differing levels of search dog performance and reduce range restriction. To test the value of verbal benchmarking we compared inter-rater reliability, raters' ability to discriminate between different levels of search dog performance, and their use of the whole scale before and after being presented with benchmarked scales for the same traits. Raters scored the performance of two separate types of explosives search dog (High Assurance Search (HAS) and Vehicle Search (VS) dogs), from short (~30 s) video clips, using 11 previously validated traits. Taking each trait in turn, for the first five clips raters were asked to give a score from 1, representing the lowest amount of the trait evident to 5, representing the highest. Raters were given a list of adjective-based benchmarks (e.g., very low, low, intermediate, high, very high) and scored a further five clips for each trait. For certain traits, the reliability of scoring improved when benchmarks were provided (e.g., Motivation and Independence), indicating that their inclusion may potentially reduce ambivalence in scoring, ambiguity of meanings, and cognitive difficulty for raters. However, this effect was not universal, with the ratings of some traits remaining unchanged (e.g., Control), or even reducing in reliability (e.g., Distraction). There were also some differences between VS and HAS (e.g., Confidence reliability increased for

OPEN ACCESS

Edited by:
Nathaniel James Hall, Texas Tech University, United States

Reviewed by:
Erik Wilsson, Swedish Armed Forces, Sweden
Samuel Gosling, University of Texas at Austin, United States

***Correspondence:**
Nicola J. Rooney
nicola.rooney@bristol.ac.uk

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences, a section of the journal Frontiers in Veterinary Science

Received: *24 March 2020*
Accepted: *08 January 2021*
Published: *02 February 2021*

Citation:
Clark CCA and Rooney NJ (2021) Does Benchmarking of Rating Scales Improve Ratings of Search Performance Given by Specialist Search Dog Handlers? Front. Vet. Sci. 8:545398. doi: 10.3389/fvets.2021.545398

VS raters and decreased for HAS raters). There were few improvements in the spread of scores across the range, but some indication of more favorable scoring. This was a small study of operational handlers and trainers utilizing training video footage from realistic operational environments, and there are potential cofounding effects. We discuss possible causal factors, including issues specific to raters and possible deficiencies in the chosen benchmarks, and suggest ways to further improve the effectiveness of rating scales. This study illustrates why it is vitally important to validate all aspects of rating scale design, even if they may seem inconsequential, as relatively small changes to the amount and type of information provided to raters can have both positive and negative impacts on the data obtained.

Keywords: search dog, detection dog, rating scale, performance, measurement, scale design, quantification

INTRODUCTION

Rating scales are used across numerous fields to assess differences between individuals (human and animal), e.g., in the occurrence of particular behaviors or medical conditions (1, 2), the degree of pain experienced (or inferred in the case of animals) (3, 4), mood and quality of life (5–7), marketing preferences (8, 9), as well as being widely used to assess performance in specific tasks or roles (10, 11). They are used widely when quantifying the performance of working dogs both in selection tests [e.g., (12–14)] and in their working role [e.g., (15, 16)].

Search, or detection, dogs are used for many purposes, for example: to locate target species in wildlife conservation (17); in human medicine, to identify patients with cancer (18); or to assist people with medical conditions (19–21); and by various law enforcement agencies to find people, drugs, money and explosives [e.g., (15, 22, 23)]. Monitoring of search dog performance is essential to maintain the effectiveness of individual dog-handler teams by highlighting any short-term training needs, but it is also critical to direct longer term strategies for improving ability of working dogs. To ensure that ratings provide accurate and reliable information it is important that any performance measurement tool is designed appropriately, with relevant and quantifiable measures, which accurately reflect differences between the subjects being rated. Irrespective of the purpose there are two elements to the rating process, and therefore two main potential sources of error or variance: the design of the measurement tool (e.g., rating scale, questionnaire or survey), and factors associated with the observer/rater. Elsewhere we deal with the latter (24), here we are primarily interested in the former, the measurement tool, and how manipulating specific aspects of rating scale design can affect the ability of observers (in this case dog handlers) to distinguish between differing levels of performance.

There is a growing body of research exploring and validating scales for rating dog behavior during temperament and behavior tests. Researchers have systematically examined how best to quantify dog behaviors [e.g., (14, 25–27)], demonstrating, for example, that rating scales used by trained observers (14), or researchers (28) provide ratings similar to those made by working dog experts and that scales are as successful as behavioral coding in predicting which dogs would be successfully selected as odor detection dogs at 12 months of age (27). However, when investigating which factors affect and best predict working ability, behavioral measures are usually compared against training or practitioners' organizations' own measures of success. Whilst some studies explore predictors of successful acceptance into training (27), other explore predictors of successful certification (23), binary outcomes, which although practically very important, lack granulation. Other studies rely on scales devised by the working dog organization, such as those used in competitive hunting trials (16), which were often formed historically without scientific input and without thorough consideration of how their design may influence potential sources of error and the quality of information obtained. There is considerable evidence that seemingly small changes in scale design can alter the way raters interpret and use scales, therefore affecting the reliability and validity of data (29, 30). Application and investigation of these principles when rating working dog performance will allow us to devise meaningful scales for investigating factors impacting performance.

In Rooney and Clark (31), we detailed a systematic process of selecting and testing suitable behavioral trait measures (e.g., Motivation to search, Confidence in the environment) for dogs trained to search for explosives on/in vehicles (VS dogs), and high assurance search (HAS) dogs, trained to detect buried improvised explosive devices (IEDs). These instruments were designed as a method of recording day-to-day variation in performance, using the most appropriate traits for each search classification. We found good reliability for assessing dog performance within a group of raters, but it would appear that some raters in the group were better able to use the 1–5 scales reliably than others and the predicted reliability if a single rater were to provide scores was poor. Therefore, we could not be confident that individual handlers could provide comparable data. As handlers often work alone, it is important to make the measures practically applicable and viable for a single a handler to use and our aim here was to find a method of improving the use of the scales at the individual level and increase single rater reliability to an acceptable threshold.

One plausible reason for low single rater reliability (31) was that the raters may not have agreed on, or even understood some of the measurement traits, making it difficult to reliably categorize the behavior they observed into points on a 1–5 scale. Where raters have difficulty in conceptualizing an aspect of performance they may resort to careless rating, bias (e.g., halo or leniency), or using a restricted range of values on the scale - typically mid-range to positive range (net acquiescence) or at the extremes (8, 32). Our previous study of handler ratings found evidence for range restriction in the use of the scales (24, 31), which could reduce accuracy (real performance compared to scores), agreement between raters, and the ability to distinguish between performances - as ratees are, in effect, being scored on much smaller scales than intended. In order for ratings to be reliable, meanings of traits and any other performance measures must therefore be clear.

How raters encode, organize, integrate, recall and evaluate information involves an "on-line" (33) or internal evaluation, where categorization or judgements are made, based on raters' own idiosyncratic understanding of a trait or concept (34). To assign a performance score, behaviors which generally occur on a continuum are assessed according to which of several discrete (i.e., non-continuous) categories they most closely match. To maximize the accuracy of this categorization it is therefore important to bring each rater's idiosyncratic categorization (based on their own internal mental representation) closer to a common understanding of which behaviors constitute specific levels of each dimension of performance. If this can be achieved then we would expect raters shown examples of behavior at each level of performance to be able to utilize the full range of the scale.

Providing verbal benchmarks (anchors or labels) as opposed to just providing a numerical scale, has been shown to help raters to select an appropriate category, increasing reliability and validity [e.g., (35, 36)]. These verbal anchors can be single words or short descriptions and are generally "adjective-based" (good, poor, high, low, average) or "descriptive," providing details about the construct and what each level of performance means (37). Adding descriptive anchors for each level of the behavioral measure should help to clarify the meaning of traits, removing an interpretive step where meanings could be confused between raters, also reducing the cognitive burden of the rater in interpreting their internal mental representation into a point on a scale (38). The use of verbal anchors has been shown to alter the way in which raters use scales (39) and has been recommended to improve agreement between raters (40) and reduce rater bias (32). Descriptive anchors in particular, may be effective at preventing leniency error (37), which is commonly reported in the literature [e.g., (10)] and has also been found to occur when dog handlers rate their own dogs (24). However, the selection of appropriate benchmarks requires careful consideration as providing insufficient detail, too much detail, or altering scales so that they that become emotionally valenced (e.g., too critical) can in fact be detrimental to accuracy and discrimination between ratees (37, 39, 41). In Rooney and Clark (31) raters were asked to apply identical, basic (one or two word) adjective-based descriptors (very low, low, intermediate, high, and very high) for every behavioral measure. The lack of rater reliability may have been a consequence of providing inadequate benchmarks to provide raters with a common concept of each level of performance. This may be a particular issue for behavioral traits that are harder to conceptualize, or where raters are likely to disagree with each other in their meaning. The next stage in developing a dog performance rating tools was therefore to explore the value of providing observers with more detailed verbal benchmarks for each of the levels within every behavioral performance measure, as a method of potentially bringing rater categorizations of performance closer together.

There are of course other considerations and aspects of scale design which can be affected by the addition of benchmarks. These include, the number of scale points, whether some or all points on the scale are benchmarked, and also whether scales are balanced (where the midpoint of the scale equates to the conceptual midpoint). We previously decided that 5-point scales would provide a reasonable trade-off between obtaining enough information and the practicalities of field-based assessment (31). Numerous studies have reported that using scales of typically 5 or 7 points minimizes variability in scale use, maximizes inter and intra test reliability, as well as optimizing cognitive comprehension by raters (41–46). We chose to label all points on the scale, as this is generally considered better than benchmarking only selected points such as the extremes (41, 45). It is also practically feasible to benchmark scales of this size, whereas benchmarking larger scales (particularly if all points are labeled) will increase cognitive burden on the rater (29) and may lead to reduced accuracy due to rating fatigue. Unbalanced scales, where the midpoint of the scale does not equate to the conceptual midpoint of the level of performance, may be useful in discriminating between ratees in a negatively skewed population (47), i.e., where none of the ratees is expected to score at the lowest extremes. But this was not relevant here, as we required discrimination across the full range of the scale.

Although benchmarking is often used in rating scales, including those used on working dogs its efficacy is rarely assessed. Here, we tested the value of providing benchmarks to performance rating scales for two types of explosives search dog. Our aim was to test if, by providing observers with benchmarks for every level (1–5 scale) of search performance, we can: [1] bring raters' conceptualization of the traits closer together, as evidenced by an increase in inter-rater reliability; and [2], increase the ability of raters to discriminate between levels within performance measures, reducing the effect of rating range restriction as measured by a greater use of the 1–5 scale (increased standard deviation). Our observers rated the performance of VS and HAS dogs in training searches using 11 performance measures derived previously (31). Previous work showed that raters of VS dogs assigned differing importance to each of the 11 traits, as well as showing differing levels of reliability for each trait, compared to a raters of HAS dogs. Thus, the impact of benchmarking is likely to differ with the type of search dog being rated. We compared ratings for 10 videoed searches per trait, selected to show as wide a range of performance as possible. The first five for each trait were rated without benchmarks and the second set of five were rated with benchmarks. This was repeated for all 11 traits, therefore each group of raters (VS and HAS) watched and scored 110 video clips in total. This was an opportune study conducted on experienced

TABLE 1 | Behavioral measures of performance (traits) in the order they were scored.

Behavior measures	
Short title	**Full title and description of behavioral trait**
Control	Control (responsiveness to verbal and or physical commands). The proportion of commands obeyed and speed of response.
Motivation	Motivation (enthusiasm to search). How keen or eager the dog is to search – assessed from the dog's behavior leading up to and at the start of the search.
Distraction	Distraction when searching. A distraction is anything that takes the dog's attention away from searching or from starting to search, including urinating.
Search pattern	Ability to follow search pattern. How well the dog follows the correct search pattern, without missing areas or needing constant correction. Not following search pattern would include: HAS, pulling off-line, wide back-seek, or following visual cues; VS, pulling/moving away from vehicle being searched, searching ground, or not searching "overlap."
Stamina	Stamina throughout search. How much motivation or enthusiasm decreases over the search, e.g., due to tiredness or loss of confidence.
Indication	Strength of indication.
Confidence	Confidence (absence of fear/anxiety) How confident or relaxed the dog is.
Thoroughness	Thoroughness of search. How much of the search the dog is actively searching: HAS, sniffing with its head down and nose to the ground for the entire search, including on the back-seek and searching right up to the handler; VS, sniffing with nose to the vehicle.
Independence	Independence. Ability of the dog to search without guidance, (not needing, or looking for, constant guidance), including being able to continue searching when further away from handler and on back-seek.
Speed	Speed of search
Detect & locate	Ability to detect and locate scent to source

Full titles and descriptions are as presented to the raters, but for the sake of brevity the behaviors are referred to in the text by their shortened title (in bold).

dog handlers observing dogs in operational environments as part of their own training. It was therefore not possible to randomize the order of video presentation.

METHODS

Behavioral Measures of Performance

The behavioral measures had been obtained by a systematic process of scale derivation, involving detailed interviews and questionnaires with stakeholders (e.g., trainers, handlers, senior staff) [see (31)]. From this, 12 behavioral trait measures were derived, we selected 11 of these, which could also be scored from short videoed searches on a 1–5 scale (**Table 1**). Consistency in searching behavior was not included as it could only be assessed from whole searches, not short clips. We did not include the composite measure "Overall Performance" for the same reason, and also because it does not represent a single independent dimension of performance.

Search Videos

Video recordings were made (using Sony Handycam DCR-SR58) of 200 training and accreditation searches (117 VS, 91 HAS), performed by 62 different dogs (35 VS, 27 HAS) in 100 different handler-dog pairings (50 VS; 50 HAS). The same videos had been used to make 17 5 min clips to develop behavioral scales in Rooney and Clark (31); although to avoid repetition different searches or sections within searches were used wherever possible. For each of the 11 behavioral traits we extracted 10 short video clips (each ∼30 s), with the aim of illustrating each point on the 1–5 scale, or as wide a range of performance as possible (110 short clips in total). These clips were to be used in a training resource for military dog handling personnel ahead of overseas placement. Both authors viewed and rated the clips and based on their assessments, videos showing a range of ratings were balanced across the pre and post benchmark conditions, with the order shown in the particular set of clips randomized by performance level (i.e., 1, 2, 3, 4, 5).

Raters

The majority of raters were military (or ex-military) personnel, with the exception of two raters per group who were civilian trainers working within a military establishment. Many of the raters had experience of both VS and HAS, but individuals were assigned to either classification observation group, with only one person appearing in both groups. Raters were all experienced in the classification being studied (16 VS, mean experience with VS 3.2 years, 11 HAS, mean experience 2.3 years) as either dog trainers, course instructors (training search-dog handlers), or as dog handlers. Many had experience of assessing and recording performance, but not using the methods or rating scales used here, although most had been raters in another study on one previous occasion (see Video observations).

Video Observations

All observations were performed at the Defense Animal Center (DAC) (Leicestershire, UK), in three sessions (April, May and July 2013). Subjects attended in groups of between 1 and 11 participants. Each session lasted ∼3 h, with two breaks in each as close to an hour apart as possible without disrupting the task. All but three (two HAS, one VS) of the observers had taken part in a previous rating experiment (31) using the same behavioral traits. For 11 subjects this had been the day before, for 11 observers it had been between 5 and 12 weeks previously, and for 2 VS observers, 2 h previously. The first task gave some experience of rating the behavioral traits, but the video clips were longer (∼6 min) and observers were required to rate all of the traits at once. They had received the same definitions of the traits as in the pre-benchmark condition here, without the detailed descriptions of each performance level, so it was assumed that this would not affect the question of whether the descriptors were effective. As the raters who had taken part the previous task had been briefed on the purpose of rating performance and on common errors to try to avoid (e.g., halo), the three new subjects underwent a similar briefing. The instructions below were also reiterated to all participants before observations began.

When rating performance subjects were urged to:

a) assess each performance trait in isolation;
b) avoid being affected by an overall good or bad impression (halo effect), or being overly influenced by individual events;
c) avoid being influenced by any prior knowledge they had of the dog;
d) use the whole 1–5 scale whenever possible (e.g., avoid using just middle ranges);
e) assess the performance of the dog (not handler) in the particular search shown (not prior knowledge);
f) watch the whole clip before scoring any behaviors and assess performance based on the entire clip;
g) score the videos in silence to avoid influencing each other's scores.

Subjects were shown 10 videos of ∼30 s in duration for each of 11 performance measures, starting with Control (for order of presentation see **Table 1**) and moving sequentially through to Detect & Locate. When each clip ended, the observers were asked to write their score for the particular performance trait on a recording sheet and this was repeated until five videos were rated. Subjects were then handed a list of benchmarks, or anchored terms, describing the 1–5 levels of that particular trait and asked to rate a further five videos with the anchored benchmarks to aid them. Benchmarks had been derived by the authors after watching and discussing the range of performance for each trait. They were deliberately kept as short adjective based sentences expanding on the original basic (one/two word) anchors. For example: Distraction, from [1] Very low - not distracted at all, through to [5] Very high - highly distracted; Motivation, from [1] Very low - no enthusiasm to search, to [5] Very high - very enthusiastic to search. After each trait had been scored, subjects were encouraged to discuss within the group how easy or difficult they found using the benchmarks and whether they felt that the benchmarks correctly described the different levels of performance. Due to time limitations of using expert handlers, all subjects watched the videos in the same order. We did not randomize or balance the order of the two conditions (benchmarked or not), as we anticipated there would be strong carry over effects after benchmarks were introduced.

Statistical Methods

Analyses were performed for each trait within each classification (IBM SPSS Statistics 21), to answer the following questions:

1) **Does providing benchmarks increase rater reliability?**
We expected between-rater agreement, or reliability, to increase when subjects had the benchmarks for reference, as their idea of what constitutes the different levels should become more similar. We tested this by visual comparison of intra-class correlation coefficients (ICCs, two-way random effects with absolute agreement) in pre- and post-benchmark conditions. Average measure ICCs indicate how reliably a group of raters rated each of the traits (48), but cannot be generalized to indicate how well a single rater would perform. We therefore used single-rater ICCs, although average rater values are included to allow comparison with previous studies [e.g., (31)]. Reliabilities of > 0.7 were taken to indicate strong agreement.
2) **Does providing benchmarks change the range of ratings?**
   I. Range restriction. If raters were using a greater range of the 1–5 scale (less range restriction) the spread of scores around the mean (standard deviation) will increase post-benchmark, as indicated by a significant change in standard deviation (SD) from pre- to post- benchmarking conditions (paired *t*-tests).
   II. Mean scores. Univariate GLM, with pre/post benchmarks as the fixed factor and rater ID as a random factor, were used to test for a change in mean scores in pre- and post-benchmark conditions. As the videos were balanced across conditions the mean should "∼3" for each trait in both pre and post conditions, but mean ratings might change as raters adjust their perception of the 1–5 categories within each behavior. If the raters were restricting ratings to a particular part of the scale then we would expect the mean score to change in the post-benchmarked condition; for example, if benchmarking reduces net acquiescence (use of mid to higher end of scale), mean scores should decrease.

RESULTS

Overall, average rater reliabilities were very high for both HAS and VS ratings; indicating good agreement amongst the group of raters (**Table 2**). Single rater ICCs were above the 0.7 threshold for strong reliability for 7 VS traits and 9 HAS traits, indicating that we could expect individual raters to produce reliable ratings. The exceptions to this were Independence and Speed in both classifications, as well-Stamina and Detect & locate for VS.

Does Providing Benchmarks Increase Rater Reliability?

In the post-benchmarking condition, four VS traits improved noticeably in agreement (Motivation, Confidence, Independence, and Detect & Locate; **Table 2**), but three had lower levels of agreement (Distraction, Thoroughness and Speed). Distraction and Thoroughness did not reach 0.7 after adding benchmarks, despite both exceeding this threshold in the pre-benchmark condition. There were small positive changes in Stamina (enough to bring it over the 0.7 threshold for strong reliability) and Search Pattern, and similar changes - in the opposite direction - for Control and Indication.

The reliability of four HAS traits noticeably improved post-benchmarks (Motivation, Independence, Speed and Indication), whereas three had lower agreement (Distraction, Confidence and Search Pattern). Although Confidence and Independence did not reach 0.7 with the addition of benchmarks, the latter improved considerably (from 0.385 to 0.629). There were negligible changes in reliability for Control and Thoroughness (decreased), and Stamina and Detect & Locate (increased).

TABLE 2 | Single rater agreement (ICC) between 11 raters, with 95% confidence intervals (upper and lower bounds) and average rater ICC for comparison.

VS *N* = 17	ICC	Control	Motivation	Distraction	Search pattern	Stamina	Indication	Confidence	Thoroughness	Independence	Speed	Detect & locate
Pre-benchmarks	Single value	0.930	0.744	0.837	0.755	0.690	0.900	0.722	0.719	0.561	0.662	0.597
	Lower bound	0.817	0.484	0.628	0.500	0.416	0.748	0.452	0.452	0.280	0.384	0.315
	Upper bound	0.991	0.961	0.977	0.963	0.950	0.987	0.957	0.956	0.918	0.944	0.928
	Average value	0.995	0.979	0.988	0.980	0.973	0.993	0.976	0.976	0.953	0.969	0.960
Post-benchmarks	Single value	0.877	0.892	0.633	0.782	0.764	0.854	0.850	0.554	0.826	0.780	0.772
	Lower bound	0.702	0.733	0.353	0.539	0.512	0.657	0.651	0.277	0.609	0.534	0.525
	Upper bound	0.984	0.986	0.937	0.968	0.965	0.980	0.980	0.915	0.976	0.968	0.966
	Average value	0.991	0.992	0.965	0.983	0.981	0.989	0.898	0.952	0.987	0.983	0.982
HAS *N* = 11	**ICC**	**Control**	**Motivation**	**Distraction**	**Search pattern**	**Stamina**	**Indication**	**Confidence**	**Thoroughness**	**Independence**	**Speed**	**Detect & locate**
Pre- benchmarks	Single value	0.843	0.706	0.874	0.912	0.770	0.797	0.826	0.922	0.385	0.687	0.857
	Lower bound	0.629	0.422	0.686	0.768	0.503	0.550	0.596	0.789	0.129	0.390	0.651
	Upper bound	0.979	0.954	0.983	0.989	0.966	0.971	0.976	0.990	0.620	0.950	0.985
	Average value	0.983	0.964	0.987	0.991	0.974	0.977	0.981	0.992	0.873	0.960	0.985
Post-benchmarks	Single value	0.823	0.912	0.823	0.786	0.779	0.852	0.682	0.911	0.629	0.837	0.884
	Lower bound	0.592	0.768	0.590	0.532	0.522	0.645	0.392	0.767	0.327	0.618	0.708
	Upper bound	0.975	0.989	0.975	0.969	0.968	0.980	0.949	0.989	0.937	0.978	0.985
	Average value	0.981	0.991	0.975	0.981	0.959	0.949	0.976	0.991	0.983	0.988	0.984

Highlighting indicates an increase (darker shading) or decrease (paler) in ICC of at least 0.1 in the post-benchmark condition. All ICCs were significant at P < 0.001.

