Detection dogs towards Predator Free 2050

A post-doctoral project funded by Predator Free 2050 Ltd. and hosted by Manaaki Whenua - Landcare Research

2024

Dr Emma Feenstra
Emmafeenstra@gmail.com

CONTENTS

1. INTRODUCTION

Detection dogs are an invaluable tool for population monitoring. Dog teams are useful for surveying over large areas and detecting rare animals, which makes them particularly suitable for monitoring during eradication projects, where locating every individual is important (Becker et al., 2017; Bennett et al., 2022; Cowan, 1992; Jamieson et al., 2021; Long et al., 2007).

Possum detection dogs are used to detect live Brushtail possums (*Trichosurus vulpecula*) and their scat, providing critical data on possum presence and distribution. Possum dogs have been integral to the success of large-scale pest control programmes, for example on Kapiti Island, where they were considered essential in declaring eradication success (Cowan, 1992; Sherley, 1992). Scat detection is particularly useful in the latter stages of eradication projects when possum densities are low. Detection becomes increasingly difficult as density drops (Hurt et al., 2014), and dogs can provide higher detection rates than other monitoring methods (Harrison, 2006).

Dog teams are a powerful monitoring tool, however, dog survey data can be difficult to translate into robust conclusions, is vulnerable to subjective interpretation, and difficult to standardise (Johnen et al., 2017). There are many variables that impact the detection ability of dogs and subsequent survey results that need to be better understood. Dogs are also underutilised within population monitoring projects, and their utility is worth developing further.

This project aimed to help clarify some of the variables involved in possum scat surveys using detection dogs. Specifically, to estimate the persistence of possum scat in the environment (which is available to be detected by dog teams), whether there are any scat characteristics that can assist with estimates of scat age, and the effect that age of scat has on dogs' ability to detect it. The findings of this report contribute to the understanding of how scat surveys can enhance targeted pest control efforts, ultimately improving the efficiency of possum management operations.

The project spanned the period from May 2023 to June 2024. The dog trials and scat survey data were in collaboration with private contractors from Hoegh Hunting in Taranaki.

2. SCAT SURVEY SURVEILLANCE

Possums can be difficult to detect even when known to be present (Breedt & King, 2022). However, an individual can leave evidence of its presence that offers increased detection probability. Possum scats are more likely to be detected than an individual as they are present over a wider range at any one time and can be detected efficiently using a scat dog. Increasing the probability of detection is particularly important for low density populations and where declaring eradication is a goal.

The presence or absence of scat in an area can be used as an indicator of possum presence and distribution (Cowan, 1992; Hickling & Thomas, 1990). Possums at lower densities are more likely to be aggregated in small, fragmented groups than distributed evenly across the landscape (Sweetapple & Nugent, 2009; Sweetapple & Nugent, 2020), particularly in agricultural land, as possums generally avoid crossing open farmland (Veale & Etherington, 2023). In these contexts, targeting control based on the location of detections is a more cost effective approach than blanket control area wide (Sweetapple & Nugent, 2020).

Scat surveys are rapid; only a single visit is required to map the locations of scat. This means that control can be implemented quickly, whereas other monitoring toolsneed to be left in place for a period of time before being collected and analysed for detections (Sweetapple & Nugent, 2020). Dog teams also cover an area more thoroughly than stationary monitoring tools such as traps or cameras (Glen et al., 2016), and have detected scat in areas where camera traps deployed over long periods have failed to detect the target species (Jamieson et al., 2021; Thomas et al., 2020).

The efficiency and utility of scat surveys can be increased by applying a grid system to the landscape (Thompson et al., 2012; Wasser et al., 2012). Establishing a spatial grid over an area and searching the grid cells for the presence of possums can reduce the effort associated with determining the distribution of individual animals. Cells where possums are deemed present can be targeted for control, which reduces the costs and effort of control by ensuring it is only applied where it is most needed. This type of targeted control is an effective strategy for ground-based management of large-scale, low density possum populations (Sweetapple & Nugent, 2020).