Does Providing Benchmarks Change the Range of Ratings Used?

There was a significantly greater spread of scores around the mean (standard deviation) for Motivation in the post-benchmark condition, but lower spread for Control and Confidence and a tendency for Distraction (**Table 3**). With benchmarks, the VS observers rated Confidence, Independence and Thoroughness higher and tended to also rate Indication higher; whereas, they rated Distraction (and tended to rate Stamina) as lower. There were significant effects of rater identity on ratings for Stamina ($p = 0.011$), Confidence ($p = 0.033$), and Speed ($p = 0.012$).

For HAS raters, the only behavior where there was a significant difference in the spread of scores around the mean (standard deviation) was Motivation, where observers used a wider range of values in the post-benchmark condition. There was a trend in the same direction for Stamina, but in the opposite direction for Distraction, with observers tending to use a narrower range of scores when benchmarks were included. They rated Stamina and Distraction lower with benchmarks, and rated Confidence, Search Pattern, Speed, Detect & Locate, and Indication higher. There were significant effects of rater identity on Motivation ($p = 0.014$), Distraction ($p = 0.010$), Search Pattern ($p = 0.036$), Indication ($p < 0.001$), which with the exception of Confidence and Independence, coincides with the behaviors showing the greatest change in ICC.

DISCUSSION

In the pre-benchmarking condition average rater reliabilities were generally very high (>0.7) for both classifications, with single rater ICCs above the 0.7 threshold for strong reliability for most behavioral traits. This means that an individual rater within the group is likely to show good reliability in their ratings. The exceptions to this were Independence and Speed in both classifications, and for VS, ratings for Stamina (although this was very close to the threshold at 0.69) and Detect & Locate. Using benchmarks has the potential to alter how the handlers rated several of the traits, as evidenced by changes in rater agreement and in differences between scores. Motivation for example, showed an increase in reliability amongst raters of both VS and HAS classifications when benchmarks were provided, and the spread of scores increased in the post-benchmarking condition without any change in mean scores, suggesting that as well as bringing raters together in their understanding of the trait, they were also better able to use the full range of the scale. Thus, for Motivation the use of benchmarks achieved the initial aims. However, the benefit of benchmarking was not universal, with the size and direction of effects varying between the VS and HAS groups and according to the behavioral trait being rated.

Does Providing Benchmarks Increase Rater Reliability?

Improvements in reliability occurred when benchmarks were provided for Motivation and Independence (HAS and VS), Speed and Indication (HAS only), and Confidence and Detect & Locate (VS only). As agreement was higher, the benchmarks appeared to bring the raters' interpretation of category meanings for these traits closer together. It seems logical that the relative improvement in the post-benchmark condition should be greater for traits that may be conceptually harder to rate, such as

TABLE 3 | Difference in mean scores and standard deviation (within observer) between pre- and post-benchmark conditions for each performance trait (Univariate GLM, paired *t*-test); where the difference is significant the higher mean value is shown in bold and trends[t] (also in italics).

VS	**Difference between mean scores**				**Difference in standard deviation**			
Behavior	**Mean pre-benchmarks**	**Mean post-benchmarks**	***F*-statistic**	**Significance**	**SD pre-benchmarks**	**SD post-benchmarks**	***t*-statistic**	**Significance**
Control	2.713	2.763	0.556	0.468	**1.701**	1.470	2.780	0.014
Motivation	2.930	2.912	0.072	0.791	1.407	**1.587**	−2.40	0.030
Distraction	**3.550**	3.163	5.333	0.036	*1.654*[t]	1.418	2.113	0.052
Search pattern	2.838	2.900	3.021	0.103	1.413	1.441	−0.227	0.823
Stamina	*3.275*[t]	3.088	4.494	0.051	1.339	1.236	1.139	0.273
Indication	3.075	*3.275*[t]	3.750	0.072	1.585	1.433	1.637	0.122
Confidence	2.988	**3.363**	32.767	<0.001	**1.843**	1.453	4.543	<0.001
Thoroughness	2.825	**3.362**	16.44	0.001	1.351	1.303	0.390	0.702
Independence	3.225	**3.613**	13.874	0.002	1.266	1.408	−1.420	0.176
Speed	3.212	3.206	0.004	0.952	1.271	1.203	0.632	0.537
Detect & locate	2.863	2.913	0.128	0.725	1.552	1.492	0.777	0.449
HAS	**Difference between mean scores**				**Difference in standard deviation**			
Behavior	**Mean pre-benchmarks**	**Mean post-benchmarks**	***F*-statistic**	**Significance**	**SD pre-benchmarks**	**SD post-benchmarks**	***t*-statistic**	**Significance**
Control	2.545	2.655	0.803	0.391	1.506	1.460	0.571	0.581
Motivation	2.982	2.982	0.000	1.000	1.387	**1.560**	−2.893	0.016
Distraction	**3.491**	3.055	20.426	0.001	*1.633* [t]	1.495	2.029	0.070
Search pattern	2.964	**3.218**	12.564	0.005	1.625	1.376	1.874	0.090
Stamina	**3.618**	3.073	14.063	0.004	1.068	*1.300* [t]	−2.144	0.058
Indication	2.636	**3.309**	190.139	<0.001	1.380	1.421	−0.247	0.810
Confidence	2.945	**3.405**	7.183	0.023	1.440	1.476	−0.492	0.633
Thoroughness	2.691	2.782	2.119	0.176	1.670	1.640	0.368	0.720
Independence	2.782	2.927	0.907	0.363	1.189	1.29	0.649	0.531
Speed	2.564	**3.382**	21.182	0.001	1.101	1.100	0.019	0.985
Detect & locate	2.873	**3.291**	9.446	0.012	1.548	1.509	0.388	0.706

those that are more abstract and less easily quantifiable [see (31)]. For example, raters are likely to hold clearer a-priori representations of the difference between a score 3 and a score 4 for Control, a trait with high observability (49), where we expect them to already have a concept of differing levels of dogs' responsiveness to commands, compared to traits such as Independence and Motivation, which are conceptually more abstract or more evaluative (49). This was the case for these behavior traits: reliability for both Motivation and Independence improved with benchmarks for both dog classifications, whereas for Control ICCs changed very little and in fact, declined very slightly.

Whilst we may have expected any change in the reliability of less evaluative traits (e.g., Control) to be of smaller magnitude compared to traits where there was greater room for improvement, it is not immediately clear why benchmarks had no impact at all on traits such as Control (HAS and VS); or why agreement for Distraction (HAS and VS), Confidence and Search Pattern (HAS), and Thoroughness and Speed (VS), decreased. One possibility is that where there was negligible change in reliabilities, rater conceptions of the trait levels may have already matched the provided benchmarks, hence leading to no improvement. Alternatively, the lack of change or decrease in agreement for some behaviors could indicate a reluctance of some observers to change their a-priori assumptions (non-compliance) about what constitutes each level of performance even with the benchmarks in front of them. This could, in fact, prove to be a particular issue with very quantifiable traits, where raters might hold steadfast ideas of performance, and especially in this group of raters which included many with considerable experience and expertise, whilst less experienced raters may show differential effects. It is also possible that our adjective-based descriptors were insufficient to make a difference to these ratings. This requires further investigation.

There are also limitations to the study, in that the design was inevitably unbalanced, with the post-benchmarking condition having to come after the pre-benchmarking. This design was deliberately selected to avoid carry-over and memory effects, however a consequence is that some of the changes post-benchmarking may have resulted from raters having more practice with scoring. This could have been particularly important for those raters with less experience, who may not have seen many examples of dogs performing at the very poor end of the spectrum previously. Although the lack of any universal increase in agreement suggests that any "practice" effects were limited.

Were Reliabilities Comparable to Previous Studies?

Both average and single measures agreement was generally high, and higher than the author's previous studies using the same traits (24, 31), and variability between raters was lower. Even without benchmarking, the predicted ability of a single rater to reliably score performance was within the levels of acceptability (> 0.7) for most traits in this task, unlike in Rooney and Clark (31). This is considerably higher than most estimates of inter-rater agreement when rating aspects of dog behavior [e.g., see (50); see (25)], but similar to that found by Fratkin et al. (14) when rating search dog performance. Raters may have been better because, in this task the vast majority had previous experience of using the ratings; although as there seemed to be no universal benefit from practicing ratings between the pre- and post-benchmarked conditions (see above), the more likely explanation is that the task itself was inherently easier. Raters were focusing on one trait at a time here, rather than trying to remember multiple traits at once, and perhaps more importantly, here raters assessed behavior from very short clips designed to illustrate a particular level of performance, whereas in Rooney and Clark (31) they had to make an assessment based on assimilating 6 min of behavior during which time performance level could fluctuate. They may also be artificially inflated as the videos used in this study were selected to reflect the full range of each of the rating scales. This was deliberate as the same clips were to be used to train personnel how to rate searches using the full extent of the scale, however it is likely easier for judges to distinguish between behaviors that vary greatly in magnitude than those that are close to a midpoint, and normally encountered. Caution should therefore be exercised when extrapolating the results of this study to actual performance in the field, as the methodology in Rooney and Clark (31) is a closer representation to the actual task facing handlers. This raises an important point when developing measurement tools encompassing rating scales for several aspects of performance, as although raters may be reliable at scoring individual traits in an experimental setting, this may not equate to ability in the field. Additionally we further need to test the value of benchmarks when rating numerous traits simultaneously.

Does Providing Benchmarks Change the Range of Ratings Used?

The addition of benchmarking either did not affect the mean score, or lead to an increase, so that in the post-benchmarking condition most traits were scored higher than "3" on average (as opposed to lower than 3); which was more evident in HAS ratings. The opposite was true for Distraction, but unlike the other traits, low scores for this trait are more positively valenced (no or low Distraction is ideal). This suggests that some raters became more lenient when provided with benchmarks.

The quality of ratings and the degree of rater error are a reflection of the measurement tool, the rater, and the interaction between the two. Although our treatment altered the way that the measurement tool was presented to raters, it is likely that the impact of this change will vary between raters based on characteristics inherent to the individual. Previous studies have demonstrated that observers differ in their ability to rate traits (24, 31); therefore, it also seems likely that they will differ in their ability to effectively use benchmarks, and here we did detect several main effects of rater identity. Differences may simply be due to factors such as differing levels of experience of working with search dogs and assessing performance. However, the psychology literature commonly reports the occurrence of raters resorting to particular styles of responding [e.g., (51)], which can be pervasive, despite instruction on avoiding bias (52). Such response styles include scoring only within mid-to-positive range values (net acquiescence) and responding only at the extremes of the scale (8, 32).

While benchmarking might make all points on the scale equally salient and accessible, thus potentially reducing rater biases, it could also increase response style bias in some people. For example, negative extremes become more salient in benchmarked scales compared to when categories are unlabelled, which may lead to greater net acquiescence (41). Although our scales were labeled based purely on the amount of a particular trait rather than being explicitly positively or negatively valenced (i.e., good, poor etc.), dog handlers have shown leniency bias in ratings using these scales (24), and it is likely that they will naturally associate favorable and negative connotations with particular ends of the scale. For example, VS observers used a narrower range of scores for Control and Confidence when presented with benchmarks, with scores for the latter also being higher post-benchmarking. So for these traits, benchmarking appeared to cause some raters to be more reluctant to use scale extremes; and for Confidence, this was also associated with more positive ratings, which could be attributed to greater net acquiescence. Therefore, we would recommend that benchmarks cannot be universally applied with global benefit. The next step should be to try to understand changes as a result of benchmarking at the individual rater level and between classifications (e.g., comparing why HAS and VS observers differed).

Further Development to Increase Effectiveness of Benchmarks

The lack of improvement or decrease in agreement for some traits may also have been the result of deficiencies in those specific benchmarks used, leading to greater uncertainty in trait meanings or greater non-compliance in using them by observers who disagreed with the descriptions provided. Some benchmarks were perhaps not detailed enough, or too "generic." Our benchmarks were adjective based (e.g., Very, Low, Intermediate, High, and Very High) and although a "descriptive" statement followed, this was intentionally short (to allow inclusion in an operational data recording instrument) and in most cases also adjective-based (e.g., rarely, often, sometimes, usually, always). They may therefore have had limited benefit in helping raters to categorize behaviors and were open to subjective interpretation, e.g., when discussed during the session, raters felt that the benchmarks for Detect & Locate needed some alteration. A further point to consider is that ratings may also change over time, e.g., if the quality of dogs drifts, and it is therefore important to have quality control in place such as the video based training resource developed here which could be used to standard set and ensure temporal standardization in ratings.

In general, feedback from observers on the value of the benchmarks was positive: raters generally preferred scales with benchmarks; hence they may prove valuable at increasing compliance even if the scale use improvement is variable or unproven. We suggest that including more descriptive terminology in the benchmarks may be valuable. Descriptive statements elaborating on examples of the precise behaviors that constitute particular levels of performance have been found to increase reliability in ratings more than purely adjective-based benchmarks (37). The development of descriptive benchmarks is more time consuming than simple adjective-based terminology and care needs to be taken to ensure the validity of the scale [see (53)]: e.g., in scales intended to be balanced, that the scale midpoint equates to this conceptual performance midpoint (39, 53). Care must also be taken not to make the descriptions too lengthy as to increase rater cognitive burden and subsequent non-compliance or careless rating [e.g., (54)]. Descriptive benchmarks need to be derived using feedback from and discussion with raters to ensure the distinction between levels is meaningful and acceptable to the rating population. This could be achieved through an iterative process involving consultation with the end-users. For example, we could add a further step to examine whether benchmarked levels match rater perceptions, by asking raters to describe the different levels of performance in their own words. If levels are found to differ then the benchmarks should either be altered to match the observer levels (if these are in fact deemed to be the correct categorisations), or alternative methods employed to persuade and train raters to use the benchmarked levels.

Rater Training

A further and potentially valuable approach to improving reliability, in conjunction with benchmarking, is rater training. Simply providing the list of benchmarks was insufficient for at least some of the raters, who may benefit from being able to discuss and see examples of the difference between performance levels. Evaluative accuracy training focuses on increasing validity by moving respondent ratings closer to a reference or gold standard, through dimension-relevant judgements (33). Instead of the handlers comparing their internal categorization with a written list of benchmarks, their internal representation of the different levels of behaviors is altered to match the desired levels, making the process of categorizing behaviors using the new shared framework automatic and internal. Frame-of-reference (FOR) training is widely used in other disciplines and teaches raters to use a common conceptualization (or frame of reference) when observing and evaluating performance (38) providing gold-standard examples of the different levels has been proven to improve rater accuracy in many different scenarios (55–57), particularly when combined with anchored rating scales (38).

Limitations to the Methodology

The practical constraints of working with a time-limited operational cohort of raters meant that it was impossible to counterbalance the experimental design and randomize the order of presentation of videos. Therefore, all participants watched the searches in the same order, and for each trait the same examples were benchmarked. Efforts were made to balance the two sets of video based on previous ratings by handlers (31) and the authors' assessments. However, it remains possible that there were inherent differences between the sets of videos selected, which led to some of the detected differences between conditions. The sample sizes of both the number of videos observed per trait and the number of participants were relatively small. The results show traits displaying differential effects of benchmarking, which could in theory be attributed to initial differences between videos. In the absence of a control group, it is impossible to draw strong conclusions regarding the effect of benchmarking. We therefore conclude that similar to the way in which codings of behaviors have been demonstrated to lack the often presumed value relative to more subjective ratings [e.g., (27, 58)], the addition of adjective based descriptors here did not demonstrate clear universal value to rating scales.

There may be value in further investigating these concepts using larger groups of observers. Here, we did not have adequate sample sizes to explore the impact of rater experience and it may be that this will impact rater's ability to reliably use rating scales based on adjectives. Although, as discussed, pre-conceived ideas may also make experienced handlers less receptive to using the scales. Using researchers as well as trained dog handlers could facilitate obtaining a larger sample size as it has been demonstrated that their ability to rate searches is comparable to expert dog handlers (14, 28). This could enable a replication of this study with balanced presentation of benchmarked and unbenchmarked scales, randomization of videos and potentially examine the effect of experience rating dogs in using the scales. It would, however, not be possible to use the realistic operational training searches from the field used in this study as these are only viewable by military personnel. Therefore, despite the limitations in design, this study provides a rare opportunity to measure the application of rating scales in a realistic military environment.

CONCLUSIONS

Rating scales with and without benchmarks, are widely used in human and animal sciences, yet variable levels of consideration are given to how aspects of the design of the scale might alter the validity of the results obtained. In Rooney and Clark (31) we illustrated the importance of looking beyond overall correlations between behaviors when assessing scale validity and here we demonstrate that even relatively small additions to the amount of information given to raters can have important consequences for the data obtained. This study illustrates that to produce optimal performance measures, it is important to validate all aspects of design of the measurement tool and consider the amount and type of information provided to raters, as this can have both positive and negative impacts on ratings. The changes seen here were equivocal, but the feedback received from subjects suggests that handlers can benefit from additional information when scoring, especially for certain traits where providing benchmarks may potentially reduce ambivalence in scoring, ambiguity of meanings, and cognitive difficulty. However, benchmarking was not demonstrated to be universally valuable, and simply providing very basic adjective-based anchors may result in limited overall improvement and potentially more disagreement for some terms. Before performance measurement tools, such as the example developed here for the working dog community, are used, we recommend iterative development of benchmarks, given in conjunction with training such as Frame Of Reference, whereby raters can view and discuss differing levels of performance. This is likely to be the most effective method of improving rater reliability, by training those inexperienced in assessing performance as well as altering any pre-conceived ideas and bringing all raters closer to common conceptualization of the meaning of traits.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because they contain information on military working dog performance and are thus sensitive. Requests to access the datasets should be directed to Nicola J. Rooney, nicola.rooney@bristol.ac.uk.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Faculty of Medical & Veterinary Sciences (FMVS) Research Ethics Committee. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. The animal study was reviewed and approved by Animal Welfare & Ethical Review Board. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

CC and NR: conception of idea, methods development, and preparation of manuscript. CC: majority of data collection and analysis.

FUNDING

This study was supported by UK Ministry of Defence (MOD). Quantification of Performance – VS and HASD dog roles Ref DFAS RK7283.

ACKNOWLEDGMENTS

We would like to thank personnel of the 1st Military Working Dogs Regiment (Royal Army Vet Corps) and personnel of the Defence Animal Centre and the Royal Air Force for willingly giving their time and sharing their expertise for during both the video collection and rating experiment. We would also like to thanks Professor Bill Browne for statistical advice, and Colonel Neil Smith and Dr. Steve Nicklin for valuable comments on the manuscript.