Grid cells over the landscape also provide a metric for reporting throughout a control operation (Bengsen et al., 2014). The number of cells where animals are present before and after control, and the number of cells over time where possums are not detected, can be used to document and report progress. Anderson et al. (2013) used a grid as a novel approach for the surveillance of bovine tuberculosis in possums in New Zealand, estimating the probability of freedom from disease per grid cell based on post-mortems of animals caught in leg-hold traps, providing landscape scale estimates for the presence of disease.

The presence and amount of possum scat as a metric for possum abundance has been used to report on populations for eradication projects (Cowan, 1992), pre-and post-control operations (Hickling & Thomas, 1990), and distribution on public conservation land (Gormley et al., 2015). Hickling and Thomas (1990) recommended that possum pellet counts be used to audit possum control, where the reduction in possum numbers is calculated from the reduction in mean pellets per plot.

2.1. Objectives

In Taranaki, scat dog teams are used as a surveillance and monitoring tool when possum populations are at lower densities (after live possum dogs and trapping have reduced abundance). The locations of scats are used to target further control, which has anecdotally been found to increase trapping success. Scat surveys and targeted trapping can occur repeatedly per area until there are no scat detections. This system of ‘smart control’ means that less traps are used, and the traps strategically target areas where possums have been detected. The system is flexible and adaptive and avoids blanket control over an area.

The objective was to provide an example of ‘smart control’ whereby scat survey data reduces the area where control is required.

2.2. Methods

2.2.1. Study area

The Taranaki Regional Council is committed to reducing the possum population to zero in the ‘Zero Possum Block’, a 10,270 ha area in the northwest of the Taranaki Region, as part of the Predator Free 2050 initiative. The area is dominated by pastureland and interspersed by native and pine forest patches along rivers. It is bounded to the north-south by the Tasman Sea and Mt. Taranaki National Park, and east-west by the Oākura and Hangatāhua rivers. Scat survey data was used from part of the ‘Zero Block Extension’, the western portion of the total block, which is a 3,311 ha area bounded to the east by the Timaru stream. The rivers were used to define block boundaries due to their potential to act as semi-porous barriers to possum dispersal.

2.2.2. Scat surveys

Scat survey data was provided by Hoegh Hunting Ltd., a private contracting company specialising in possum detection dogs (live possum and scat dogs) from the period May 1st 2023 to May 1st 2024. Scat surveys took place over the Zero Extension Block at different time intervals throughout the study period, as Hoegh Hunting worked in conjunction with Taranaki Regional Council to progressively eliminate possums from the area.

Scat dog teams consisted of one dog and handler. During a scat survey the dog was usually able to roam freely searching for possum scat, or was on a leash when necessary, and the handler would aim for reasonable coverage over the area being searched (usually a private farm) and target likely areas of possum habitation. If the dog detected scat it would indicate by lying down and pointing towards it with its nose, whereby the handler would approach and on confirmation the indication was correct, reward the dog with a toy and short session of play. The location of the scat was recorded on handheld Garmin GPS, and all scat detected was collected and removed. Once scat had been detected in a distinct area (e.g., a bush fragment, hedge, scrub patch), the handler usually encouraged the dog to move on to the next area, as the purpose was to map distribution, not count or record every scat present.

2.2.3. Data processing

To provide an example of the use of scat surveys to target subsequent possum control, all the scat survey tracks and waypoints were combined from the study period to represent the areas covered by scat dog teams and the locations of scat detections. These were taken from GPX files and reprojected to NZGD 2000 New Zealand Transverse Mercator. Data processing was conducted in ArcGIS Pro 3.2.1, using the associated Python environment (version 3.9.18). Grid cells of 8 ha were created, which represented the lower end of recent estimates of local possum home range size (O'Malley, 2023). Using estimates of possum home range size for grid cell size can improve the sensitivity of grid cells for detection (Anderson et al., 2013), and should be applied based on local estimates, and at the current status of the population, as home range size can change as density changes (Sweetapple & Nugent, 2020). Different symbology was then applied to the cells to represent the presence or absence of scat survey tracks, scat waypoints, and the amount of scat found.

Opportunistically, we attempted to explore the relationships between the presence/amount of scat and success of traps. We also explored whether the decrease in possum scat presence could be used as a metric for the success of possum control.