REFERENCES

1. Flynn D, van Schaik P, van Wersch A. A comparison of multi-item Likert and visual analogue scales for the assessment of transactionally defined coping function. *Eur J Psychol Assess*. (2004) 20:49–58. doi: 10.1027/1015-5759.20.1.49
2. Kaler J, Wassink GJ, Green LE. The inter- and intra-observer reliability of a locomotion scoring scale for sheep. *Vet J*. (2009) 180:189–94. doi: 10.1016/j.tvjl.2007.12.028
3. Bellamy N, Campbell J, Syrotuik J. Comparative study of self-rating pain scales in osteoarthritis patients. *Curr Med Res Opin*. (1999) 15:113–9. doi: 10.1185/03007999909113371
4. Cambridge AJ, Tobias KM, Newberry RC, Sarkar DK. Subjective and objective measurements of postoperative pain in cats. *J Am Vet Med Assoc*. (2000) 217:685–90. doi: 10.2460/javma.2000.217.685
5. Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of likert and visual analogue scales for measuring change in function. *J Chronic Dis*. (1987) 40:1129–33. doi: 10.1016/0021-9681(87)90080-4
6. McCormack HML, Horne DJ, Sheather S. Clinical applications of visual analogue scales: a critical review. *Psychol Med*. (1988) 18:1007–19. doi: 10.1017/S0033291700009934
7. Mazaheri M, Theuns P. Effects of varying response formats on self-ratings of life-satisfaction. *Soc Indic Res*. (2009) 90:381–95. doi: 10.1007/s11205-008-9263-2
8. Shulman A. A comparison of two scales on extremity response bias. *Pub Opin Quart*. (1973) 37:407–12. doi: 10.1086/268102
9. Westbrook RA, Oliver RL. Developing better measures of consumer satisfaction: some preliminary results. *Adv Consumer Res*. (1981) 8:94–9.
10. Jawahar IM, Williams CR. Where all the children are above average: the performance appraisal purpose effect. *Pers Psychol*. (1997) 50:905–25. doi: 10.1111/j.1744-6570.1997.tb01487.x
11. Davis DA, Mazmanian PE, Fordis M, Van Harrison RR, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA: J Am Med Assoc*. (2006) 296:1094–102. doi: 10.1001/jama.296.9.1094
12. Goddard ME, Beilharz RG. Early prediction of adult behaviour in potential guide dogs. *Appl Anim Behav Sci*. (1986) 15:247–60. doi: 10.1016/0168-1591(86)90095-X
13. Weiss E, Greenberg G. Service dog selection tests: effectiveness for dogs from animal shelters. *Appl Anim Behav Sci*. (1997) 53:297–308. doi: 10.1016/S0168-1591(96)01176-8
14. Fratkin JL, Sinn DL, Thomas S, Hilliard S, Olson Z, Gosling SD. Do you see what i see?: Can non-experts with minimal training reproduce expert ratings in behavioral assessments of working dogs? *Behav Process*. (2015) 110:105–16. doi: 10.1016/j.beproc.2014.09.028
15. Haverbeke A, Messaoudi F, Depiereux E, Stevens M, Giffroy JM, Diederich C. Efficiency of working dogs undergoing a new human familiarization and training program. *J Vet Behav Clin App Res*. (2010) 5:112–9. doi: 10.1016/j.jveb.2009.08.008
16. Pallotti S, Riganelli S, Antonini S, Valbonesi A, Renieri C. Estimates of non-genetic effects for measures of hunting performance in short-haired and rough-haired Italian hound. *Italian J Animal Sci*. (2020) 19:439–46. doi: 10.1080/1828051X.2020.1748526
17. Arnett EB. A Preliminary Evaluation on the use of dogs to recover bat fatalities at wind energy facilities. *Wildl Soc Bull*. (2006) 34:1440–5. doi: 10.2193/0091-7648(2006)34[1440:APEOTU]2.0.CO;2
18. Willis CM, Church SM, Guest CM, Cook WA, McCarthy N, Bransbury AJ, et al. Olfactory detection of human bladder cancer by dogs: proof of principle study. *BMJ*. (2004) 329:712. doi: 10.1136/bmj.329.7468.712
19. Rooney NJ, Morant S, Guest C. Investigation into the Value of trained glycaemia alert dogs to clients with type I diabetes. *PLoS ONE*. (2013) 8:e69921. doi: 10.1371/journal.pone.0069921

20. Rooney N, Guest C, Swanson L, Morant S. How effective are trained dogs at alerting their owners to changes in blood glycaemic levels? Variations in performance of glycaemia alert dogs. *PLoS ONE*. (2019) 14:e0210092. doi: 10.1371/journal.pone.0210092
21. Wilson C, Morant S, Kane S, Pesterfield C, Guest C, Rooney NJ. An owner-independent investigation of diabetes alert dog performance. *Front Vet Sci*. (2019) 6:91. doi: 10.3389/fvets.2019.00091
22. Rooney NJ, Bradshaw JWS, Almey H. Attributes of specialist search dogs - A questionnaire survey of UK dog handlers and trainers. *J Forensic Sci.* (2004) 49:300–6. doi: 10.1520/JFS2003048
23. Sinn DL, Gosling SD, Hilliard S. Personality and performance in military working dogs: reliability and predictive validity of behavioral tests. *App Animal Behav Sci*. (2010) 127:51–65. doi: 10.1016/j.applanim.2010.08.007
24. Clark CCA, Sibbald N, Rooney NJ. Search dog handlers show positive bias when scoring their own dog's performance. *Front Vet Sci Working Dog Form Funct*. (2020) 7:612. doi: 10.3389/fvets.2020.00612
25. Jones AC, Gosling SD. Temperament and personality in dogs (Canis familiaris): A review and evaluation of past research. *Appl Anim Behav Sci.* (2005) 95:1–53. doi: 10.1016/j.applanim.2005.04.008
26. Kubinyi E, Gosling SD, Miklósi Á. A comparison of rating and coding behavioural traits in dogs. *Acta Biol Hung.* (2015) 66:27–40. doi: 10.1556/ABiol.66.2015.1.3
27. McGarrity ME, Sinn DL, Thomas SG, Marti CN, Gosling SD. Comparing the predictive validity of behavioral codings and behavioral ratings in a working-dog breeding program. *Appl Anim Behav Sci.* (2016) 179:82–94. doi: 10.1016/j.applanim.2016.03.013
28. Rooney NJ, Gaines SA, Bradshaw JWS, Penman S. Validation of a method for assessing the ability of trainee specialist search dogs. *Appl Anim Behav Sci.* (2007) 103:90–104. doi: 10.1016/j.applanim.2006.03.016
29. Krosnick JA. Question and questionnaire design. In: Marsden PV, Wright JD, editors. *Handbook of Servey Research.* Bingley: Emerald Group Publishing Ltd. (2010). p. 263–313.
30. Conti G, Pudney S. Survey design and the analysis of satisfaction. *Rev Econ Statist.* (2011) 93:1087–93. doi: 10.1162/REST_a_00202
31. Rooney NJ, Clark CCA. Development of a performance monitoring instrument for rating explosives search dog performance. *Front Vet Sci.* (in press).
32. Hoyt WT, Kerns MD. Magnitude and moderators of bias in observer ratings: a meta-analysis. *Psychol Methods.* (1999) 4:403–24. doi: 10.1037/1082-989X.4.4.403
33. Woehr DJ, Huffcutt AI. Rater training for performance appraisal: a quantitative review. *J Occup Organiz Psychol.* (1994) 67:189–205. doi: 10.1111/j.2044-8325.1994.tb00562.x
34. Roch SG, Woehr DJ, Mishra V, Kieszczynska U. Rater training revisited: an updated meta-analytic review of frame-of-reference training. *J Occup Organ Psychol*. (2012) 85:370–95. doi: 10.1111/j.2044-8325.2011.02045.x
35. Smith DE. Training programs for performance appraisal: a review. *Acad Manage Rev*. (1986) 11:22–40. doi: 10.5465/amr.1986.4282615
36. Krosnick JA. Survey research. *Annu Rev Psychol.* (1999) 50:537–67. doi: 10.1146/annurev.psych.50.1.537
37. Kuhlemeier H, Hemker B, van den Bergh H. Impact of verbal scale labels on the elevation and spread of performance ratings. *Appl Measure Educ.* (2013) 26:16–33. doi: 10.1080/08957347.2013.739425
38. Melchers KG, Lienhardt N, Von Aarburg M, Kleinmann M. Is more structure really better? A comparison of frame-of-reference training and descriptively anchored rating scales to improve interviewers' rating quality. *Pers Psychol.* (2011) 64:53–87. doi: 10.1111/j.1744-6570.2010.01202.x
39. French-Lazovik G, Gibson CL. Effects of verbally labeled anchor points on the distributional parameters of rating measures. *Appl Psychol Meas.* (1984) 8:49–57. doi: 10.1177/014662168400800106
40. Jonsson A, Svingby G. The use of scoring rubrics: Reliability, validity and educational consequences. *Educ Res Rev*. (2007) 2:130–44. doi: 10.1016/j.edurev.2007.05.002
41. Weijters B, Cabooter E, Schillewaert N. The effect of rating scale format on response styles: the number of response categories and response category labels. *Int J Res Market*. (2010) 27:236–47. doi: 10.1016/j.ijresmar.2010.02.004
42. Masters JR. The relationship between number of response categories and reliability of likert-type questionnaires. *J Educ Measure.* (1974) 11:49–53. doi: 10.1111/j.1745-3984.1974.tb00970.x
43. Birkett NJ. Selecting the number of response categories for a likert-type scale. In: *Proceedings of the Survey Research Methods Section*. Washington, DC: American Statistical Association (1986). p. 488–92.
44. Preston CC, Colman AM. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychol.* (2000) 104:1–15. doi: 10.1016/S0001-6918(99)00050-5
45. Weng L-J. Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educ Psychol Meas.* (2004) 64:956–72. doi: 10.1177/0013164404268674
46. Garratt AM, Helgeland J, Gulbrandsen P. Five-point scales outperform 10-point scales in a randomized comparison of item scaling for the patient experiences questionnaire. *J Clin Epidemiol.* (2011) 64:200–7. doi: 10.1016/j.jclinepi.2010.02.016
47. Kasten R, Weintraub Z. Rating errors and rating accuracy: a field experiment. *Hum Perform*. (1999) 12:137–53. doi: 10.1080/08959289909539864
48. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. (1979) 86:420–8. doi: 10.1037/0033-2909.86.2.420
49. John OP, Robins RW. Determinants of interjudge agreement on personality traits: the big five domains, observability, evaluativeness, and the unique perspective of the self. (1993) *J Pers*. 61:521–51. doi: 10.1111/j.1467-6494.1993.tb00781.x
50. Diesel G, Pfeiffer DU, Brodbelt D. Factors affecting the success of rehoming dogs in the UK during 2005. *Preven Vet Med.* (2008) (2008) 84:228–41. doi: 10.1016/j.prevetmed.2007.12.004
51. Van Vaerenbergh Y, Thomas TD. Response styles in survey research: a literature review of antecedents, consequences, and remedies. *Int J Pub Opin Res.* (2013) 25:195–217. doi: 10.1093/ijpor/eds021
52. Bernardin HJ, Cooke DK, Villanova P. Conscientiousness and agreeableness as predictors of rating leniency. *J Appl Psychol*. (2000) 85:232–6. doi: 10.1037/0021-9010.85.2.232
53. Landy FJ, Farr JL. Performance rating. *Psychol Bull*. (1980) 87:72–107. doi: 10.1037/0033-2909.87.1.72
54. Cook DA, Beckman TJ. Does scale length matter? A comparison of nine-versus five-point rating scales for the mini-CEX. *Adv Health Sci Educ*. (2009) 14:655–64. doi: 10.1007/s10459-008-9147-x
55. Spool MD. Training programs for observers of behavior: a review. *Pers Psychol.* (1978) 31:853–88. doi: 10.1111/j.1744-6570.1978.tb02128.x
56. Noonan LE, Sulsky LM. Impact of frame-of-reference and behavioral observation training on alternative training effectiveness criteria in a Canadian military sample. *Hum Perform*. (2001) 14:3–26. doi: 10.1207/S15327043HUP1401_02
57. Schlientz MD, Briesch AM, Riley-Tillman TC, Walcott CM, Chafouleas SM. The impact of training on the accuracy of direct behavior ratings (DBR). *Sch Psychol Q*. (2009) 24:73–83. doi: 10.1037/a0016255
58. Vazire S, Gosling SD, Dickey AS, Schaprio SJ. Measuring personality in nonhuman animals. In: Robins RW, Fraley RC, Krueger R, editors. *Handbook of Research Methods in Personality Psychology*. New York, NY: Guilford (2007). p. 190–206.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 22 April 2021
doi: 10.3389/fvets.2021.543463

The End of the Partnership With a Guide Dog: Emotional Responses, Effects on Quality of Life and Relationships With Subsequent Dogs

Janice Lloyd[1]*, Claire Budge[2], Steve La Grow[2] and Kevin Stafford[3]

[1] Discipline of Veterinary Sciences, James Cook University, Townsville, QLD, Australia, [2] College of Health, Massey University, Palmerston North, New Zealand, [3] Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand

Guide dogs are mobility aids that facilitate independent travel of people who are blind or visually impaired. Additional benefits imparted to the guide dog handler include companionship, and increased: social-function, self-esteem and confidence. Some evidence shows that the end of the guide dog partnership can result in reduced mobility, and may have profound psychosocial effects on the handler due to feelings of bereavement and loss of self-esteem. However, this evidence is limited. This study examined the experiences and feelings of 36 people across New Zealand, who experienced the ending of at least one partnership with a guide dog (77 pairings), to explore issues arising at the end of the partnership and how this may impact on relationships with subsequent dogs. Results indicate that the majority of handlers experienced a reduction in their quality of life due to a decrease in independent mobility followed by the loss of a friend and companion, curtailment of social interactions, and loss of self-esteem/confidence. The end of the partnership affected people in different ways. Most handlers "accepted" the partnership had ended, but some felt guilty or angry with the guide dog school. Most applied for another dog immediately, as the need for mobility was high, while others preferred to wait and a smaller number did not reapply. Feelings at this time also affected the handlers' relationships with subsequent guide dogs, with over a quarter expressing a negative effect. Retiring a guide dog (for whatever reason) is not only difficult for the handler, but also for the handler's family, friends, co-workers, and doubtlessly, the dog. The majority of handlers expressed feelings of extreme grief when the partnership ended, whether it was successful or not. Feelings of extreme grief were more common for first than subsequent dogs. The depth of emotion was compared to losing a family member or other loved one, which has been reported in some person and pet relationships. A better understanding of issues surrounding the end of the partnership, including the human-animal bond, will help inform the guide dog industry of how best to support their clients during this time and when transitioning to another dog. Findings may be applied to other service/assistance dog users and the pet owning community.

Keywords: guide dog, human—animal relationships, quality of life, grief, attachment, assistance dog, blind mobility aid, vision impaired

OPEN ACCESS

Edited by:
Lynette Arnason Hart, University of California, Davis, United States

Reviewed by:
George E. Moore, Purdue University, United States
Marie Jose Enders-Slegers, Open University of the Netherlands, Netherlands

***Correspondence:**
Janice Lloyd
janice.lloyd@jcu.edu.au

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences, a section of the journal Frontiers in Veterinary Science

Received: *17 March 2020*
Accepted: *15 March 2021*
Published: *22 April 2021*

Citation:
Lloyd J, Budge C, La Grow S and Stafford K (2021) The End of the Partnership With a Guide Dog: Emotional Responses, Effects on Quality of Life and Relationships With Subsequent Dogs. Front. Vet. Sci. 8:543463. doi: 10.3389/fvets.2021.543463

INTRODUCTION

Guide dogs are primary mobility aids intended to enhance the lifestyle of people with a visual disability (blind or visually impaired) by facilitating independent travel. Additional benefits imparted to the guide dog handler include friendship, companionship, increased social-function, and improved self-esteem and confidence (1–10). More has been published about guide dog usage in the scientific literature than about other types of service/assistance dogs. According to York and Whiteside (11) this work has leaned toward aspects of training, health and reproduction, and benefits to mobility and well-being, with less on the experience of owning a guide dog. Compatibility and the success or failure of the relationship between a person and their first guide dog was assessed by Lloyd et al. (12), and the complexities of successful and unsuccessful guide dog matching and partnerships were examined by Lloyd et al. (13) and Lloyd et al. (14). These, and other studies (11, 15) indicate that factors other than orientation and mobility, such as the dog's social behavior in and out of the home environment, need to be considered in the process of matching a dog to its new owner to promote a successful outcome. However, limited evidence exists that discusses how the handler might be affected when a guide dog retires, is returned to the guide dog training establishment (for whatever reason) or dies.

The end of the guide dog partnership can result in reduced mobility (1, 2) and may have profound psychosocial effects on the handler due to feelings of bereavement (6, 16–19), and loss of self-esteem (17, 20, 21). A seminal study by Nicholson et al. (22) examined distress arising from the end of a guide dog partnership and concluded that the emotions experienced by the handler at this time could be likened to feelings that follow the death of a pet, the loss of a close friend or relative or the loss of sight. These findings are supported by Kwong and Bartholomew (23) who explored individuals' relationships with an assistance dog and concluded that when confronted with the loss of their dog, people experienced intense grief consistent with the loss of a caregiving relationship. More recently, Uccheddu et al. (24) conducted a comprehensive analysis of grief responses in dog owners after the death of a pet dog and found that owners tended to humanize their pets and experienced a negative view of life after the death of their pet. The grief response to losing a dog, be it a pet or an assistance dog, is still an underestimated issue. Given the increasing number of service/assistance dogs being used across the world (25–27), this type of grief is of major concern for the welfare of the people who use them.

The present study[1] builds on these findings by discussing how feelings at the end of the guide dog partnership affect the handlers' relationships with subsequent guide dogs, and indicates trends in the dataset concerning multiple dog use. The effects of being without a guide dog, after experiencing guide dog mobility, on quality of life will also be discussed. A better understanding of the effects of the end of the partnership will help inform the guide dog industry of how best to support their clients during this time and when transitioning to another dog. Findings may be applied to other service/assistance dog organizations and pet (companion) dog ownership as the impact of the separation is similar in some aspects (28).

METHOD

Participant Recruitment

Customarily, guide dogs are well-accepted in New Zealand and guide dogs are provided to a wide-range of applicants with varying visual conditions and mobility needs (17). In order to apply for a guide dog, the applicant should be eligible to receive services from Blind Low Vision NZ (formerly Blind Foundation/Royal New Zealand Foundation of the Blind) by being blind or markedly sight impaired. Blind Low Vision NZ Guide Dogs is a member school of the International Guide Dog Federation (IGDF), and as such is accredited to the highest international standards. The population of interest, as previously described in Lloyd et al. (1), was all people living in New Zealand who were, or had been, clients of Blind Low Vision NZ Guide Dogs since its establishment in 1973. At the time of participant recruitment, this was ∼210 people. No exclusion criteria were applied. For reasons of privacy, a Blind Low Vision NZ staff member mailed the invitations to participate on behalf of the researcher (first author). The invitations consisted of an information document (supplied in the person's preferred format of Braille, audiotape, e-mail, or regular or large print), plus a consent form and a pre-paid, addressed envelope. Potential participants returned the signed consent form directly to the researcher, thus maximizing confidentiality and anonymity. Seventy two percent ($n = 151$) of the target group responded, from which 50 participants were randomly selected (i.e., around one quarter of the entire population of guide dog users in New Zealand at this time). Those not selected were notified and thanked.

Participants

Fifty people from across New Zealand (as described above) who had used one or more guide dogs were interviewed by the researcher either by telephone (78%) or face-to-face (22%) regarding their experiences with guide dogs. The total sample (people and dogs) is described in Lloyd et al. (13) and Lloyd et al. (14). Of these 50 people, 36 had experienced the ending of at least one partnership. These 36 people constitute the participants for the present study that explored experiences associated with the end of the guide dog partnership and how they affected subsequent matches. Just over half the sample identified as female (20, 55.6%), and the majority (25, 69.4%) identified themselves ethnically as New Zealanders of European decent, 6 (16.6%) as Māori (the indigenous people of Aotearoa/New Zealand), and the remainder as "other." They ranged in age from 28 to 80 years, with a mean age of 50.6 years (SD = 14.0). All were registered members of Blind Low Vision NZ, with an affiliation from 4 to 66 years, and an average membership of 29.1 years (SD = 15.9). These characteristics were in accordance with Blind Low Vision NZ's estimation of its client base at the time of the study. At the time of the study 27 of the 36 participants were currently using a

[1]Preliminary results of the present study were published in the proceedings of the 13th International Mobility Conference (18).

dog. Nine were not, and of these, seven had decided not to use a dog in the future due to not wanting a dog currently ($n = 3$) or at all ($n = 3$), or due to having a poor relationship with the guide dog school ($n = 1$). The other two were on the waiting list for a replacement dog. Nearly all participants (32, 88.9%) had used more than one dog, the average being 2.9. Consequently there were more handler-dog partnerships than there were participants in the study. Of the 36 participants; 17 had experienced a single partnership end (dog loss), 11 had experienced two losses, three had experienced three, one had experienced four, five and six losses, respectively, and two people had experienced the ending of seven partnerships. This makes a total of 77 ends of partnerships that were rated by the 36 participants.

Data Collection

All 36 participants had encountered the ending of at least one partnership with a guide dog. Participants were asked to rate their experiences in terms of (a) what became of the dog (the dog's fate), (b) their feelings when the partnership ended, (c) how this affected their application for a subsequent dog, and (d) how the end of the partnership influenced the relationship with subsequent dogs. Issues concerning how being without a guide dog, after having experienced using one, affected quality of life were also explored. Participants were asked to describe, in order of importance, how the absence of the dog affected their quality of life in general. All questions were open ended. Thus, the participants provided unique, unanticipated answers, which were recorded in written form and analyzed descriptively to show how often a response was given *via* measures of frequency, including count and percent.

RESULTS: THE END OF THE PARTNERSHIP

Fate of the Dog

Participants ($N = 36$) had experienced the ending of at least one and up to seven partnerships with dogs, giving a total of 77 "losses" (**Table 1**). Of these 77, 13 dogs (16.9%) were kept as pets by their handlers[2] and another 24 were retired to live with a friend (10, 13.0%), family member (2, 2.6%) or rehomed by the guide dog school (12, 15.6%). Twenty-three (29.9%) of the dogs were returned to the guide dog school; Of these, 19 were rematched with another handler, three were withdrawn from the guide dog program and rehomed by the guide dog school, and one had a successful "change of career" with a different national working dog program. The remaining 17 (22.1%) dogs died of old age, underwent euthanasia for health problems or had a fatal accident before reaching retirement. Following the end of the majority of the partnerships (57, 74.0%) the handlers either wanted to or did keep in touch with their dogs. It was notable that people were most likely to want to retain contact with their first dog.

[2]According to some participants in this study, before 1988 the guide dog school regulations did not always permit handlers to apply for a replacement dog if they kept a retired guide dog as a pet, nor was contact between the former handler and the person/family that adopted a retired guide dog allowed.

Application for a Replacement Dog

Overall, the end of a partnership did not put handlers off applying for another dog immediately in 62 (80.5%) of the 77 cases (**Table 1**). However, on six (7.8%) occasions people were put off indefinitely and nine (11.7%) chose to wait from a couple of months up to 5 years. Most of the people who had experienced a mismatch and who had applied for a replacement dog stated that they were optimistic about getting a better dog next time.

The people who wanted a replacement dog immediately had wanted guide dog assisted mobility as soon as possible and/or had kept their previous dogs as pets. For most this was the right decision, even if they had experienced an unsuitable dog, but some regretted not taking more time to come to terms with the loss of their previous dog before acquiring its replacement. On six occasions people indicated they would never get another dog and this was because they did not expect to get over the loss of their previous dog and/or they did not wish to repeat the painful experience of receiving an unsuitable dog. One who had initially felt this way declared that the guide dog school had "forced" another dog on her, which she was ultimately grateful for as it turned out to be a very good match. When people elected to wait some time before requesting another dog they did so as they needed time to grieve, wanted a break from the responsibility of owning a dog or had temporary changes in personal circumstances such as increased social support or an alternative means of travel.