2.3. Results

The study area grid had 427 cells of 8 ha. The proportion of cells with at least one scat survey at the time of this report was 60% (256 cells). Of the cells that had been surveyed by a scat dog team, 30% (76 cells) had scat detections (Figure 1).

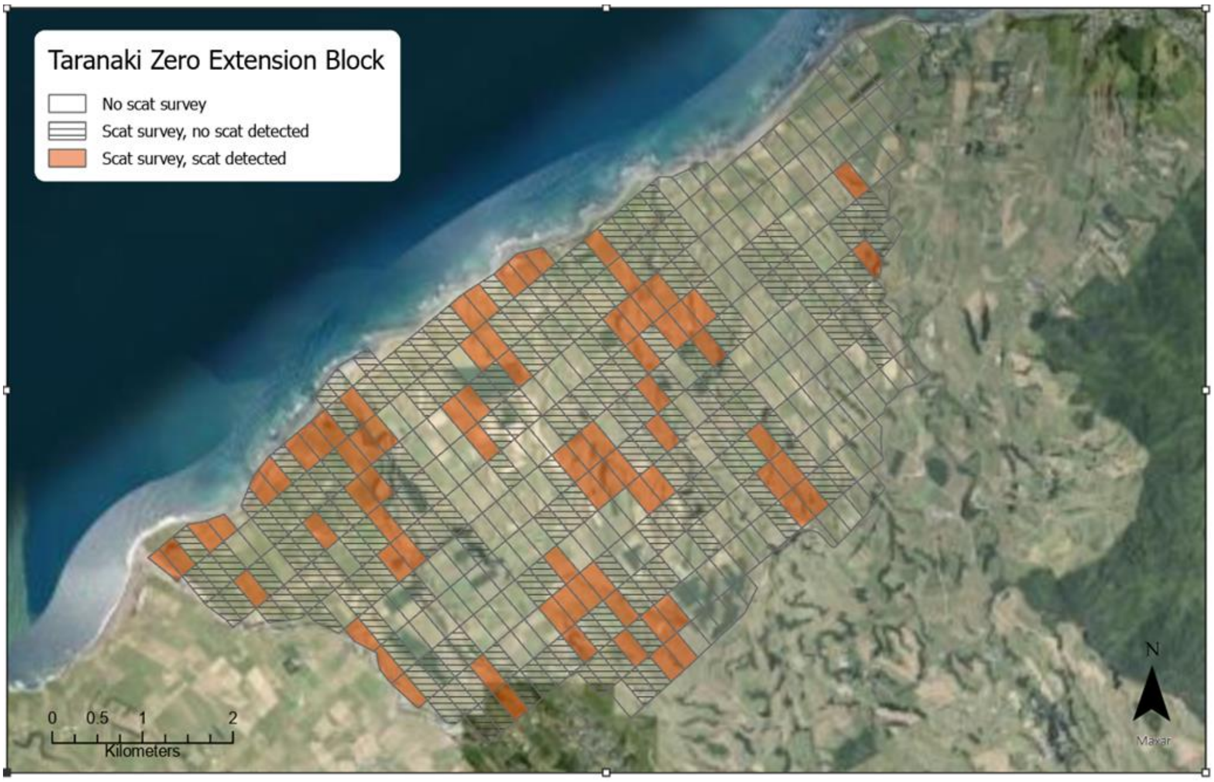


Figure 1. The study area grid on the Taranaki Zero Extension Block with cells 8 ha in size. Blank cells had no scat survey, lined cells had at least one scat survey, and orange cells are where scat was detected.

Calculating the total area of the cells that had at least one scat survey, 1999 ha were surveyed, and the total area of the cells that had at least one scat detection was 589 ha.

The number of possum scat detections per grid cell showed differences across the cells (Figure 2).