Of the 36 participants, 26 predicted using another guide dog at some point, seven people did not, and three people said they would consider it. Of the 26 wanting to use another dog, eight did so solely because they preferred a dog to a long cane as a mobility aid, while 16 people, who also preferred the dog to the long cane, gave equal importance to the companionship and social interactions that the dog provided. Reasons given by the seven people not envisaging the use of a dog in the future included: changed mobility needs ($n = 2$); lack of trust in the guide dog school ($n = 2$); considered it not worth the effort of retraining with a new dog ($n = 1$); felt there was no difference in life with and without a guide dog ($n = 1$); or was in an unsuitable living environment ($n = 1$). The three participants who were undecided were enjoying a break from dog ownership ($n = 1$), unsure about future need ($n = 1$) or under pressure from family not to get another dog ($n = 1$).

Six participants (16.7%) were currently on the guide dog school's waiting list for a replacement dog; three currently had no dog, one had a temporary dog, another a poorly matched dog and the remaining person had a dog that was due to retire. The other 30 participants not currently on the waiting list comprised 19 people who were happily using their current dogs, three who were debating returning their current dogs (mainly for canine reasons of distraction, aggression and poor health), and eight who had previously used a dog and decided not to use another, or were having a break, as described earlier.

Feelings at the End of the Partnership

Amongst the 36 participants, the end of 77 guide dog/handler partnerships (losses) had been experienced. **Table 2** presents the responses to the question of how people felt at the end of a

TABLE 1 | Issues regarding the end of the handler-dog partnership ($n = 77$) for the handlers' first (1st dog) and subsequent guide dogs (up to the 7th dog used).

Issues	1st dog ($n = 36$)	2nd dog ($n = 19$)	3rd dog ($n = 8$)	4th dog ($n = 5$)	5th dog ($n = 4$)	6th dog ($n = 3$)	7th dog ($n = 2$)
Dogs' fate *n* (%)							
Retired-family	2 (5.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Retired-friend	7 (19.4)	1 (5.3)	1 (12.5)	0 (0.0)	0 (0.0)	1 (33.3)	0 (0.0)
Retired-RNZFB home	7 (19.4)	3 (15.8)	1 (12.5)	1 (20.0)	0 (0.0)	0 (0.0)	0 (0.0)
Returned-rematched*	2 (5.6)	7 (36.8)	2 (25.0)	2 (40.0)	3 (75.0)	2 (66.7)	1 (50.0)
Kept as pet by handler	6 (16.7)	3 (15.8)	2 (25.0)	1 (20.0)	1 (25.0)	0 (0.0)	0 (0.0)
Returned-RNZFB's home	2 (5.6)	0 (0.0)	1 (12.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Returned-new career	0 (0.0)	1 (5.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Deceased/euthanasia	10 (27.8)	4 (21.1)	1 (12.5)	1 (20.0)	0 (0.0)	0 (0.0)	1 (50.0)
Kept in touch-if possible (%)	32 (88.9)	12 (63.2)	6 (75.0)	3 (60.0)	2 (50.0)	1 (33.3)	1 (50.0)
Time "put off" applying for a replacement dog (%)							
No time	28 (77.8)	15 (78.9)	7 (87.5)	5 (100.0)	4 (100.0)	2 (66.6)	1 (50.0)
Up to 3 months	3 (8.3)	1 (5.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
3–6 months	1 (2.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
6 months−1 year	0 (0.0)	1 (5.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (50.0)
2–5 years	2 (5.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Indefinitely	2 (5.6)	2 (10.5)	1 (12.5)	0 (0.0)	0 (0.0)	1 (33.3)	0 (0.0)

No missing responses.

**Rematched dogs are dogs that are returned to the guide dog school and matched to another handler.*

TABLE 2 | Feelings expressed about the end of the handler-dog partnership according to number of losses experienced (1–7).

Feelings* *n* (%)	1st dog ($n = 36$)	2nd dog ($n = 19$)	3rd dog ($n = 8$)	4th dog ($n = 5$)	5th dog ($n = 4$)	6th dog ($n = 3$)	7th dog ($n = 2$)	Total ($n = 77$)
Grief-extreme	22 (61.1)	8 (42.1)	3 (37.5)	1 (20.0)	1 (25.0)	1 (33.3)	2 (100.0)	38 (49.4)
Grief-somewhat	3 (8.3)	1 (5.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	4 (5.2)
Neutral	1 (2.8)	1 (5.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (2.6)
Relieved	1 (2.8)	4 (21.1)	1 (12.5)	3 (60.0)	2 (50.0)	2 (66.6)	0 (0.0)	13 (16.9)
Angry with the guide dog school	8 (22.2)	6 (31.6)	1 (12.5)	1 (20.0)	2 (50.0)	1 (33.3)	0 (0.0)	19 (24.7)
Shocked to "fail"	3 (8.3)	4 (21.1)	0 (0.0)	2 (40.0)	1 (25.0)	1 (33.3)	0 (0.0)	11 (14.3)
Guilty	11 (30.6)	2 (10.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (50.0)	14 (18.2)
Accepting	23 (63.9)	12 (63.2)	7 (87.5)	2 (40.0)	2 (50.0)	0 (0.0)	0 (0.0)	46 (59.7)
Reassured re. Pet home	16 (44.4)	6 (31.6)	7 (87.5)	2 (40.0)	2 (50.0)	0 (0.0)	0 (0.0)	33 (42.9)
Worry re. mobility	3 (8.3)	2 (10.5)	1 (12.5)	0 (0.0)	1 (25.0)	0 (0.0)	1 (50.0)	8 (10.4)
Resentful/denial	2 (5.6)	1 (5.3)	2 (25.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	5 (6.5)
No self-blame	1 (2.8)	2 (10.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (3.9)
Family devastated	6 (16.7)	3 (15.8)	1 (12.5)	1 (20.0)	0 (0.0)	1 (33.3)	1 (50.0)	13 (16.9)
Hoping for better next time	2 (5.6)	0 (0.0)	0 (0.0)	1 (20.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (3.9)

**Total percent adds to more than 100 due to open-ended questions/multiple responses.*

partnership with a guide dog with responses to the loss of the first (column 2) and subsequent dogs (columns 3-8) provided separately. The combined (total) responses are presented in column 9. From the total column it can be seen that the main feelings experienced at the end of partnerships were acceptance, extreme grief and feeling reassured about the dog's future home. There was a high degree of acceptance expressed, because people had enjoyed a successful relationship, and/or understood the rationale for an early ending to the partnership and/or needed another dog for mobility. Extreme grief was expressed at the end of around half of the relationships and was likened to losing a family member or other loved one. Grieving was not limited to the handlers; many of those expressing profound grief said that their family members and some work colleagues who had spent a good deal of time with the dog also suffered a great loss. Feelings of extreme grief were expressed at the end of 38 partnerships and the fate of these 38 dogs was: rehoming ($n = 15$); death during working life ($n = 15$); being rematched to another handler by the guide dog school ($n = 5$); and remaining with their handlers as pets ($n = 3$). While feelings of extreme grief were very common, experiencing a lesser degree of grief or feeling neutral was rare. Relief was expressed at the end of several partnerships (16.9%)—all of which were considered to be poor matches. It is notable that feelings of extreme grief were expressed most often after the loss of the first dog and feelings of relief that the partnership had ended were lower for first dogs compared to subsequent dogs combined.

Acknowledging that the working relationship was over was easier for the majority of participants if they knew the dog was going to a good home. This was especially true if the dog was being replaced because of work, or was being kept as a pet. A few people found the situation very hard to accept, and resented and in some cases denied that the dog was getting too old to work. Loss of mobility was a concern at the end of eight partnerships as people worried about losing their freedom and independence.

Anger directed toward the guide dog school was experienced at the end of 19 partnerships because people felt that they had not been supplied with a suitable dog in the first place and/or were not fully informed that a dog had a problematic history. Other reasons for ire at this time came from people who felt abandoned when their partnerships ended because of the dogs' ill health, due to a perceived lack of emotional support from the guide dog school. Guilt was experienced at the end of 14 partnerships, 11 of which were first dogs, not only due to having to give up the "old" for the "new," but in some cases where dogs had died people felt guilty that the dog had not been able to enjoy retirement and when a partnership had failed, some people felt guilty that it may have been their (or their family's) fault. "Failure" had not been an option considered by the 11 people who were shocked when their relationship did not work out necessitating in the return of the dog.

Relationships With Subsequent Dogs

Not everyone continued on to use another guide dog when a partnership was over so responses to the question of how a relationship with a previous dog had influenced the next was relevant following the end of 71 partnerships. Of these, 19 (26.8%) were reported to have had a negative effect in that the old dog was considered to be a better mobility aid and/or less puppy-like in general, and/or that the memory of the old dog inhibited bonding with the new. The latter was reason enough to put a few handlers off acquiring a replacement dog indefinitely. There were 13 (18.3%) examples of a positive effect through the handler knowing what to expect through experience and realizing that the new dog was an improvement over the previous one. No comparison was made following the end of 12 (16.9%) partnerships as the dogs were appreciated as individuals despite the associated feelings of loss. However, the largest number 27 (38.0%) were said to have had no effect on the subsequent relationship. Breed-specific behaviors also played a role, as exemplified by two people saying they did not want to repeat the experience of having a Labrador retriever due to the scavenging behaviors exhibited.

TABLE 3 | The participants' ($N = 36$) comments on how being without a guide dog after experiencing guide dog assisted mobility affected quality of life.

Effect on quality of life	Comments	*N* (%)
Reduced	Loss of independent mobility*	31 (86.1)
	Loss of friend/companion**	21 (58.3)
	Curtailing of social function/interactions***	10 (27.7)
	Loss of self-esteem/confidence	6 (16.7)
	Concern regarding when dog will be replaced	5 (13.9)
	Feelings of failure re mismatch/guilt at giving up dog	3 (8.3)
Neutral	Handler's decision—change in circumstances/mobility needs	4 (11.1)
	No effect—good cane traveler and not very attached to dog	3 (8.3)
Increased	Appreciated a break from responsibilities of dog ownership	3 (8.3)
	Return of unsuitable dog liberating	2 (5.6)
	Long cane skills improved	1 (2.8)

Total percent does not add to 100 due to open-ended questions/multiple responses.
**Most important effect; **Second most important effect; ***Third most important effect.*

Post-guide Dog Assisted Mobility: Effects on Quality of Life

Responses to the question of how quality of life was affected by being without a guide dog after experiencing guide dog mobility are presented in **Table 3**. Most people provided more than one response so the number of comments (89) was greater than the number of participants ($N = 36$) and percentages add to more than 100. Seventy-six of the 89 comments (85.4%) indicated that quality of life was reduced when participants were without a dog after experiencing guide dog mobility. This outcome was mainly due to a reduction in independent mobility for 31 (86.1%) participants, followed by losing a friend and companion for 21 (58.3%). Other reductions in quality of life included the effect on social interactions for 10 (27.7%) and loss of self-esteem and/or confidence for six (16.7%). Quality of life had increased for the three (8.3%) who enjoyed a break from the responsibilities of looking after a dog, the two (5.6%) who appreciated not having to deal with an unsuitable dog and the one (2.8%) whose long cane skills improved through the opportunity to practice. The remaining seven comments indicated that quality of life did not change with four people (11.1%) explaining that their requirement for a dog had altered as a result of a change in circumstances re their mobility needs, and three (8.3%) said that they were good cane travelers and had not felt particularly emotionally attached to the dog. In responding to this question participants were asked to describe the most important effect first and, as shown by the asterisks in **Table 3**, the results mirror the overall findings. Not only was loss of individual mobility the most commonly cited effect of no longer having a canine guide, it was also considered to be the most important effect on quality of life. This was followed by loss of friend or companion being the second most important effect and the curtailment of social interactions the third most important and frequent.

DISCUSSION

One way of understanding humans' relationships with companion animals is through attachment theory i.e., the concept that we become emotionally attached to our companion animals in a similar way as we do to people. The human-animal bond has existed for thousands of years. The American Veterinary Medical Association (AVMA) defines this bond as "a mutually beneficial and dynamic relationship between people and animals that is influenced by behaviors essential to the health and well-being of both. This includes, among other things, emotional, psychological, and physical interactions of people, animals, and the environment." (42). Dogs (and other animals) have been helping people with physical disabilities and providing emotional support for centuries (25, 29). The results of the present study shows that the existence of this bond has major significance for the well-being of both the handler and the dog as it influences how people feel about getting a subsequent guide dog and the relationship formed between the dyad, as well as directly impacting on the handlers' well-being at the end of the partnership.

Distress at the End of the Partnership

Almost all the participants had used more than one dog. As the transition from one dog to the next is a recurring feature, guide dog handlers probably experience the end of more relationships than the average pet dog owner (22). Retiring a guide dog is not only difficult for the handler, but also for the handler's family and friends, and doubtlessly the dog. Most people expressed feelings of grief when a working partnership ended, whether it was successful or not. Extreme grief was the feeling most frequently reported and this was shared by family members, friends and co-workers. The depth of emotion was compared to losing a family member or other loved one; a comparison that was also reported by Fogle (30) and Stewart (31) at the end of some person and pet relationships, and more recently by Uccheddu et al. (24). The distress caused by the end of the partnership between a handler and a guide dog might be more intense than that experienced between a person and a pet (8) due to the interdependent nature of the relationship (7, 10), the time spent together and because the dog helps the handler to do things that could not be accomplished alone. However, the grieving process over the loss of any companion animal may be hampered by the lack of validation for the mourning of non-human animals by the general public and some professionals, as well as the lack of socially sanctioned grief rituals that typically accompany the loss of a human (32). Many guide dog schools around the world recognize this significance and have memorial gardens for people to visit, remember and honor their guide dogs who have passed.

A study by Barnard-Nguyen et al. (33) that measured people's responses to the loss of a pet (dog or cat), rather than a human, found that people had different types of grief (sorrow, anger, and guilt). While a reaction of sorrow on the loss of the pet was considered to be part of a "normal" psychological process, some people developed "complicated" grief manifesting as depression and other mental health problems. From the attachment theory perspective, it would be expected that people with a stronger attachment to their guide dog or pet would feel more grief when the pet dies, which was the case in both the present study and Barnard-Nguyen et al. (33) study. Barnard-Nguyen et al. (33) found people who were more emotionally attached to their pet reported more grief and sorrow, and also more feelings of anger (e.g., toward the veterinarian for not being able to save the pet), but not guilt. This contrasts somewhat with the end of the working relationship with a guide dog as described in the next paragraph.

The participants that described the end of the partnership with their guide dog as a relief were commenting on dogs that they felt they had been mismatched with, suggesting a strong emotional bond had not been formed. This finding was also discussed in a focus group prior to the present study (34). However, some participants in the present study reported anger at being mismatched, shock in having "failed," guilt in case it had been their fault, and a few lost self-esteem and confidence. This was true even if the team had not bonded. These findings parallel those of Nicholson et al. (22) who found that the end of the partnership was an upsetting experience, even if there had been problems in the relationship. The exception to this was mismatches that ended after a relatively short period with no real bonding.

A related study by Lloyd et al. (13) and Lloyd et al. (14) on the complexities of successful and unsuccessful guide dog partnerships found that most dogs that were considered to be mismatched were returned to the guide dog school by the handler after just 3 months. An earlier study by the same authors (34) reported that the bond between a handler and their new guide dog could take 6 months or longer to develop. Therefore, guide dog instructors should inform handlers who may be frustrated with a new partnership that the relationship might take from 6 months up to a year to improve. Lloyd et al. (13) and Lloyd et al. (14) also found that some handlers who returned their problem dog did not feel that a mis-match had occurred because the guide dog instructor discussed potential issues at the outset, thus enabling the handlers to make informed choices. Hence, the opportunity for person and dog to work through problems together may actually strengthen the bond.

The grieving process was easier if handlers were reassured about the dogs' destiny, an outcome also reported by Nicholson et al. (22). This held true whether dogs were being kept as pets or were going to approved retirement homes. As Sanders (6) illustrated, the latter option was seen as a better alternative for dogs in that they were not expected to cope with the presence and/or the role of a new guide dog. It is noteworthy that almost all who experienced the death of a dog in the present study while it was still working experienced feelings of extreme grief at the end of the partnership. This may not only be due to attachment dynamics but may also be due to the loss of the equally important caregiving relationship as described by Kwong and Bartholomew (23). According to Bowlby (35), the care giving system is designed to provide protection and support to a dependent—goals that are no longer viable on the death of a dog and which may lead to feelings of intense despair. Identifying those who may be at greater risk for problematic grief reactions is of considerable value to guide dog schools. While guide dog schools should be prepared to support all clients in their grief responses, recognizing that someone is highly attached to their dog or that the dog died during its working life should trigger additional support.

Applying for a Replacement Dog

As for those who mourn the end of a relationship with humans, feelings of grief for the end of the relationship with a service dog or pet dog should be acknowledged and respected. The need for a period of adjustment before committing to a relationship with a new service dog or pet vs. prompt replacement is controversial. A study by Jarolmen (36) compared grief and bereavement responses of children, adolescents and adults who had lost a pet within a 12-month interval. Her findings indicated that children and adolescents are similarly attached to their pets and that children grieved more than adults did. Jarolmen (36) also concluded that the more recent the loss the more intense the response, but if the loss is anticipated grief is allayed. In her doctoral thesis abstract, Jarolmen (37) makes the interesting statement that those who have another pet in the home at the time of loss grieve the same for the lost pet as those who do not have another pet in the home, but those who replace the pet have a higher grief response. Unfortunately, it was not possible to obtain a copy of the entire thesis to read more about this relationship. Stewart's (31) findings suggested that deferring replacement was not warranted, even when a highly significant pet died, provided the pet's death was not trivialized and the new pet was introduced in a sensitive manner.

The majority of handlers in the present study said that they would continue to use guide dogs in the future, as they preferred and/or had become dependent on guide dog assisted mobility. Those who did not anticipate using another dog had a limited workload, unsuitable living conditions, family pressure, or did not have a trusting relationship with the guide dog school. Others, who were undecided, enjoyed not having the responsibility of ownership, were concerned that they might experience another mismatch or were impartial concerning cane or dog. Although the grief response was high, this did not preclude most handlers from applying for another dog right away, including those who had experienced a mismatch. The decision about when to replace a dog is personal, but some handlers regretted replacing their dog before adjusting to the loss of the previous one. Regardless, it appears that the need for mobility is the force behind the desire to replace a dog quickly.

Many handlers felt that the end of the partnership with the previous dog had a negative effect on the relationship with the new one. This was due to the old dog being considered a better mobility aid, less puppy-like and the memory of the previous dog inhibiting bonding. The role of the guide dog instructor re the human-animal bond is vital at this time to maximize the potentials of the relationship between their clients and their new dogs. Practical implications for the guide dog industry include that feelings of grief at the end of a partnership should not be trivialized and the new dog should be introduced in a sympathetic manner. Guide dog schools that offer grief-counseling sessions to enable handlers to share their feelings over the loss of the previous dog, have found that these handlers form a healthy relationship and train more quickly with the replacement dog (J. Campbell, Leader Dogs for the Blind, Michigan, USA, personal communication, August, 2000). Thus, the grief response and its expression may in fact be necessary in order for the next bond to form.

Quality of Life

Lloyd (17) showed that using a guide dog improved the quality of life for handlers due to enhanced mobility, social interactions, fitness, mental and physical health, and adjustment to loss of vision. These findings were echoed by a recent longitudinal study by McIver et al. (38) who demonstrated that guide dog owners perceived their quality of life to increase over time in similar terms. However, as shown by the present study, quality of life can also decrease at the end of some partnerships. Lessening of quality of life in the present study was due to a reduction in independent mobility followed by the loss of a friend and companion and curtailment of social interactions. People experience a reduction in the quality of their independent mobility for a variety of reasons. For example, Lloyd et al. (2) found that it was troublesome for those who were accustomed to guide dog assisted mobility to be without a dog because their cane skills had deteriorated. The present study also reveals that confidence and self-esteem were reduced when an unsuitable dog was received, or when a dog was retired or died. These effects were also found in children (39) and adults (32, 40) who mourned the loss of a pet, and are described for guide dog owners in Schneider (21) and Gosling (20).

It would behoove guide dog schools to be empathic to the emotional and practical challenges their clients face at this time. Many guide dog schools provide access to councilors or information sources that may be helpful to people dealing with the loss of a guide dog such as helplines, books, on-line resources etc. Since this research was conducted Blind Low Vision NZ set in motion a client-driven national system for grief management, including sharing memories/experiences, for members who have lost or retired a guide dog, and Ward and Pierce (19) created a client driven information resource for second time guide dog applicants to help them transition to a new dog.

Practical implications for the guide dog industry can also be found in a practice report written by an experienced guide dog handler (21). Schneider suggests: being honest, letting the person know that one is open to hearing about one's grief, and reminding others that all dogs were once new and young. Allen (16) has written a heart wrenching account entitled "Letting go of the harness for the last time" that provides advice for guide dog schools and for veterinarians who care for guide dogs (and thus the person-dog team). Allen (16) study illustrates how peoples' experiences can influence broader social issues including policies for agencies that provide guide dogs, and the role of the veterinarian when it is time to make decisions about retiring or euthanizing the dog.

First vs. Subsequent Dogs

"Second dog syndrome" (SDS) is a phenomenon seen whereby a person's second guide dog is significantly less favored than the first dog (17) and is apparent on a number of levels in the present study. Lloyd (17) found that handlers' expectations regarding mobility were met less often and handlers were less compatible with their second dogs than their first dogs, with little apparent difference between the first and the third dog. Lloyd et al. (13) Lloyd et al. (14) showed that the number of dogs that were deemed to be mismatched and/or returned to the guide dog school were highest for second dogs compared to first or third. Similar trends were apparent in the present study regarding the handlers' feelings at the end of the partnership with their first dogs compared to subsequent dogs, where people grieved more; reported more feelings of guilt and less of relief when the partnership ended; and more desire to keep in touch with their first dogs. As explained in Lloyd et al. (13) and Lloyd et al. (14), it is understandable why there should be a "first-dog" effect in the handlers' affections; this dog initialized/improved independent mobility, thus being the catalyst for life changing events. Another explanation proposed by Lloyd et al. (18) is "distortion of memory" where handlers may have forgotten that their first dog was as boisterous and exuberant in its youth as the new dog now was, and that it took time for the dog (and the partnership) to mature. In addition, handlers may forget or deny that the previous dog made the same mistakes as the inexperienced new one, or harbor the unrealistic expectation that the new dog can take over precisely where the old dog left off (41). This is exemplified in a touching and elegantly written account by a graduate of Leader Dogs for the Blind on her experiences concerning the loss of several guide dogs:

> Jack was my second dog. As I was training with him, we waited for the [traffic] light to change… I stepped from the curb with confidence. Jack executed a perfect (sic) diagonal crossing… [(41), pp.10-11].

Although this experience left the handler "dazed and frightened at the (wrong) corner," this "green" dog eventually grew into a conscientious worker and Smith (41) emphasizes the importance of patience and understanding within any new relationship. Allen [(16), p.9] suggests that SDS can happen with any dog as "left overs from a previous dog can get in the way of accepting a new dog." Knowing that a second, or any subsequent, dog is likely to be considered second best is valuable knowledge for guide dog instructors to help prepare clients to receive their next dog. Knowing that SDS is a tangible occurrence may alleviate some negative feelings the replacement dog engenders simply by not being the other dog.

CONCLUSIONS

This study has contributed to the small body of literature concerning how the end of the partnership with a guide dog affects the handlers' quality of life and their relationships with subsequent dogs. Results support the negative view of life after the loss of a guide dog (16, 22) or other assistance dog (24), and mirrors many aspects of grief responses in pet dog ownership (28, 30, 31, 33, 36)—issues that remain underestimated. Although the grief response was high, this did not preclude most handlers from applying for a replacement dog quickly as the need for mobility was high. Future research could attempt to further tease apart any differences in a guide dog handler's reactions to the loss of their subsequent dogs—an exercise that was not feasible with the small number of successive dogs used in this study. It would also be beneficial to look at the experiences of handlers at various stages of working with dogs vs. having a break from guide dog use. Concerning how individuals handle the end of the partnership with their guide (or other) dogs: in the words of Gosling [(20), p. 12] "there is no right or wrong; it just is." However, findings from the present study will help inform the guide dog industry regarding how best to support their clients during this time and when transitioning to another dog. Understanding the importance of the human-animal bond and implementing strategies to help clients grieve to support quality of life when experiencing the loss of a guide dog may be applied to other assistance/service dog users and the pet owning community.

DATA AVAILABILITY STATEMENT

The datasets for this article are not publicly or otherwise available due to issues of confidentiality. However, Supplementary Material can be found online at https://www.frontiersin.org/article/10.3389/fvets.2016.00114/full#supplementary-material.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Massey University Human Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JL undertook the research and wrote the article with CB with the approval of the other authors who have critically revised the content. All authors on this publication have contributed to the conception and the design of this work, and agree to be accountable for the content.

ACKNOWLEDGMENTS

The authors wish to thank the people who participated in this study, and the staff of Blind Low Vision NZ (formerly Blind Foundation/Royal New Zealand Foundation of the Blind) Guide Dogs for their support throughout the process. We would like to note the contribution of the doctoral dissertation Exploring the match between people and their guide dogs (17).

REFERENCES

1. Lloyd JKF, La Grow S, Stafford KJ, Budge RC. The guide dog as a mobility aid part 1: perceived effectiveness on travel performance. *Int J Orient Mobil*. (2008) 1:17–33. doi: 10.21307/ijom-2008-003
2. Lloyd JKF, La Grow S, Stafford KJ, Budge RC. The guide dog as a mobility aid part 2: perceived changes to travel habits. *Int J Orient Mobil*. (2008) 1:34–45. doi: 10.21307/ijom-2008-004
3. Muldoon C. Does the presence of a guide dog enhance feelings of social acceptance in guide dog users? [CD-ROM]. In: *Proceedings of the 10th International Mobility Conference, Coventry, England*. (2000). P. 258–61.
4. Refson K, Jackson AJ, Dusoir AE, Archer DB. Ophthalmic and visual profile of guide dog owners in Scotland. *Br J Ophthalmol*. (1998) 83:470–7. doi: 10.1136/bjo.83.4.470
5. Refson K, Jackson AJ, Dusoir AE, Archer DB. The health and social status of guide dog owners and other visually impaired adults in Scotland. *Visual Impair Res*. (1999) 1:95–109. doi: 10.1076/vimr.1.2.95.4411
6. Sanders CR. *Understanding Dogs: Living and Working with Canine Companions*. Philadelphia: Temple University Press (1999).
7. Sanders CR. The impact of guide dogs on the identity of people with visual impairments. *Anthrozo?s*. (2000) 13:131–9. doi: 10.2752/089279300786999815
8. Steffans MC, Bergler R. Blind people and their dogs. In: Wilson CC, Turner DC, editors. *Companion Animals in Human Health*. Thousand Oaks, CA: Sage (1998). p. 149–57.
9. Whitmarsh LE. The benefits of guide dog ownership. *Visual Impair Res*. (2005) 7:27–42. doi: 10.1080/13882350590956439
10. Zee A. Guide dogs and their owners: assistance and friendship. In: Katcher AH, Beck AM, editors. *New Perspectives on Our Lives with Companion Animals*. Philadelphia: University of Pennsylvania Press (1983). p. 472–83.
11. York S, Whiteside H. Exploring social (non-working) behaviour in guide dogs from the perspective of guide dog owners. *AnthrozoÖs*. (2018) 31:525–36. doi: 10.1080/08927936.2018.1505259
12. Lloyd JKF, La Grow SJ, Budge RC, Stafford KJ. Matching guide dogs: mobility and compatibility outcomes. In: *Proceedings of the 11th International Mobility Conference, March 30-April 4, 2003*. Stellenbosch (2003).
13. Lloyd J, Budge C, La Grow S, Stafford K. An investigation of the complexities of successful and unsuccessful guide dog matching and partnerships. *Front Vet Sci*. (2016) 3:114. doi: 10.3389/fvets.2016.00114
14. Lloyd J, Budge C, La Grow S, Stafford K. An investigation of the complexities of successful and unsuccessful guide dog matching and partnerships. In: Otto CM, Wilsson E, Cobb M, editors. *Working Dogs: Form and Function*, 2nd edn. Lausanne: Frontiers Media SA (2020). p. 86–100. doi: 10.3389/978-2-88963-636-5
15. Koda N, Kubo M, Ishigami T, Furuhashi H. Assessment of dog guides by users in Japan and suggestions for improvement. *J Visual Impair Blindness*. (2011) 105:591–600. doi: 10.1177/0145482X1110501006
16. Allen DM. Letting go of the harness for the last time: A descriptive realism approach to exploring the ending of working relationships with guide dogs. *McNair Scholars Res J*. (2006) 2:7–13. Available online at: https://scholarworks.boisestate.edu/mcnair_journal/vol2/iss1/5
17. Lloyd JKF. *Exploring the Match Between People and their Guide Dogs* (Doctoral dissertation). Massey University, Palmerston North, New Zealand (2004).
18. Lloyd JKF, Budge RC, Stafford KJ, La Grow S. The end of the partnership with a guide dog: Emotional response and relationships with subsequent dogs. [CD-ROM]. In: *Proceedings of the 13th International Mobility Conference*. Marburg, 5 pages – unnumbered (2009).
19. Ward K, Peirce K. A client driven information resource for second time guide dog applicants. *Int J Orient Mobil*. (2010) 3:36–40. doi: 10.21307/ijom-2010-005
20. Gosling J. Guide dog handler bereavement. *Visionary*. (2009) 14:12–13. Available online at: https://www.igdf.org.uk/media/30685/visionary_july_2009_proof_v7_1_.pdf
21. Schneider KS. The winding valley of grief: when a dog guide retires or dies. *J Visual Impair Blindness*. (2005) 99:368–70. doi: 10.1177/0145482X0509900606
22. Nicholson J, Kemp-Wheeler S, Griffiths D. Distress arising from the end of a guide dog partnership. *AnthrozoÖs*. (1995) 8:100–10. doi: 10.2752/089279395787156419
23. Kwong MJ, Bartholomew K. "Not just a dog": an attachment perspective on relationships with assistance dogs. *Attach Hum Dev*. (2011) 5:421–36. doi: 10.1080/14616734.2011.584410
24. Uccheddu S, De Cataldo L, Albertini M, Coren S, Da Graça Pereira G, Haverbeke A, et al. Pet humanisation and related grief: development and validation of a structured questionnaire instrument to evaluate grief in people who have lost a companion dog. *Animals*. (2019) 9:933. doi: 10.3390/ani9110933
25. Lloyd J, Johnston L, Lewis J. Psychiatric assistance dog use for people living with mental health disorders. *Front Vet Sci*. (2019) 6:1–6. doi: 10.3389/fvets.2019.00166
26. Walther S, Yamamoto M, Thigpen AP, Garcia A, Willits NH, Hart LA. Assistance dogs: historic patterns and roles of dogs placed by ADI or IGDF accredited facilities and by non-accredited U.S. facilities. *Front Vet Sci*. (2017) 4:e00001. doi: 10.3389/fvets.2017.00001
27. Zapf SA, Rough RB. The development of an instrument to match individuals with disabilities and service animals. *Disab Rehab*. (2002) 24:47–58. doi: 10.1080/09638280110066316
28. Yamamoto M, Hart LA. Separation from assistance dogs: the complicated psychological burden during loss of the relationship. In: Kogan L, Erdman P, editors. *Pet loss, Grief and Therapeutic Interventions: Practitioners Navigating the Human-Animal Bond (Ch. 12, 21 pages)*. Routledge. Available online at: https://www.taylorfrancis.com/books/e/9780429505201 (accessed March 9, 2020).
29. Bustad LK, Hines LM. Historical perspectives of the human-animalbond. In: Anderson RK, Hart BL, Hart LA, editors. *The Pet Connection: Its Influence on Our Health and Quality of Life*. Minneapolis, MN: Center to Study Human-Animal Relationships and Environments (1984). p. 15–29.
30. Fogle B. Attachment-euthanasia-grieving. In: Fogle B, editor. *Interrelations Between People and Pets*. Springfield, IL: Charles C. Thomas (1981). p. 331–43.
31. Stewart M. Loss of a pet-loss of a person. In: Katcher AH, Beck AM, editors. *New Perspectives on Our Lives with Companion Animals*. (1983). Philadelphia: University of Pennsylvania Press. p. 390403.
32. Sharkin BS, Knox D. Pet loss: issues and implications for the psychologist. *Prof Psychol Res Pract*. (2003) 32:414–21. doi: 10.1037/0735-7028.34.4.414
33. Barnard-Nguyen S, Breit M, Anderson K, Nielsen J. Pet loss and grief: identifying at-risk pet owners during the euthanasia process. *Anthrozoös*. (2016) 29:421–30. doi: 10.1080/08927936.2016.1181362
34. Lloyd JKF, Budge RC, Stafford KJ, La Grow S. A focus group discussion on using guide dogs. *Int J Orient Mobil*. (2009) 2:52–64. doi: 10.21307/ijom-2009-005
35. Bowlby J. *Attachment (Attachment and loss, Vol. 1)*. New York: Basic Books (1969).
36. Jarolmen J. A comparison of the grief reaction of children and adults: Focusing on pet loss and bereavement. *Omega J Death Dying*. (1998) 37:133–50. doi: 10.2190/H937-U230-X7D9-CVKH
37. Jarolmen JJ. *A Comparison of the Human Grief Reaction: Focusing on Pet Loss and Bereavement (Order No. AAM9618530)*. (1996). APA PsycInfo®. (6189983541996-95016-100). Available online at: https://elibrary.jcu.edu.au/login?url=https://www.proquest.com/docview/618998354?accountid=16285 (accessed December 1, 2020).
38. McIver S, Hall S, Mills DS. The impact of owning a guide dog on owners' quality of life: a longitudinal study. *AnthrozoÖs*. (2020) 33:103–17. doi: 10.1080/08927936.2020.1694315
39. Bryant BK. The richness of the child-pet relationship: a consideration of both benefits and costs of pets to children. *AnthrozoÖs*. (1990) 3:253–61. doi: 10.2752/089279390787057469

40. Weisman AD. Bereavement and companion animals. *OMEGA J Death Dying*. (1991) 22:241–8. doi: 10.2190/C54Y-UGMH-QGR4-CWTL
41. Smith MM. *Allow Me to Grieve, Help Me to Heal*. Rochester, MI: Leader Dogs for the Blind (1991).
42. AVMA (American Veterinary Medical Association). *Human-Animal Bond*. (n.d). Available online at: https://www.avma.org/one-health/human-animal-bond Available online (accessed March 1, 2021).

Conflict of Interest: The authors declare that this study received funding from Douglas Pharmaceuticals Ltd. and the Palmerston North Medical Research Foundation. The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 07 June 2021
doi: 10.3389/fvets.2021.545382

Development of a Performance Monitoring Instrument for Rating Explosives Search Dog Performance

Nicola J. Rooney[1]* and *Corinna C. A. Clark*[1,2]

[1] Animal Welfare and Behavior Group, Bristol Veterinary School University of Bristol, Langford, United Kingdom, [2] Warwick Medical School, University of Warwick, Coventry, United Kingdom

The growing body of working dog literature includes many examples of scales robustly developed to measure aspects of dog behavior. However, when comparing behavior to working dog ability, most studies rely on training organizations' own long-established ratings of performance, or simply pass/fail at selection or certification as measures of success. Working ability is multifaceted, and it is likely that different aspects of ability are differentially affected by external factors. In order to understand how specific aspects of selection, training, and operations influence a dog's working ability, numerous facets of performance should be considered. An accurate and validated method for quantifying multiple aspects of performance is therefore required. Here, we describe the first stages of formulating a meaningful performance measurement tool for two types of working search dogs. The systematic methodology used was: (1) interviews and workshops with a representative cross-section of stakeholders to produce a shortlist of behaviors integral to current operational performance of vehicle (VS) and high assurance (HAS) search dogs; (2) assessing the reliability and construct validity of the shortlisted behavioral measures (at the behavior and the individual rater level) using ratings of diverse videoed searches by experienced personnel; and (3) selecting the most essential and meaningful behaviors based on their reliability/validity and importance. The resulting performance measurement tool was composed of 12 shortlisted behaviors, most of which proved reliable and valid when assessed by a group of raters. At the individual rater level, however, there was variability between raters in the ability to use and interpret behavioral measures, in particular, more abstract behaviors such as Independence. This illustrates the importance of examining individual rater scores rather than extrapolating from group consensus (as is often done), especially when designing a tool that will ultimately be used by single raters. For ratings to be practically valuable, individual rater reliability needs to be improved, especially for behaviors deemed as essential (e.g., control and confidence). We suggest that the next steps are to investigate why individuals vary in their ratings and to undertake efforts to increase the likelihood that they reach a common conceptualization of each behavioral construct. Plausible approaches are improving the format in which behaviors are presented, e.g., by adding benchmarks and utilizing rater training.

Keywords: working dog, performance, scales, rating, validation, construct validity, reliability, individual

Edited by:
Lynette Arnason Hart, University of California, Davis, United States

Reviewed by:
Jane Ellen Russenberger, Guiding Eyes for the Blind, United States
Cynthia M. Otto, University of Pennsylvania, United States

***Correspondence:**
Nicola J. Rooney
nicola.rooney@bristol.ac.uk

Specialty section:
This article was submitted to Veterinary Humanities and Social Sciences, a section of the journal Frontiers in Veterinary Science

Received: *24 March 2020*
Accepted: *19 April 2021*
Published: *07 June 2021*

Citation:
Rooney NJ and Clark CCA (2021) Development of a Performance Monitoring Instrument for Rating Explosives Search Dog Performance. Front. Vet. Sci. 8:545382. doi: 10.3389/fvets.2021.545382

INTRODUCTION

Working Dog Research

Working dogs are used for a large number of roles: herding, assistance, protection, and detection of an increasing number of targets including people, cadavers, insects, money, drugs, explosives, human disease, and animal species of conservation importance [e.g., (1–4)]. To optimize capability, we need to fully understand the factors that may influence a dog's ability to perform in various settings. Hence, there has developed a wealth of scientific inquiry exploring ways to predict working ability and thereby improve the cost effectiveness and potentially ultimate working ability of a range of working dogs.

Temperament and selection tests have received considerable focus, with researchers, for example, exploring which behaviors in adult dogs and puppies best predict success as a detection dog [e.g., (5–8)], police dog (9), guide dog (10–12), or hunting dog (13). These studies have required researchers to measure the behavioral variability between individual animals, and there has been a large amount of effort developing adequate instrumentation that is both reliable and valid (14–16). Studies have also, for example, compared the use of subjective ratings *vs.* behavioral coding (8), finding them comparable in their predictive value and showing that novice raters produce comparable ratings to expert trainers (17).

Measuring Dog Performance

However, when seeking to predict working dog success, or explore factors that may influence that success, there has been less focus on validating measures of success or performance. Some studies have quantified the proportion of targets found in a single standardized search task (18–20), but most have relied on training organizations' own long-established ratings of performance [e.g., (13, 19, 21, 22)], or used pass/ fail at selection (6) or certification (23) as measures of success. These approaches ensure that the outcomes of the studies have great practical relevance and validity and enable individuals responsible for training working dogs to determine predictors of successful training. However, binary pass/fail outcomes do not facilitate the exploration of factors linked to excellent as compared to adequate performance, nor do they provide a means to examine differences in ability after the end of training. Organizations' own performance scales have generally not been developed with the same degree of scientific rigor as the behavioral scales described above, and there may be aspects of their design that have unforeseen consequences.

Evidence from the social science literature demonstrates that relatively small changes in scales (for example the number of points on the scale), or information (such as including anchor or benchmarks) and the training provided, can have significant effects on the way raters use scales and on their interpretation of the underlying constructs (24, 25). What's more, working ability is multifaceted, and it is likely that different aspects of ability are differentially affected by extraneous factors. Hence, in order to improve understanding of how specific aspects of selection, training, and operations impact a dog's working ability, we suggest that accurate and validated methods for quantifying multiple aspects of performance are needed. Having such measures would also allow dogs' performance to be measured on a regular basis, and hence organizations could determine factors affecting ongoing working ability and working dog longevity. Since handlers often work alone, any such measuring instrument needs to be easy to use and reliable when applied by a single handler.

Multiple Aspects of Performance

Scientists have started to consider the varying behavioral elements of detection work (26). Our own survey of experienced handlers suggested that search dog performance could be described by a number (>30) of independent attributes, including, for example, "obedience to human control," "stamina," and "boldness" (27, 28). This survey took a novel approach in that we asked respondents to rate both the ideal level and the importance of each trait, as we believe asking only about importance, as is often done [e.g., (29)], can lead to respondents underrating the importance of undesirable traits, which may in fact be very important to avoid.

We also saw that trainers were able to rate the performance of arms and explosives search (AES) dogs using a selection of scales, and their ratings for individual behaviors were independent of one another, reasonably reliable, and showed good agreement with objective ethological measures (20). These attributes could therefore form the basis of rating scales for quantifying dog performance on a search-by-search basis in the field. Here, we further develop rating scales utilizing the knowledge and experience of stakeholders (e.g., dog trainers and handlers) and a systematic approach.

Prioritizing Important Aspects

Our aim was to develop rating scales to be part of a performance measurement tool for monitoring the ability of working dogs in the British military, where collecting accurate and reliable data is paramount for informing both short-term training needs and long-term planning and policy changes.

Providing ratings for every possible aspect of search performance after every search would be practically unfeasible. Moreover, rater buy-in is vitally important in obtaining reliable ratings (30), and expecting handlers to rate too many behaviors would be unpopular as well as unfeasible. Therefore, the first stage of scale development was to identify the most important behavioral aspects to be measured. Many essential behaviors are common across multiple types of search work. However, the importance and desirable level of specific behaviors will likely differ with discipline (28), and it is unlikely that a single "one-size-fits-all" approach would be very effective. For example, "Friendliness to humans" is more important in drug search dogs, which, compared to explosives search dogs, have greater direct contact with the public (28). Conversely, the importance of "Obedience to human control" is greater in explosives search dogs because of the potentially dangerous situations in which they work. Here, we focus on two classifications of explosives search dogs being utilized by the British Army and Royal Air Force: vehicle search (VS) dogs, trained to search for explosives on/in vehicles (e.g., at the entrance to secure locations), and high assurance search (HAS) dogs, trained to assist counter-improvised explosive device (IED) teams searching routes and buildings for buried IEDs. Both VS and HAS dogs are trained using similar reward-based methods to find similar targets and to work under close handler control. Therefore, we expected there to be some similarities in the behavioral measures reflecting optimal performance, but also some subtle differences, illustrating the need for role-specific measurement scales. We used a series of interviews and workshops with a representative cross-section of stakeholders from each discipline using a procedure that does not limit or bias which behavioral measures are selected, based upon the method developed by Rooney et al. (28).