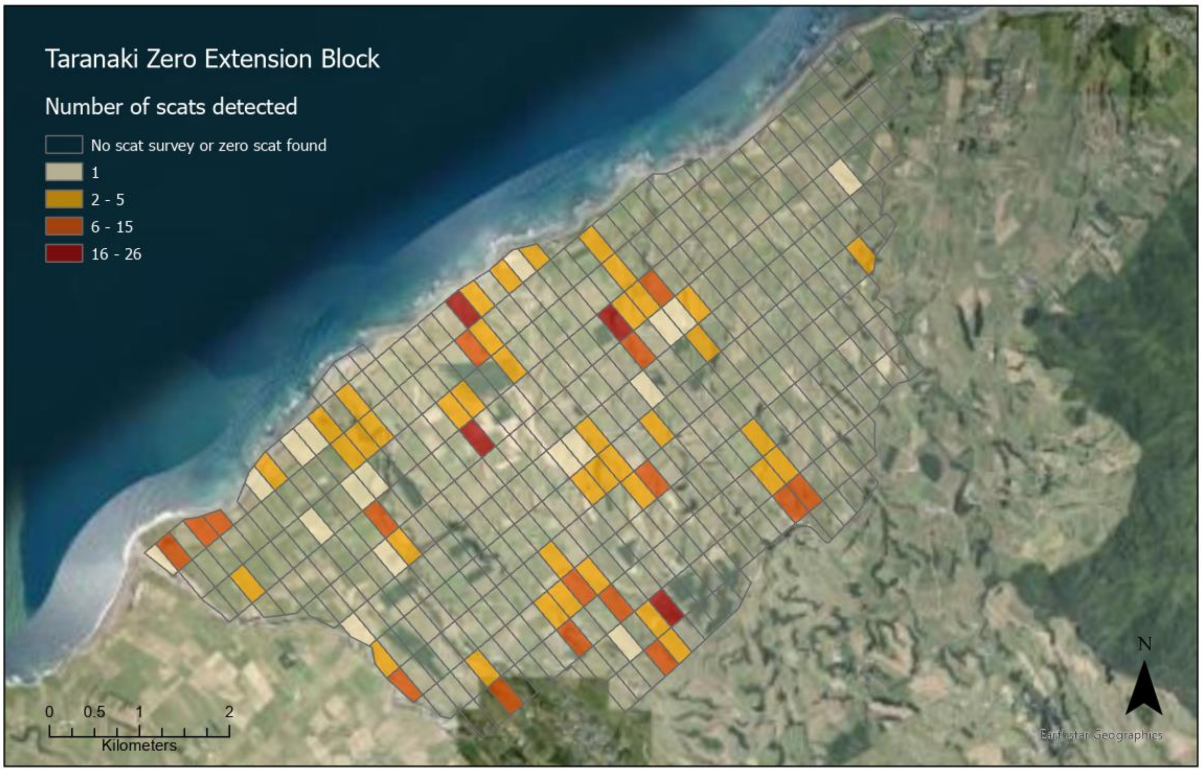


Figure 2. The study area grid over the Taranaki Zero Extension Block with cells 8 ha in size. Blank surveys show cells where there was no scat survey or no scat was found. Coloured cells represent the number of scat detections per cell in a gradient of lighter cells (less scat) to darker cells (more scat).

We were unable to determine any relationship between the scat survey data and trapping success. The predominant issue was that the data was not structured for this purpose. The timing of scat surveys and trapping differed across the landscape, and the number, placement, and trap nights of leg-hold trapping was not consistent. The distance between traps was often small, so that trap success was unlikely to be independent. It was not always clear which traps had been placed based on scat locations, and we could not identify a suitable control site where trap numbers had been reduced but not based on scat locations (as higher success per trap could also be a result of reducing trap density).

Future studies exploring these relationships would benefit from setting up a purposeful, standardised trial. One way of doing this would be to establish two sites in an area with a reasonably consistent possum population. Trapping should be set up in a consistent manner in both sites and trap rates recorded. Scat dog teams should then survey the sites, recording the locations of scat found. At the treatment site, traps should be moved to locations where scat was detected, causing a decrease in the number of traps used. Traps at the control site should also be reduced by randomly removing traps, then trap success in each area could be compared.

2.4. Discussion

The results confirm that possum scat surveys using dog teams can be used to effectively reduce the area to target with control. The benefit is that control can be more efficient and cost effective by being applied to specific areas where possum scat has been detected as opposed to blanket control over a wider area. Dog teams are particularly effective at allowing rapid detection and response.