Reliability and Validity of Scales

We next examined reliability and construct validity using ratings of filmed searches in order to assess whether the raters were able to use the behavioral measures as accurate indicators of performance. Inter-rater reliability refers to how similar raters are to each other, and intra-rater reliability indicates how reliable individual raters are at repeatedly rating the same behaviors (31–33). High levels of inter-rater reliability would indicate that there is a common understanding of the behavior (or construct) between raters, and high intra-rater reliability indicates that a rater is consistent in applying the same principles to assessing behavior. Previous studies have shown that people are able to rate multiple aspects of dog behavior with high inter-rater reliability [e.g., (12)], but similar tests on performance measures are lacking. To test inter-rater reliability, we used the intraclass correlation coefficient (ICC). Average measure ICCs indicate how reliably, in general, a group of raters score each of the behaviors (34). Where ICCs are above an arbitrary threshold [>0.7 is commonly used and is therefore used throughout this study; see (35)], we could assume that raters were generally using a shared understanding of the behavior, as more of the variability in ratings exists between dogs than between raters. This is an important step in the initial stage of selecting reliable behaviors for a performance measurement tool (PMT). However, average measure ICCs cannot be generalized to indicate how well a single rater would perform (34), and as our final measurement tool [like many other commonly used scales, e.g., those used in rehoming centers; see (36)] will ultimately be used by single raters, it was also important to consider reliability estimates at the individual rater level. Single-measure ICCs were used to indicate how reliably a randomly selected rater might have scored each of the behaviors (37). Where raters are all equally reliable, the average and single-rater ICCs will be similar, but where there is variation among raters, with some more reliable than others, average measure ICCs will produce higher reliabilities.

Another goal of this study was to determine whether the behaviors measured separate constructs and whether the raters could efficiently distinguish between these. This is referred to as disciminant validity (DV). For example, Motivation (enthusiasm to search) and Stamina (ability to maintain enthusiasm) are theoretically separate constructs. A dog can show any level of stamina irrespective of its initial level of motivation (20), so we tested whether raters were able to score the two traits independently of each other. Conversely, convergent validity (CV) is a measure of how well-different scales measure the same behavior and can be assessed using correlation coefficients, with the expectation that items measuring the same aspect of performance will have high correlation coefficients.

Previously validated scales for working dog behavior have mainly focused on testing scenarios where there is little cost to measuring extra items and later applying using data reduction techniques [e.g., (9, 23)]. The goal of this sudy was to produce a measurement tool for use during day-to-day operational searches, so including only the most essential items while avoiding multiple items recording the same aspect of performance was essential. Hence, having low convergent validity and high DV was desirable. As there are no standard threshold coefficient values for DV or CV, we used the commonly used cutoff values, that is, a DV of <0.85, indicating that items do not overlap, and a CV of 0.7 or over, indicating a strong correlation between items (35).

We also wanted an indication of what proportion of individual raters were able to see the scales as independent (use high DV). As the measures of CV and DV describe the relationship between behavior measures across raters, providing a single coefficient may disguise important variations between raters. For example, it may be that, if some raters do not score behavior measures independently, this will not be evident in the average coefficient. We therefore counted the incidence of raters who showed high CV (above 0.7) and low DV (above 0.85 threshold).

Developing the Tool Based on the Analysis Results

Behaviors that cannot be reliably rated, or are viewed by raters as so closely linked to each other that they are indistinguishable, will be redundant in a streamlined performance rating tool. These should either be excluded or, where they are deemed essential but lack discrimination due to scale design inadequacies or rater error, should be further developed and reassessed. For example, if stakeholders agree that Motivation to search is a vital determinant of search performance, however raters cannot agree on how to rate it or are unable to rate it independently of Stamina, then inclusion of the trait Motivation requires consideration. In this study, we used our analyses (ICC, CV, and DV) to produce a shortlist of scales and asked stakeholders to rate the importance of numerous aspects of search dog performance, including all our shortlisted terms. We also asked the raters to score a number of searches for Overall Ability, from 1 (very poor performance) to 10 (excellent performance), as well as for the shortlisted behaviors. The assumption was that those behaviors (subconsciously or consciously) seen as most important when forming an overall impression of a good or bad performance would be correlated most closely with the Overall Ability score.

Quantifying search dog performance on a day-to-day basis potentially has great value. In this paper, we describe our method for developing a suitable instrument. Using a series of evidence-based steps, we assess each behavioral measure in the rating instrument. We explore which behaviors should be included/excluded for both HAS and VS based on the levels of reliability, and where behaviors are deficient in these aspects (yet classed as important by practitioners), we suggest additional scale development and training to improve their reliability to acceptable levels.

METHODS

Selection of Behavioral Measures of Performance

We conducted interviews with stakeholders with varying experiences, including military personnel (senior officers, trainers, instructors, and handlers) and civilian dog trainers working with specialist search dogs. In total, we conducted 23 interviews for VS and 31 interviews for HAS classifications. The first part of the interview was "free-term" generation, where interviewees were asked to describe behaviors that they considered to be important and other factors that might influence performance. Interviewers were careful not to lead interviewees by suggesting particular behaviors, and interviewees were given as much time as needed to describe all aspects of performance. Following from this, interviewees were given a pre-generated list of 37 behaviors linked to different dimensions of performance across all types of search dog. After several interviews, it became evident from the free-term descriptions that new behaviors "Ability to follow search pattern" and "Consistency" should be added, and these were included for the remaining interviews (nine HAS and four VS). Interviewees were asked to rate the ideal level (as low as possible, low, intermediate, high, or as high as possible) and importance (not important, slightly important, important, very important, or critical) of each term. Each response was numerically coded from 1 to 5 (1 = very low to 5 = very high; 1 = not important to 5 = essential) to be able to produce mean importance and ideal levels. Behaviors were ranked according to the mean importance. We excluded those which could not be scored from behavioral observations of a single search, for example, acuity of sense of smell, ease of adaption to kenneling, and health. The 11 most important behaviors for each discipline were selected (**Table 1**), and Speed of search was added after discussion with military personnel as this was believed to be an important rating element.

TABLE 1 | Performance behavior measures, with short titles (referred to in the text) and abbreviated titles (referred to in **Table 3**).

Behavior measure	Behavioral trait name and description, as presented during the rating task. Scored as: 1 (very low), 2 (low), 3 (intermediate), 4 (high), or 5 (very high).
Control (Cont)	Control (responsiveness to verbal and/or physical commands). The proportion of commands obeyed and speed of response.
Motivation (Motiv)	Motivation (enthusiasm to search). How keen or eager the dog is to search—assessed from the dog's behavior leading up to and at the start of the search.
Distraction (Dist)	Distraction when searching. A distraction is anything that takes the dog's attention away from searching or from starting to search, including urinating.
Search pattern (S.Pat)	Ability to follow search pattern. How well the dog follows the correct search pattern, without missing areas or needing constant correction. Not following search pattern would include: HAS, pulling off-line, wide back-seek, or following visual cues; VS, pulling/moving away from vehicle being searched, searching ground, or not searching "overlap."
Stamina (Stam)	Stamina throughout search. How well motivation or enthusiasm is retained over the search, e.g., not decreasing due to tiredness or loss of confidence.
Indication (S.Ind)	Strength of indication.
Confidence (Conf)	Confidence (absence of fear/anxiety) How confident or relaxed the dog is.
Thoroughness (Thor)	Thoroughness of search. How much of the search the dog is actively searching: HAS, sniffing with its head down and nose to the ground for the entire search, including on the back-seek and searching right up to the handler; VS, sniffing with nose to the vehicle.
Independence (Inde)	Independence. Ability of the dog to search without guidance (not needing, or looking for, constant direction), including being able to continue searching when further away from handler and on back-seek.
Detect & locate (D.Loc)	Ability to detect and locate scent to source.
Speed (Spee)	Speed of search.
Consistency (Cons)	Consistent (not erratic) in performance throughout the search

Rating Performance Using Behavioral Measures

Search Videos

Video recordings were made of over 200 training searches (117 VS and 91 HAS) performed by 62 different dogs (35 VS and 27 HAS) in 100 different handler–dog pairings (50 VS and 50 HAS) using a Sony Handycam (DCR-SR58). Most of these training exercises were performed in and around the Defense Animal Centre (DAC), Leicestershire, UK (68% VS and 96% HAS), which is the training school for all UK military working dogs and where dog handlers obtain their initial training. The remaining training searches were recorded at various other Army and Royal Air Force (RAF) bases (in the UK and overseas). To obtain footage of a wide a range of performance, searches were of dogs and handlers with different levels of training and operational experiences. Where possible, an experienced observer (or the handler themselves) rated the performance of the dog immediately after the search was completed. A research scientist also rated each search (at a later date from the video recordings). These scores were used to aid selection of searches. For both HAS and VS, 16 training searches were selected illustrating a wide range of performance. These were each edited to ∼6 min duration, but always including the beginning and the end of the search.

Raters

Raters varied in experience, but all had experience of the particular discipline being studied as either dog trainers, course instructors (training search dog handlers), and/or as operational dog handlers. Although many of the subjects had experience of both classifications, individuals were assigned to either the VS or HAS observation group, with only one person appearing in both groups. The majority of subjects were military (or ex-military) personnel (14 VS and 15 HAS), although there were also two civilian dog trainers. There were 15 VS raters (five females and 10 males), with an average of 6 years working with specialist search dogs (range 3 months to 17 years) and 3 years with VS dogs (range, 2 months to 9 years). Of the 16 HAS raters (four females and 12 male), 12 were personnel from the DAC and four were current course students (handlers learning how to handle HAS dogs). Not including the course students, raters had an average of 5 years' experience working with specialist search dogs (range, 1 month to 10 years) and 2 years with HAS dogs (range, 7 months to 4 years). The difference in the maximum number of years of experience between HAS and VS personnel reflects the greater number of years that VS dogs had been operational as a specific search classification compared to HAS. The course students had been working with HAS dogs for ∼1 month, and their experience with search dogs prior to this ranged from 1 month to 4 years; therefore, we tested to see whether their ratings differed significantly from the remainder of the population.

Video Rating Protocol

The raters attended in groups of between two and 12 people and were given a task introduction, which included some background information on why we were asking them to rate searches and a list of the behaviors they would be rating. As the majority (if not all) had never used performance rating scales before, we gave the following basic instructions aimed at reducing any conscious bias in ratings:

a) Score the videos in silence to avoid influencing each other's scores.
b) Assess each behavior in isolation.
c) Avoid being affected by an overall good or bad impression (halo effect) or being overly influenced by particular event(s) during the search.
d) Resist being influenced by any prior knowledge of the dog.
e) Use the whole 1–5 scale whenever possible (e.g., try not to just use middle ranges).
f) Assess the performance of the dog (not handler) in the particular search being shown.
g) Watch the whole 6-min clip before scoring any behaviors and assess performance based on the whole clip.

All videos were displayed using an overhead projector and screen (with sound). After each clip ended, the raters used a printed sheet to score all 12 behavioral performance measures (**Table 1**) on a scale from 1 to 5, with 1 being the lowest level of the behavior and 5 the highest. They also gave an Overall Ability score from 1 (lowest) to 10 (highest). Once all subjects had scored the video, the next video was shown. Each observation session lasted ∼3 h, with two breaks of between 10 and 20 min, as close to an hour apart as possible without disrupting the task. Thus, not including the Overall Ability score, we collected 2,880 rating scores for VS and 3,072 for HAS (16 videos, 12 behaviors, 15 VS raters, and 16 HAS handlers).

To understand how valuable the raters perceived the behavioral measures *after* rating them, we asked them how easy each of the behaviors were to score (easy, okay, or difficult) and how important they felt each was for assessing performance (essential, okay, or not needed).

Ethics Statement

The project was retrospectively reviewed and approved by both the Faculty of Medical & Veterinary Sciences (FMVS) Research Ethics Committee (concerning human participants) and the Animal Welfare and Ethical Review Board (concerning canine participants) at the University of Bristol. Development of the performance monitoring tool was part of standard military procedure and, as such, was considered to be part of regular duties for participants. Consent was sought from commanding officers. Participants were fully briefed on the purpose of the study and were free to request non-participation from their officers. Data were stored anonymously.

Statistical Methods

To assess whether the four trainee HAS handlers significantly differed from the other raters (due to their relative lack of experience), their ratings for each behavior were compared to the remaining raters using t-tests with a Bonferroni correction applied (244 tests, with α set at 0.05 and significance at $P < 0.002$) (IBM SPSS Statistics).

Inter- and Intra-Reliability of Raters

ICCs were calculated (two-way random effects with absolute agreement) for both average rater and single-rater agreement (**Table 2**). Our cutoff for good reliability was 0.7, although for absolute confidence in the reliability of the raters, we would also want the lower bound of the 95% confidence interval to exceed this. To examine the range of agreement, pairwise correlations (Pearson's) were conducted producing minimum and maximum levels of agreement between raters.

Discriminant and Convergent Validity—Were Observers Able to Distinguish Between the Differing Behaviors as Measuring Separate Aspects of Performance?

Convergent Validity

Measured at the Group Level Because the study was a repeated-measures design, between-behavior correlations need to be interpreted with caution (to avoid errors due to pseudo-replication). Therefore, we used two approaches:

i) We used overall correlation coefficients between behaviors using all ratings, which do not take into account the dependence of repeated within-observer data points. Hence, factors such as clustering by rater may lead to correlations between observers (rather than between behaviors), causing artificial inflation of some coefficients.

ii) We used correlation coefficients between behaviors calculated for every rater and then averaged across all raters. These coefficients may be conservative underestimates of the level of association, as averaged coefficients are likely to be

TABLE 2 | Reliability of behaviors.

VS ***N*** **= 15**	**Control**	**Motivation**	**Stamina**	**Distraction**	**Confidence**	**Independence**	**Consistency**	**Search pattern**	**Thoroughness**	**Speed**	**Detect & locate**	**Strength of indication**	**Overall performance**
Average value	**0.752**	**0.930**	**0.866**	**0.833**	**0.912**	0.668	**0.783**	**0.804**	**0.819**	**0.879**	**0.724**	**0.801**	**0.879**
Lower bound	0.543	**0.867**	**0.745**	**0.780**	**0.823**	0.401	0.601	0.639	0.664	**0.773**	0.495	0.627	**0.771**
Upper bound	**0.898**	**0.971**	**0.946**	**0.951**	**0.967**	**0.862**	**0.909**	**0.918**	**0.925**	**0.950**	**0.886**	**0.919**	**0.950**
Single value	0.168	0.470	0.303	0.334	0.410	0.118	0.194	0.215	0.232	0.327	0.149	0.212	0.325
Min (r) coeff.	−0.659	0.044	−0.337	−0.136	−0.115	−0.423	−0.423	**−0.713**	−0.267	−0.199	−0.550	−0.497	−0.107
Max (r) coeff.	**0.717**	**0.844**	**0.847**	**0.816**	**0.882**	**0.825**	**0.764**	**0.738**	0.669	**0.904**	**0.777**	**0.805**	**0.811**
HAS ***N*** **= 16**	**Control**	**Motivation**	**Stamina**	**Distraction**	**Confidence**	**Independence**	**Consistency**	**Search pattern**	**Thoroughness**	**Speed**	**Detect & locate**	**Strength of indication**	**Overall performance**
Average value	**0.934**	**0.858**	0.605	**0.852**	0.530	0.427	**0.910**	**0.938**	**0.888**	**0.752**	**0.853**	**0.852**	**0.918**
Lower bound	**0.875**	**0.733**	0.243	**0.720**	0.202	−0.037	**0.826**	**0.876**	**0.786**	0.533	0.696	0.684	**0.837**
Upper bound	**0.973**	**0.941**	**0.858**	**0.941**	**0.791**	**0.763**	**0.964**	**0.978**	**0.955**	**0.898**	**0.952**	**0.955**	**0.969**
Single value	0.469	0.274	0.087	0.265	0.066	0.045	0.386	0.487	0.331	0.159	0.267	0.265	0.411
Min (r) coeff.	0.006	−0.338	−0.521	−0.379	**−0.710**	**−0.771**	−0.187	−0.220	−0.311	−0.681	−0.355	−0.463	0.203
Max (r) coeff.	**0.918**	**0.857**	**0.608**	**0.884**	**0.769**	0.632	**0.851**	**0.795**	**0.823**	**0.785**	**0.811**	**0.751**	**0.857**

Average intraclass correlation coefficients with 95% confidence intervals (upper and lower bounds), and predicted single observer reliabilities (coefficients over 0.70 in bold to illustrate strong reliability). Also minimum and maximum (Pearsons) coefficients between pairs of raters are shown to illustrate the range in agreement between observers.

reduced in magnitude (closer to zero) as the raw scores (correlations) can be either positive or negative.

We used Pearson's r correlation coefficient ≥ 0.7 as our threshold, above which we considered convergence to occur between behaviors, with behaviors sharing more than 49% of variation (coefficient of determination, r^2). As this is an arbitrary cutoff (there is no exact figure for convergence), we also discuss correlations in excess of 0.6 as showing enough convergence to warrant concern about behavior validity.

Measured at the Rater Level Since the evaluation instrument is designed for single raters to measure performance in their own dogs, within-rater correlations are potentially more meaningful than group-level correlations. Hence, we also calculated within-rater correlation coefficients (between behaviors). We report the number of raters with coefficients exceeding 0.7 (incidence of CV in **Tables 3**, **4**) across the 16 searches. When summarizing the data, we discuss combinations of behaviors where at least three raters (∼20%) showed strong convergence (>0.7) and/or five or more (∼30%) showed moderate convergence (>0.6) as warranting concern about the ability of raters to evaluate these behaviors independently. We also calculated the standard deviation (SD) across the within-observer and between-behavior correlation coefficients and highlighted combinations of behaviors with greater than average variation in the degree of convergence across raters (SD higher than the mean SD). Minimum and maximum within-rater correlations between pairs of behaviors are presented to illustrate the range of variation in convergence between raters.

Discriminant Validity

To assess whether there was DV between behaviors, we used the correction for correlation attenuation (38), which takes into account measurement errors when comparing the relationship between variables [although also see (39)]. We divided the reliability between behaviors by the square root of the individual behavior reliabilities, multiplied by each other. Average measure ICC values were used as the within-behavior reliability estimates. For each pair of behaviors, a threshold value of between-behavior reliability was calculated using an arbitrary cutoff value of 0.85, as is commonly used to indicate that discriminant validity cannot be assumed (40). Above this calculated within-behavior discriminant validity threshold (DVT), it is likely that apparently independent behaviors were in fact being used by raters to measure the same underlying construct. The number of raters exceeding this between-behavior reliability threshold for each behavior combination (therefore not exhibiting DV) was recorded and behavior combinations where at least three (∼20%) raters did not discriminate between behaviors were highlighted.

Deciding Which Behaviors Were Reliable, Essential, and Easy to Rate

Responses to questions concerning how easy it was to rate each behavior and how important they were to include were coded (1-3) and averaged across the participants. The mean values were rounded to the nearest whole number and converted back into the headings as they appeared on the form to represent a consensus for ease of rating and importance (**Table 4**). In addition, the correlation coefficients between each individual behavior and Overall Ability were listed (**Table 4**).

RESULTS

There was no significant difference between the four HAS course handlers and the other HAS raters in the scores given for the rated behaviors; therefore, the whole sample of handlers are analyzed together.

Reliability—How Much Did Raters Agree?

There was, in general, good agreement within the groups of HAS and VS students (average rater ICC) for most behaviors, but the level of agreement varied between classifications (**Table 2**, summarized in **Table 4**). Group-level reliability estimates exceeded 0.7 for VS raters for all behaviors, except Independence (0.668), although the lower bound of the 95% CI failed to reach the 0.7 level for Control, Consistency, Search Pattern, Thoroughness, Detect and Locate, and Indication. HAS raters did not reach good agreement when rating Independence, Stamina, or Confidence. In addition, the lower bound of the 95% CI was below 0.7 for Speed, Detect and Locate, and Indication. Visually comparing reliabilities for the classifications, there was greater agreement by VS raters for Motivation, Stamina, Confidence, and Speed compared to HAS raters, whereas HAS raters showed greater agreement in their ratings for Control, Consistency, Search Pattern, Thoroughness, and Detect and Locate.

Considering the reliability of single raters (single-measure ICCs), agreement was poor to moderate at best for both VS and HAS. This indicates considerable variation between raters, which was also demonstrated by the large range between the minimum and maximum coefficients for the correlations between pairs of raters (**Table 2**) and significantly reduces confidence in the ability of individuals (as opposed to groups of raters) to use the measures reliably.

Were Observers Able to Distinguish Between the Selected Behaviors as Measuring Separate Aspects of Performance?

Convergent Validity

Group Level

Looking at the "all ratings" (and using the 0.7 cutoff), we found no convergent validity between any of the VS behaviors (**Table 3a**, summarized in **Table 4**), indicating that the raters were able to observe the dogs in action and rate the behaviors independently of one another (i.e., as unique facets of working dog behavior), although there was some indication that both Stamina (0.631) and Speed (0.639) were seen as related to Motivation. Within the HAS ratings, there were moderate correlations (>0.6) between Control, Consistency, and Search Pattern, with the relationship between Control and Search Pattern (0.771) exceeding the threshold for convergence.

TABLE 3a | Convergent and discriminant validity for between-behavior comparisons (VS lower left section of table and HAS upper right section).

			Cont	Moti	Stam	Dist	Conf	Inde	Cons	S.Pat	Thor	Spee	D.Loc	S.Ind	O.Abi
(a) All		Cont		0.458**	0.310**	−0.324**	0.281**	0.170**	**0.652****	**0.771****	0.515**	0.128*	0.480**	0.239**	**0.809****
(b) Ave.	HAS →			0.457	0.353	−0.459	0.281	0.031	**0.698**	**0.758**	0.524	0.154	0.463	0.204	**0.800**
(c) DVT				0.761	0.639	0.758	0.598	0.537	0.784	0.796	0.774	0.712	0.759	0.758	0.787
(a) All		Moti	0.326**		0.416**	−0.361**	0.421**	0.261**	0.476**	0.472**	0.389**	0.307**	0.355**	0.253**	0.569**
(b) Ave.	VS ↓		0.284		0.456	−0.440	0.433	0.129	0.470	0.471	0.426	0.270	0.315	0.221	0.538
(c) DVT			0.711		0.612	0.727	0.573	0.514	0.751	0.763	0.742	0.683	0.727	0.727	0.754
		Stam	0.321**	**0.631****		−0.352**	0.289**	0.172**	0.386**	0.280**	0.355**	0.070	0.274**	0.191**	0.379**
			0.275	**0.630**		−0.356	0.251	0.074	0.373	0.346	0.379	0.055	0.245	0.128	0.395
			0.686	0.763		0.610	0.481	0.432	0.631	0.640	0.623	0.573	0.611	0.610	0.633
		Dist	−0.355**	−0.366**	−0.333**		−0.265**	−0.154*	−0.557**	−0.380**	−0.435**	−0.179**	−0.224**	−0.145*	−0.415**
			−0.359	−0.389	−0.276		−0.291	0.005	−0.526	−0.510	−0.515	−0.196	−0.264	−0.173	−0.555
			0.673	0.748	0.722		0.571	0.513	0.748	0.760	0.739	0.680	0.725	0.724	0.752
		Conf	0.262**	0.459**	0.390**	−0.344**		0.277**	0.379**	0.324**	0.435**	0.150*	0.253**	0.246**	0.399**
			0.149	0.348	0.298	−0.358		0.171	0.328	0.384	0.377	0.192	0.245	0.222	0.398
			0.704	0.783	0.755	0.741		0.404	0.590	0.599	0.583	0.537	0.572	0.571	0.593
		Inde	0.262**	0.419**	0.322**	−0.293**	0.496**		0.374**	0.201**	0.144*	0.158*	0.083	0.007	0.232**
			0.189	0.356	0.277	−0.377	0.384		0.160	0.087	0.057	0.084	0.060	−0.040	0.118
			0.602	0.670	0.646	0.634	0.663		0.530	0.538	0.523	0.482	0.513	0.513	0.532
		Cons	0.481**	0.456**	0.400**	−0.490**	0.447**	0.419**		**0.665****	0.580**	0.191**	0.405**	0.259**	**0.707****
			0.402	0.430	0.334	−0.509	0.388	0.449		**0.728**	0.575	0.207	0.428	0.232	**0.748**
			0.652	0.725	0.700	0.686	0.718	0.615		0.785	0.764	0.703	0.749	0.748	0.777
		S.Pat	0.497**	0.287**	0.237**	−0.408**	0.248**	0.312**	0.540**		0.561**	0.254**	0.415**	0.216**	**0.829****
			0.491	0.271	0.242	−0.394	0.183	0.297	0.478		0.568	0.287	0.475	0.231	**0.831**
			0.661	0.735	0.709	0.696	0.728	0.623	0.674		0.776	0.714	0.760	0.760	0.789
		Thor	0.488**	0.383**	0.376**	−0.479**	0.299**	0.324**	0.569**	0.563**		0.098	0.384**	0.269**	**0.635****
			0.444	0.335	0.306	−0.455	0.177	0.287	0.488	0.483		0.095	0.355	0.214	**0.621**
			0.667	0.742	0.716	0.702	0.735	0.629	0.681	0.690		0.695	0.74	0.739	0.767
		Spee	0.176**	**0.639****	0.521**	−0.225**	0.320**	0.315**	0.253**	0.184**	0.194**		0.157*	0.088	0.194**
			0.097	**0.610**	0.440	−0.178	0.170	0.182	0.176	0.134	0.094		0.106	0.026	0.202
			0.360	0.200	0.270	0.270	0.240	0.220	0.32	0.260	0.230		0.681	0.680	0.706
		D.Loc	0.151*	0.317**	0.217**	−0.253**	0.271**	0.156*	0.264**	0.262**	0.322**	219**		0.545**	0.548**
			0.172	0.241	0.124	−0.232	0.146	0.063	0.253	0.325	0.331	0.097		0.438	0.534
			0.627	0.697	0.673	0.660	0.691	0.591	0.640	0.649	0.655	0.678		0.725	0.752
		S.Ind	0.210**	0.130*	0.140*	−0.277**	0.286**	0.065	0.224**	0.196**	0.350**	−0.026	0.567**		0.383**
			0.226	0.006	−0.011	−0.197	0.216	0.038	0.175	0.234	0.317	−0.155	0.434		0.336
			0.660	0.734	0.708	0.694	0.726	0.622	0.673	0.682	0.688	0.713	0.647		0.752
		O.Abi	0.569**	**0.694****	0.558**	−0.561**	0.590**	0.440**	**0.638****	0.512**	**0.657****	0.416**	0.472**	0.419**	
			0.526	**0.672**	0.520	−0.565	0.500	0.417	0.598	0.539	**0.623**	0.360	0.460	0.351	
			0.691	0.769	0.742	0.727	0.761	0.651	0.705	0.715	0.721	0.747	0.678	0.713	

*Coefficients (CV) for (a) "all ratings" and (b) "average" ratings exceeding 0.6 highlighted in bold. DVT, discriminant validity threshold coefficient above which DV between behaviors cannot be assumed. O.Ab, overall ability. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.*

TABLE 3b | Incidence of raters exceeding thresholds for convergent and discriminant validity and range of within-rater coefficients (VS lower left section of table and HAS upper right section).

		Cont	Moti	Stam	Dist	Conf	Inde	Cons	S.Pat	Thor	Spee	D.Loc	S.Ind	O.Abi
(a) CV 0.7(0.6)	Cont		**2 (6)**	0 (1)	1 (4)	0 (2)	**4 (8)**	**10 (12)**	**11 (15)**	**3 (7)**	0 (0)	**2 (7)**	0 (1)	**15 (15)**
(b) >DVT	HAS→			1		2	**8**	**8**	**6**			1		**11**
(c) Range			−0.074–0.725	−0.039–0.665	−0.719–0.066	**−0.354–0.628**	**−0.864–0.824**	0.238–0.929	0.561–0.929	0.074–0.763	**−0.387–0.583**	−0.108–0.785	−0.426–0.639	0.462–0.900
CV 0.7(0.6)	Moti	1 (2)		**3 (6)**	1 (3)	**4 (6)**	1 (2)	**4 (6)**	**3 (8)**	**3 (5)**	0 (0)	2 (3)	0 (1)	**6 (9)**
>DVT	VS ↓			**6**	1	**7**	**4**	**3**		2		2		**5**
Range		**−0.210–0.728**		**−0.064–0.889**	−0.816–0.003	**−0.147–0.796**	**−0.531–0.758**	−0.152–0.855	−0.204–0.742	−0.078–0.804	−0.198–0.553	**−0.334–0.881**	−0.284–0.624	0.185–0.887
	Stam	0 (1)	**7 (9)**		0 (1)	1 (1)	0 (1)	1 (3)	1 (1)	**4 (5)**	0 (0)	0 (1)	0 (0)	1 (3)
			4		1	2	**3**	2	1	**4**	1	1		**3**
		−0.119–0.653	0.334–0.873		−0.695–0.125	−0.121–0.844	**−0.487–0.646**	0.000–0.702	−0.104–0.763	**−0.046–0.871**	**−0.478–0.597**	−0.036–0.614	−0.596–0.583	−0.028–0.832
	Dist	2 (3)	1 (4)	0 (3)		0 (1)	1 (2)	**4 (8)**	**2 (7)**	**4 (6)**	0 (0)	0 (2)	0 (0)	**3 (7)**
		2	1			**3**	**4**	2	1	**4**				**3**
		−0.761–0.077	**−0.781–0.211**	**−0.683–0.143**		**−0.661–0.278**	**−0.590–0.734**	−0.766–0.105	−0.789–0.000	−0.827–0.142	−0.545–0.096	−0.644–0.366	−0.408–0.298	−0.828–0.106
	Conf		**3 (3)**	0 (1)	1 (3)		1 (3)	**2 (5)**	**2 (6)**	1 (3)	0 (0)	0 (0)	0 (2)	**3 (7)**
			1				**6**	**5**	**6**	**3**		2	2	**7**
		−0.448–0.523	**−0.162–0.821**	**−0.107–0.602**	−0.732–0.063		**−0.500–0.717**	**−0.237–0.709**	**−0.357–0.773**	−0.150–0.783	−0.322–0.498	**−0.366–0.594**	**−0.239–0.645**	**−0.326–0.827**
	Inde		1 (3)	1 (3)	2 (4)	1 (2)		**4 (5)**	**5 (7)**	0 (1)	0 (1)	1 (3)	0 (1)	**6 (8)**
			2	2		1		**8**	**8**	2	2	**3**	1	**10**
		−0.341–0.523	**−0.175–0.818**	**−0.273–0.755**	**−0.800–0.156**	−0.113–0.718		**−0.854–0.867**	**−0.854–0.891**	**−0.641–0.586**	**−0.494–0.610**	**−0.744–0.688**	**−0.508–0.607**	**−0.914–0.840**
	Cons	**3 (4)**	**2 (5)**	1 (2)	**4 (7)**	1 (3)	2 (4)		**12 (13)**	**7 (8)**	1 (1)	**2 (5)**	0 (0)	**13 (14)**
		3	1	1		1	**4**		**9**	**5**	1	2		**11**
		−0.632–0.829	**−0.159–0.787**	**−0.516–0.72**	**−0.872–0.071**	−0.114–0.759	0.102–0.751		0.164–0.968	0.219–0.831	**−0.386–0.721**	0.065–0.792	−0.149–0.509	0.474–0.928
	S.Pat	**1 (6)**	0 (1)		**1 (8)**		1 (3)	**5 (9)**		**5 (7)**	1 (1)	2 (3)	0 (0)	**14 (15)**
		3					**3**	**6**		**3**	1	1		**13**
		0.167–0.731	−0.241–0.663	−0.190–0.619	**−0.726–0.398**	−0.208–0.513	−0.177–0.77	**−0.395–0.899**		0.232–0.887	**−0.474–0.722**	0.098–0.777	−0.201–0.521	0.503–0.934
	Thor	0 (3)	0 (1)		**3 (4)**		0 (2)	**2 (6)**	**5 (8)**		0 (2)	**1 (5)**	0 (1)	**8 (9)**
		1						2	**5**			1		**4**
		0.129–0.670	−0.257–0.695	**−0.430–0.590**	−0.821–0.116	−0.064–0.583	0.035–0.657	−0.334–0.883	**−0.284–0.906**		**−0.620–0.514**	**−0.345–0.802**	−0.113–0.666	0.307–0.911
	Spee	1 (1)	**7 (10)**	2 (4)		0 (1)	0 (1)	0 (1)				1 (1)	0 (0)	0 (0)
		1	2	2								1		
		−0.744–0.537	0.227–0.888	**−0.223–0.793**	**−0.557–0.162**	−0.406–0.672	−0.242–0.615	**−0.476–0.610**	**−0.395–0.457**	−0.281–0.438		**−0.493–0.739**	−0.417–0.508	**−0.393–0.566**
	D.Loc	1 (1)	1 (2)		0 (1)			0 (3)	0 (2)	0 (1)			**3 (5)**	**6 (9)**
		1	1					1	1	1			**3**	**3**
		−0.257–0.712	**−0.526–0.719**	**−0.349–0.487**	**−0.681–0.462**	**−0.422–0.587**	−0.301–0.42	**−0.255–0.644**	**−0.324–0.680**	−0.183–0.689	−0.324–0.448		**−0.292–0.907**	−0.116–0.830
	S.Ind	1 (1)	1 (1)			0 (1)		0 (1)	0 (1)		0 (1)	**3 (6)**		0 (4)
		1										**5**		
		−0.234–0.709	**−0.579–0.703**	−0.384–0.368	−0.542–0.349	−0.122–0.641	−0.395–0.451	−0.025–0.666	−0.237–0.617	−0.121–0.567	**−0.696–0.412**	−0.024–0.726		−0.168–0.671
	O.Abi	**4 (6)**	**6 (12)**	**2 (5)**	**6 (9)**	**4 (6)**	**0 (3)**	**5 (9)**	**6 (8)**	**9 (10)**	**0 (3)**	**1 (7)**	0 (3)	
		5	**2**	**1**		**6**	2	**5**	**5**	**7**		**2**	1	
		0.136–0.917	0.539–0.799	0.251–0.755	**−0.872–0.137**	**0.048–0.856**	0.142–0.694	**−0.320–0.895**	**−0.086–0.793**	0.259–0.837	−0.241–0.612	**−0.211–0.739**	−0.152–0.693	

Number of raters exceeding (CV > 0.7), between traits (a), incidence of moderate correlations (>0.6) in brackets; and incidence of within-rater coefficients exceeding DV Threshold (b) (See ***Table 3a****). Behavior combinations where more than 3 raters exceeded CV (and/or more than 5 >0.6), or a lack of DV are highlighted in bold. (c) range of within-rater coefficients (Pearson's r), where in bold the standard deviation is greater than the sample mean (VS = 0.255, HAS = 0.248).*

TABLE 4 | Summary table for outcomes of evidence-based 3-step methodology for rating scale development.

	(1) Reliability			**(2) Variability in rating**		**(3) Importance**			
	Within behavior (ICC)		**Correlated with other behaviors**[a]	**Association with other behaviors**[b] **(CV and/or lack of DV)**	**Of 11 behavior comparisons, number where SD > mean SD**	**Correlation with overall ability**[c]	**Original ranking of importance**	**Post observation importance (Essential, OK, Not-needed)**	**Ease of rating (Easy, OK, Difficult)**
	>0.7	**CI > 0.7**							
VS									
Control	✓	x		Cons, S.Pat	8	0.569	6.5	1.1 (Ess)	1.2 (Easy)
Motivation	✓	✓	Stam, Spee	Stam, Conf, Cons, Spee	7	**0.694**	1	1.1 (Ess)	1.5 (Easy)
Stamina	✓	✓	Moti	Mot	7	0.558	5	1.4 (Ess)	1.7 (OK)
Distraction	✓	✓		Cons, S.Pat, Thor	8	−0.561	11	1.3 (Ess)	1.3 (Easy)
Confidence	✓	✓		Moti	4	0.590	8	1.5 (Ess)	1.7 (OK)
Independence	x	x		Cons, S.Pat	4	0.440	10	1.8 (OK)	1.9 (OK)
Consistency	✓	x		Cont, Moti, Dist, Indep, Cons, S.Pat	7	**0.638**	9	1.5 (Ess)	1.5 (Easy)
Search pattern	✓	x		Cont, Dist, Indep, Cons, Thor	5	0.512	3	1.5 (Ess)	1.5 (Easy)
Thoroughness	✓	x		Dist, Cons, S.Pat	2	**0.657**	4	1.7 (Ess)	1.7 (OK)
Speed	✓	✓	Moti	Motiv	6	0.416	-	2.3 (OK)	2.1 (OK)
Detect & locate	✓	x		S.Ind	7	0.472	6.5	1.4 (Ess)	2.1 (OK)
Strength of indication	✓	x		D.Loc	3	0.419	2	1.2 (Ess)	1.4 (Easy)
HAS									
Control	✓	✓	Cons, **S.Pat**	Moti, Inde, Cons, S.Pat, Thor, D.Loc	3	**0.809**	3.5	1.0 (Ess)	1.1 (Easy)
Motivation	✓	✓		Moti, Conf, Inde, Cons, S.Pat, Thor	4	0.569	1	1.2 (Ess)	1.2 (Easy)
Stamina	x	x		Moti, Inde, Thor	4	0.379	6	1.7 (OK)	1.9 (OK)
Distraction	✓	✓		Conf, Inde, Cons, S.Pat, Thor	2	−0.415	9	1.3 (Ess)	1.3 (Easy)
Confidence	x	x		Moti, Dist, Inde, Cons, S.Pat, Thor	8	0.399	8	1.3 (Ess)	1.6 (OK)
Independence	x	x		Cont, Moti, Stam, Dist, Conf, Cons, S.Patt, D.Loc	11	0.232	11	2.5 (NN)	2.9 (Diff)
Consistency	✓	✓	Cont, S.Pat	Cont, Moti, Dist, Conf, Inde, S.Pat, Thor, D.Loc	2	**0.707**	10	1.7 (OK)	1.5 (OK)
Search pattern	✓	✓	**Cont**, Cons	Cont, Moti, Dist, Conf, Inde, S.Pat, Thor	3	**0.829**	3.5	1.2 (Ess)	1.3 (Easy)
Thoroughness	✓	✓		Cont, Moti, Stam, Dist, Conf, Cons, S.Patt, D.Loc	4	**0.635**	2	1.1 (Ess)	1.3 (Easy)
Speed	✓	x			7	0.194	-	1.7 (OK)	1.4 (Easy)
Detect & locate	✓	x		Cont, Inde, Cons, Thor, S.Ind	5	0.548	7	1.3 (Ess)	1.8 (OK)
Strength of indication	✓	x		D.Loc	3	0.383	5	1.0 (Ess)	1.5 (Easy)

(1) reliability, or agreement between raters (within-behavior reliability (ICC) > 0.7, and lower bound of confidence interval > 0.7). [a]*Convergence at r > 0.7 (bold); or moderate correlation r > 0.6 (**Table 3a**).*

(2) variability in rating, ability of raters to distinguish between behaviors using correlation and construct validity (high CV and low DV), and how variable raters were in their ability to distinguish between behaviors (greater than mean standard deviation). [b]*Considered to be where behaviors converged for least 3 observers at > 0.7, or at least 5 observers at > 0.6, and/or where at least 3 observers did not discriminate between dimensions (**Table 3b**).*

(3) importance and ease of use (the lower the mean score the more important to include, or the easier to rate) compared to relative rank importance from pre-observation interviews (no ranks were available for Speed as this was not included on the original list of behaviors compiled before searches were rated, and Search Pattern was added later and was rated by a small number of interviewees) and behavior correlations with Overall Ability. [c]*All ratings correlation coefficient, bold where > 0.6 (**Table 3a**).*

Rater Level

For both classifications, some raters had greater difficulty than others in distinguishing between the independent dimensions of performance. This was more evident in the HAS group: of the 66 possible comparisons between the 12 behaviors, the number of correlations exceeding ± 0.7, and therefore indicating a strong association, were between 1 and 19 per rater (mean = 7.6) for HAS compared with between 1 and 9 per rater (mean = 3.5) for VS. One particular HAS rater scored the same value (4) for Stamina and Confidence in all 16 video clips, meaning that no correlation coefficients between these behaviors and the other performance measures were possible.

The ratings for Speed and Motivation were moderately correlated (>0.6) for 10 of the 15 VS raters, as seen similarly at the group level (**Table 3**, summarized in **Table 4**). However, Search Pattern and Consistency, which were not significantly correlated within the whole group, were seen as related behaviors by nine of the 15 raters. This was presumably because the raters differed in their interpretation of the relationship between the behaviors, as illustrated by the within-rater coefficients varying between −0.395 and 0.899, therefore bringing the overall coefficient below the 0.7 threshold. There were several other pairs of behaviors where correlations were found at the rater, but not group, level (Distraction with both Consistency and Search Pattern, Indication with Detect and Locate). This was also true for the HAS raters. As expected from the group-level correlations, several raters did not rate Control, Search Pattern, and Consistency independently of each other; but several raters also saw considerable associations between behaviors, including Independence and Thoroughness with Search Pattern, and Distraction and Thoroughness with Consistency.

For the HAS raters, there was clearly much variation in the interpretation of Independence, with the SDs for all 11 comparisons between this and the other behaviors having higher than the average values (**Tables 3**, **4**). Other behaviors where there was much variation in the degree of convergence across the HAS raters were Confidence and Speed. Several behaviors showed above-average variability in the degree of correlation with other behaviors within VS ratings (Control, Motivation, Stamina, Distraction, Consistency, and Detect and Locate).

Discriminant Validity—Was There Greater Variation Between Behaviors Than Within Behaviors?

Figure 1 summarizes where there was some "interpretive overlap" between pairs of behaviors (individual rater correlations exceeding the DVT) by at least three observers (for the actual numbers, see **Table 3**, summarized in **Table 4**). Some VS raters evidently saw significant associations between Thoroughness, Search Pattern, Control, Distraction, and Consistency. For the HAS raters, fewer raters showed discriminant validity between behaviors compared to the VS ratings, and the relationships between behaviors were slightly more complex. The strongest links (affecting at least half of the raters) were between Consistency, Search Pattern, Control, and Independence.

Which Behaviors Are Essential to Include and Easy to Rate?

Of the 12 behaviors, 10 were classed as "essential" to include as VS performance measures (excluding Independence and Speed) (**Table 4**), and eight behaviors were considered to be essential for measuring HAS performance. Control, Motivation, Thoroughness, Levels of Distraction, and Strength of Indication were generally seen as essential for both VS and HAS, but there were differences in the relative importance of the other behaviors between the classifications. Consistency of behavior and Stamina were more important to the VS raters, for example, while Confidence and Ability to follow Search Pattern were ranked more highly for the HAS dogs. Of the essential behaviors within each classification, those considered harder to rate than others (okay rather than easy) include Thoroughness, Detect and Locate, Stamina, and Confidence for VS and Indication, ability to Detect and Locate, and Confidence for HAS. Both the VS and HAS raters thought Control, Motivation, ability to follow Search Pattern, and Speed were easily rated measures.