Grid cells provide an effective spatial tool for the application of survey effort and subsequent control. As individual cells are searched, they can be classified with different risk factors depending on scat detections as well as the amount of optimal habitat, previous control, distance to project area boundaries, and risk of reinvasion. These rankings can be used to rapidly allocate resources per cell after a scat survey.

Searching grid cells can improve the efficiency of monitoring over large areas, as once a detection has been obtained, the dog team can move on to the next cell (Glen et al., 2016). However, when densities are very low or there are no detections in a cell, methods need to be adjusted so that declaring a cell free of the target is reliable. Search intensity within cells can be increased over time to improve confidence that the last animals have been detected, for example, by using a structured grid search (Gsell et al., 2010). Other monitoring methods can be combined with scat surveys to increase the probability of detection per cell (Gormley et al., 2011). Multiple surveys and using novel transect lines each visit can also increase the probability of detection (Mackenzie & Royle, 2005; Wasser et al., 2012). The size of cells can be reduced to smaller than the average home range size to ensure an opportunity for all possums to be detected (Anderson et al., 2013). Increasing the intensity of monitoring in this targeted way as density declines is still a cost-effective strategy (Sweetapple & Nugent, 2020). By comparison, other monitoring tools must be deployed for longer periods to increase their chance of detection (Comer et al., 2018; Glen et al., 2023).

The application of cells to the landscape can provide a simple metric for reporting as shown here. Additionally, the presence or absence of detections per cell can also be used in occupancy models, which provide a robust statistical measure of changes in occupancy (the proportion of sites in an area that are occupied) over time and estimate detection probabilities (MacKenzie et al., 2003). Gormley et al. (2015) used scat surveys and occupancy modelling to estimate the proportion of sites on public conservation land where possums were present. Changes in site occupancy can be used to measure the effectiveness of control operations (Comer et al., 2018), and occupancy models can also provide estimates of extinction and recolonisation rates (Bengsen, 2014), and persistence (MacKenzie et al., 2003; Robley et al., 2014).

As research regarding the relationship between the presence and amount of scat and the presence and abundance of possums progresses, e.g., Nugent et al. (1997), this could be incorporated into the intensity of control required per cell. The relative amount of scat in certain areas could also be used a proxy for habitat preference (Smith et al., 2005). This kind of information can then be used to model predictive occupancy over large areas (Gormley et al., 2011; Long et al., 2011), which could negate the need for large, landscape-scale monitoring pre-control.

Scat surveys assume that possums are present where scat is detected (Cowan, 1992; Gormley et al., 2015; Hickling & Thomas, 1990; Sweetapple & Nugent, 2009). This assumption may not always hold true, for example, if scats persist in the landscape after the animal that deposited them has died. Experiments are required to determine how long scats persist in the environment, and how long they remain detectable by dogs. Future research should also test whether traps placed based on scat locations have a higher trap rate than randomly placed traps.

In summary, this simple example shows how the use of scat dog teams to monitor possums at low densities provides opportunities to target control efforts, reducing cost and improving expediency. Cells can be declared possum free as the operation progresses, providing a valuable and motivating reporting metric.

3. SCAT DEGRADATION

Scat detection dogs are used to survey for the scat of target animals in the landscape. The presence and amount of scat can be used to make estimates regarding the population of interest (Barnes & Dunn, 2002; Hickling & Thomas, 1990; Smith et al., 2006). In New Zealand, possum scat dogs are used to determine the presence and distribution of possums. However, the longevity of possum scat in the environment and how long it is available to be detected by dogs has not previously been measured. This is important because scat that is assumed to be from the current population could in fact be from animals that are now removed from the landscape.

How long scat lasts after deposition is variable depending on location and habitat (Harestad & Bunnell, 1987; Hibert et al., 2011; Jung & Kukka, 2016). Some of the factors that impact scat degradation and disappearance include precipitation, temperature, humidity, substrate, season, and invertebrate activity (Barnes et al., 1997; Harestad & Bunnell, 1987; Jachmann & Bell, 2008; Reed et al., 2011). Not knowing how long pellets last can bias population estimates that are based on pellet presence (Campbell et al., 2004; Gormley et al., 2015).

Further compounding this is the difficulty in aging scat detected in the field. The value of being able to estimate the age of scat in the field has been