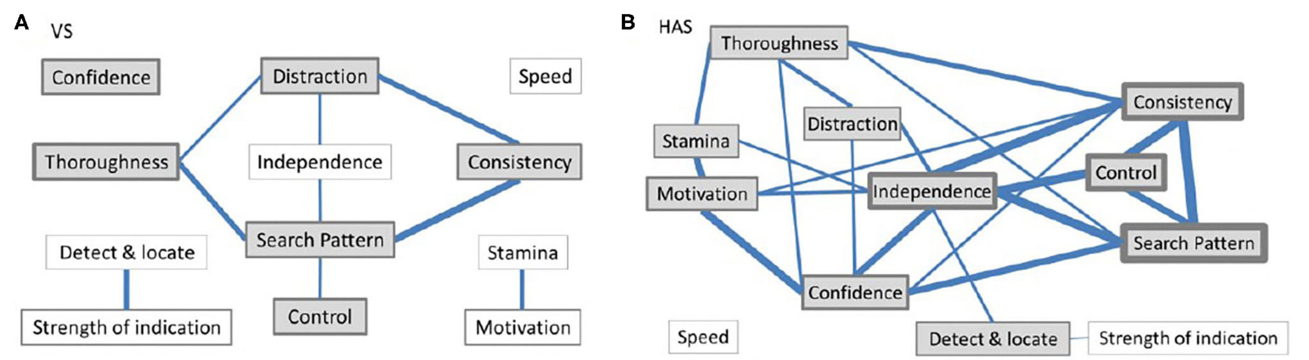


FIGURE 1 | Diagrammatic representation of discriminant validity (or lack of) between **(A)** VS and **(B)** HAS performance traits. Behaviors where at least three raters (~20%) did not discriminate between them are connected, with thicker lines indicating a greater number of raters violating discriminant validity assumptions. Behaviors that correlated with Overall ability (using "all ratings") are shaded, with the relative thickness of the border indicating the incidence of correlations (convergent validity). For values and coefficients see **Tables 2**, **3**.

Which Behaviors Correlated With Overall Performance?

Looking at the "all ratings" correlations, the only behaviors that the VS raters associated with Overall Ability (above 0.6) were Motivation, Consistency, and Thoroughness (**Table 3**, summarized in **Table 4**), although when considering individual raters, several associated Overall Ability with Distraction, Search Pattern, Confidence, and Control. For the HAS raters, "all ratings" correlations between Overall Ability and each of Control, Consistency, Search Pattern, and Thoroughness exceeded 0.6. Several individual HAS raters also associated Motivation, Distraction, Confidence, Independence, and Detect and Locate with Overall Ability. Speed and Indication were not greatly correlated with Overall Ability in either classification, despite Indication being described as essential to performance.

DISCUSSION

The behaviors generated by stakeholder interviews and workshops showed good reliability at a group level and, in general, measured independent dimensions of search performance, making them potentially suitable to include in a performance measurement tool. However, there are several indications that using this first step alone, or simply assessing reliability and validity at the level of a group of raters without further validation, would not provide an effective and reliable performance measurement tool for single raters. Despite good agreement between observers at the group level (average ICCs), the single-rater ICC values were low, which was also reflected by the large variation between raters in the within-rater correlations for individual behaviors, from near-perfect agreement to a strong negative agreement. At present, this means that our measures of performance are useful, valid, and reliable when used by multiple raters, but that reliability, when a single rater is used, is potentially below the levels of acceptability. This is especially important for evaluation instruments, which are ultimately to be used by single raters, such as this. Some measures were more reliable than others, and individual observers showed considerable variability in their ability to distinguish between behaviors and to reliably rate them. This means that further development is required, for example by providing the raters with training, which was not done here. Any such development should be followed by a reassessment of the measures using the methods described. We also found differences between the raters for the two search dog classifications, the reasons for which are discussed below.

At the group level, the reliability of the behaviors was generally acceptable, with some exceptions. When considering VS dogs, Independence was not adequately reliable (at the 0.7 cutoff), and since it was also not rated as "essential," we suggest that this behavior should be removed. Stamina, Speed, and Motivation also showed some convergence, which, although less than the 0.7 cutoff, was > 0.6 and is therefore of concern. Considering the HAS dogs, three behaviors—Confidence, Independence, and Stamina—all fell short of adequate reliability. We would suggest that the latter two could be removed at this stage as neither was considered "essential" by the group of raters as a whole. Confidence was, on average, rated as essential by the HAS handlers, but the handlers did not agree on how "confident" an individual dog was. Control and Search Pattern were convergent, but were classed as essential, and there was marginal convergence (between 0.6 and 0.7) with both of these behaviors and Consistency, although this was not seen as an essential behavior. This initial stage, therefore, detected some behaviors for both disciplines that fall short of the required levels of reliability and validity at the group level; hence, even if the instrument were to be used by a panel of raters, we would recommend the removal of one and three behaviors, respectively, for the VS (Independence) and HAS (Independence, Stamina, and Consistency) instruments. There were further behaviors, which, while also lacking reliability and validity, were considered essential to include. We will discuss how this might be addressed after the second stage, assessing reliability and validity at the single-rater level.

At the single-rater level, none of the behaviors attained our predetermined levels of reliability (single-value ICCs). In addition, almost all behaviors showed convergence and a lack of discriminant validity with at least one other measure, as determined by our cutoff values for the number of raters reaching the CV and DV thresholds. We could not, therefore, recommend using the behaviors tested here as a rating scale to be used by single dog handlers, or those without training on scale use, as is commonly done.

This finding is perhaps unsurprising. Previous studies suggest that, without training, raters are likely to vary in their ability to use dog performance rating scales (41), and hence there is potential for rater error. Although it is likely that this can be ameliorated with training, as is sometimes employed when rating dog behavior [e.g., (42)], we did not provide rater training here as we wanted to assess the existing differences within our sample population prior to external influences. We deliberately provided raters with very little information on what constituted each behavior to avoid biasing their ratings and to mimic what may realistically occur in the field. Although raters were generally experienced in observing and assessing dog performance, they were not experienced in using scales of this type. Once presented with the terms, there was no discussion permitted between raters, as this may have facilitated them deriving common conceptualization. The interpretation of terminology varied between individuals naive to the testing scale, which was also evident at the initial term derivation stage with stakeholders. A low rater agreement likely reflects the absence of common understanding, as individuals use their own idea of the constructs. It is likely that some behaviors are inherently more difficult to rate than others because they are harder to conceptualize or are more evaluative (43). We predicted Independence to be cognitively harder to rate as it is a more abstract concept [less observable; (43)] compared to the more quantifiable behaviors such as Control. This appeared to be the case as the mean coefficients for the former were low. Raters also reported Independence as being difficult to rate (**Table 4**). Rating Confidence (seen as essential by the HAS handlers) relies on recognizing the signs that a dog is not fearful, which are known to be difficult to spot without training, even for dog owners and careers (44–46). Where behaviors, such as Confidence, are judged to be an important element of performance, we suggest that efforts should be made to improve the reliability of their rating. This could be achieved by improving the recognition of behavioral signs, for example using training resources detailing the subtle signs of fear in dogs, as are now available (47).

The next development stage for this working dog performance measurement tool would therefore be to explore techniques to develop a common understanding of behavior terminology among raters. Possible methods, beyond the scope of this study, include group discussions, workshops and training sessions, or benchmarking the scales [see (48)] presenting detailed descriptions of each level of each behavioral scale. It is important to emphasize that following this development process, the reliability and validity of any behavioral measure would need to be retested using our suggested methods to assess whether it now reached the required cutoff values to give confidence in the data obtained.

The current study demonstrates that simply assessing reliability at the group level is insufficient if the ratings are subsequently to be made by individuals, as group-level (or average) coefficients can disguise a multitude of issues. As the behavior measures tested here were to form part of a tool designed to be used by single handlers, it was vital to examine how well-individual raters make judgments on performance and whether they reached an acceptable threshold. In this study, they did not. Given the solitary nature of many search tasks, handler-completed subjective rating scales of search dog performance are the most practically feasible method of monitoring search-by-search performance. However, if decisions are to be made on the basis these ratings, there must be confidence in the data obtained. Unreliable measures will, at best, add to the rater (handler) burden without providing any additional information and, at worst, could result in misleading information being collected.

Our results demonstrate that there is unlikely to be a "one-size-fits-all" measurement tool capable of capturing important aspects of performance across search dog classifications. We started with common behaviors as, for practical efficiencies, managers would much prefer a single performance instrument. However, although there were similarities between the two search classifications, our process confirmed that there were significant differences in the importance, reliability, and validity of specific behaviors between the two groups of raters. Not only were the classifications different, but it seems that our raters may also have differed. The HAS observers showed very strong average reliabilities for several behaviors, but also considerable convergence between behaviors. It is possible that the videoed HAS searches did not contain enough variation in performance and that the behaviors assessed were, in reality, correlated, in which case the raters were simply reflecting this "true halo" (where behaviors are not independent but covary). Efforts had been made to avoid this by using training searches of dogs that had only just begun training. Also, the ratings supplied by the trainers/handlers in the live searches indicated that we did have a wide range of search performance. Alternatively, there may have been differences between the way that the classifications operate and are trained, which altered the way the handlers use the rating scales or induce different biases, for example, operating a more rigid thought process that may mimic the more rigid search requirements of a high assurance search. This could be an interesting avenue for future investigation, for example, investigating the effectiveness of different training methods with populations of handlers who may vary in their backgrounds, openness, and flexibility to altering their internal conceptualizations. What is clear is that working dog performance scales need to be derived for specific search tasks.

SUMMARY

There are two elements to obtaining reliable performance measures: producing an effective instrument and ensuring effective scoring. The development method used here followed a considered, effective, and clear process for testing reliability and validity, which is essential to enable confidence in any data obtained. Because of the nature of the task, using untrained handlers and providing little information, the method demonstrated that the behavior measures given in their current format to naive handlers would not produce reliable and repeatable results. This is an important demonstration for researchers and practitioners using rating scales without full validation, especially where reliabilities are tested at the group level but the end user is an individual. Overall, most of the behaviors were reliable and showed good construct validity at the group level. Therefore, after removal of a small number, the measures could be useful and applicable when assessed by a group of raters. But this was not true at the individual rater level, which is ultimately the target for an instrument of this kind. It is therefore important to look at the individual rater level for convergence, discriminant validity, and reliability, not just at the group level. To be practically valuable, individual rater performance needs to be improved to ensure that the instrument is utilized effectively. The next steps for the development of the search dog performance measurement tool are therefore to understand why raters vary and to undertake measures to improve the ability of individual raters by increasing the likelihood that they form a common conceptualization of each behavior construct. It is then vital to retest the validity and reliability using the method described here.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because they contain information on military working dog performance and are thus sensitive. Requests to access the datasets should be directed to Nicola J. Rooney, nicola.Rooney@bristol.ac.uk.

ETHICS STATEMENT

The project was retrospectively reviewed and approved by both the Faculty of Medical & Veterinary Sciences (FMVS) Research Ethics Committee (concerning human participants) and the Animal Welfare & Ethical Review Board (concerning canine participants) at the University of Bristol.

AUTHOR CONTRIBUTIONS

NR conceived the idea. CC and NR developed the methods and conducted rater workshops. CC recorded the behavior and performed the behavioral analysis with support from NR. Both authors prepared the manuscript.

FUNDING

This study was supported by UK Ministry of Defence (MOD). Quantification of Performance – VS and HASD dog roles Ref DFAS RK7283.

ACKNOWLEDGMENTS

We would like to thank personnel of the 1st Military Working Dogs Regiment (Royal Army Vet Corps), the Defence Animal Centre, and the RAF Police for their invaluable inputs throughout this project. We would also like to thank Professor Bill Browne for statistical advice and Dr. Jo Hockenhull, Dr. Steve Nicklin, and Col. Neil Smith for very useful comments on the manuscript.

REFERENCES

1. Adams GJ, Johnson KG. Sleep, work, and the effects of shift work in drug detector dogs canis familiaris. *Appl Anim Behav Sci.* (1994) 41:115–26. doi: 10.1016/0168-1591(94)90056-6
2. Almey H, Nicklin S. How does your dog smell? A review of canine olfaction. *J Defence Sci.* (1996) 1:345–52.
3. Cablk ME, Heaton JS. Accuracy and reliability of dogs in surveying for desert tortoise (*Gopherus agassizii*). *Ecol App.* (2006) 16:1926–35. doi: 10.1890/1051-0761(2006)0161926:AARODI2.0.CO;2
4. DeMatteo KE, Davenport B, Wilson LE. Back to the basics with conservation detection dogs: fundamentals for success. *Wildlife Biol.* (2019) 1:1–9. doi: 10.2981/wlb.00584
5. Lazarowski L, Waggoner LP, Krichbaum S, Singletary M, Haney P. Selecting dogs for explosives detection: behavioral characteristics. *Front Vet Sci.* (2020) 7:597. doi: 10.3389/fvets.2020.00597
6. McGarrity ME, Sinn DL, Thomas SG, Marti CN, Gosling SD. Comparing the predictive validity of behavioral codings and behavioral ratings in a working-dog breeding program. *Appl Anim Behav Sci.* (2016) 179:82–94. doi: 10.1016/j.applanim.2016.03.013
7. Wilsson E, Sundgren PE. The use of a behaviour test for the selection of dogs for service and breeding, I: Method of testing and evaluating test results in the adult dog, demands on different kinds of service dogs, sex and breed differences. *Appl Anim Behav Sci.* (1997) 53:279–95. doi: 10.1016/S0168-1591(96)01174-4
8. Wilsson E, Sinn. D. L. Are there differences between behavioral measurement methods? A comparison of the predictive validity of two rating methods in a working dog program. *Appl Anim Behav Sci.* (2012) 141:158–72. doi: 10.1016/j.applanim.2012.08.012
9. Arvelius P, Strandberg E, Fikse WF. The Swedish Armed Forces temperament test gives information on genetic differences among dogs. *J Vet Behav.* (2014) 9:281–9. doi: 10.1016/j.jveb.2014.06.008
10. Goddard ME, Beilharz RG. Genetics of traits which determine the suitability of dogs as guide dogs for the blind. *Appl Anim Ethol.* (1983) 9:299–315. doi: 10.1016/0304-3762(83)90010-X
11. Goddard ME, Beilharz RG. Early prediction of adult behaviour in potential guide dogs. *Appl Anim Behav Sci.* (1986) 15:247–60. doi: 10.1016/0168-1591(86)90095-X
12. Serpell JA, Huse Y. Development and validation of novel method for evaluating behavior and temperament in guide dogs. *Appl Anim Behav Sci.* (2001) 72:347–64. doi: 10.1016/S0168-1591(00)00210-0
13. Pallotti S, Riganelli S, Antonini S, Valbonesi A. Renieri. C. Estimates of non-genetic effects for measures of hunting performance in short-haired and rough-haired Italian hound. *Italian J Anim Sci.* (2020) 19:8526. doi: 10.1080/1828051X.2020.1748526
14. Brady K, Cracknell N, Zulch H, Mills DS. A systematic review of the reliability and validity of behavioural tests used to assess behavioural characteristics important in working dogs. *Front Vet Sci.* (2018) *5:103*. doi: 10.3389/fvets.2018.00103
15. Gosling SD, Rentfrow PJ, Swann WB. A very brief measure of the Big-Five personality domains. *J Res Personal.* (2003) 37:504–28. doi: 10.1016/S0092-6566(03)00046-1
16. Svartberg K, Tapper I, Temrin H, Radesäter T, Thorman S. Consistency of personality traits in dogs. *Anim Behav.* (2005) 69:283–91. doi: 10.1016/j.anbehav.2004.04.011
17. Fratkin JL, Sinn DL, Thomas S, Hilliard S, Olson Z, Gosling SD. Do you see what I see?: Can non-experts with minimal training reproduce expert ratings in behavioral assessments of working dogs? *Behav Proc.* (2015) 110:105–16. doi: 10.1016/j.beproc.2014.09.028
18. Cristescu R, Foley E, Markula A, Jackson G, Jones D, Frère C. Accuracy and efficiency of detection dogs: a powerful new tool for koala conservation and management. *Sci Rep.* (2015) 5:8349. doi: 10.1038/srep08349
19. Porritt F, Shapiro M, Waggoner P, Mitchell E, Thomson T, Nicklin S, et al. Performance decline by search dogs in repetitive tasks, mitigation strategies. *Appl Anim Behav Sci.* (2015) 166:112–22. doi: 10.1016/j.applanim.2015.02.013
20. Rooney NJ, Gaines SA, Bradshaw JWS, Penman S. Validation of a method for assessing the ability of trainee specialist search dogs. *Appl Anim Behav Sci.* (2007) 103:90–104. doi: 10.1016/j.applanim.2006.03.016
21. Haverbeke A, Laporte B, Depiereux D, Giffroy JM, Diederich C. Training methods of military dog handlers and their effects on the team's performances. *Appl Anim Behav Sci.* (2008) 113:110–22. doi: 10.1016/j.applanim.2007.11.010
22. Haverbeke A, Messaoudi F, Depiereux E, Stevens M, Giffroy JM, Diederich C. Efficiency of working dogs undergoing a new human familiarization and training program. *J Vet Behav.* (2010) 5:112–9. doi: 10.1016/j.jveb.2009.08.008
23. Sinn DL, Gosling SD, Hilliard S. Personality and performance in military working dogs: reliability and predictive validity of behavioral tests. *Appl Anim Behav Sci.* (2010) 127:51–65. doi: 10.1016/j.applanim.2010.08.007
24. Krosnick JA. Survey research. *Annu Rev Psychol.* (1999) 50:537–67. doi: 10.1146/annurev.psych.50.1.537
25. Melchers KG, Lienhardt N, Von Aarburg M, Kleinmann M. Is more structure really better? A comparison of frame-of-reference training and descriptively anchored rating scales to improve interviewers' rating quality. *Person Psychol.* (2011) 64:53–87. doi: 10.1111/j.1744-6570.2010.01202.x
26. Troisi C, Mills D, Wilkinson A, Zulch H. Behavioral and cognitive factors that affect the success of scent detection dogs. *Compar Cogn Soci.* (2019) 14:51–76. doi: 10.3819/CCBR.2019.140007
27. Rooney NJ, Bradshaw JWS. Breed and sex differences in the behaviour of specialist search dogs - a questionnaire survey of trainers and handlers. *Anim Behav Sci.* (2004) 86:123–35. doi: 10.1016/j.applanim.2003.12.007
28. Rooney NJ, Bradshaw JWS, Almey H. Attributes of specialist search dogs - A questionnaire survey of UK dog handlers and trainers. *J Forensic Sci.* (2004) 49:300–6. doi: 10.1520/JFS2003048

29. Rocznik D, Sinn DL, Thomas S, Gosling SD. Criterion analysis and content validity for standardized behavioral tests in a detector-dog breeding program. *J Fore Sci.* (2015) 60:S213–21. doi: 10.1111/1556-4029.12626
30. Roch SG, Woehr DJ, Mishra V, Kieszczynska U. Rater training revisited: an updated meta-analytic review of frame-of-reference training. *J Occupat Organizational Psychol.* (2012) 85:370–95. doi: 10.1111/j.2044-8325.2011.02045.x
31. Fradenburg LA, Harrison RJ, Baer DM. The effect of some environmental factors on inter observer agreement. *Res Dev Disabil.* (1995) 16:425–37. doi: 10.1016/0891-4222(95)00028-3
32. Fuller CJ, Bladon BM, Driver AJ, Barr ARS. The intra- and inter-assessor reliability of measurement of functional outcome by lameness scoring in horses. *Vet J.* (2006) 171:281–6. doi: 10.1016/j.tvjl.2004.10.012
33. Preston CC, Colman AM. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychol.* (2000) 104:1–15. doi: 10.1016/S0001-6918(99)00050-5
34. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* (1979) 86:420–8. doi: 10.1037/0033-2909.86.2.420
35. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ.* (2011) 2:53–5. doi: 10.5116/ijme.4dfb.8dfd
36. Taylor KD, Mills DS. The development and assessment of temperament tests for adult companion dogs. *J Vet Behav.* (2006) 1:94–108. doi: 10.1016/j.jveb.2006.09.002
37. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods.* (1996) 1:30–46. doi: 10.1037/1082-989X.1.1.30
38. Spearman C. The proof and measurement of association between two things. *Am J Psychol.* (1987) 100:441–71. doi: 10.2307/1422689
39. Padilla MA, Veprinsky A. Correlation attenuation due to measurement error: A new approach using the bootstrap procedure. *Educ Psychol Meas.* (2012) 72:827–46. doi: 10.1177/0013164412443963
40. Gau JM. The convergent and discriminant validity of procedural justice and police legitimacy: an empirical test of core theoretical propositions. *J Crim Justice.* (2011) 39:489–98. doi: 10.1016/j.jcrimjus.2011.09.004
41. Clark C, Sibbald NJ, Rooney NJ. Search dog handlers show positive bias when scoring their own dog's performance. *Front Vet Sci.* (2020) 7:612. doi: 10.3389/fvets.2020.00612
42. Weiss E, Greenberg G. Service dog selection tests: Effectiveness for dogs from animal shelters. *Appl Anim Behav Sci.* (1997) 53:297–308. doi: 10.1016/S0168-1591(96)01176-8
43. John OP, Robins RW. Determinants of interjudge agreement on personality traits: the big five domains, observability, evaluativeness, and the unique perspective of the self. *J Pers.* (1993) 61:521–51. doi: 10.1111/j.1467-6494.1993.tb00781.x
44. Tami G, Gallagher A. Description of the behaviour of domestic dog (Canis familiaris) by experienced and inexperienced people. *Appl Anim Behav Sci.* (2009) 120:159–69. doi: 10.1016/j.applanim.2009.06.009
45. Mariti C, Gazzano A, Moore JL, Baragli P, Chelli L, Sighieri C. Perception of dogs' stress by their owners. *J Vet Behav.* (2012) 7:213–9. doi: 10.1016/j.jveb.2011.09.004
46. Rooney NJ, Clark C, Casey R. Minimising fear and anxiety in working dogs: a review. *J Vet Behav Clin Appl Res.* (2016) 16:53–64. doi: 10.1016/j.jveb.2016.11.001
47. Loftus BA, Rooney NJ, Casey RA. *Recognising Fear and Anxiety in Dogs.* (2012). Available online at: http://www.bris.ac.uk/vetscience/services/behaviour-clinic/dogbehaviouralsigns/ (accessed March 20, 2020).
48. Clark CCA, Rooney NJ. Does benchmarking of rating scales improve ratings of search performance given by specialist search dog handlers? *Front Vet Sci.* (2021) 8:545398. doi: 10.3389/fvets.2021.545398

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership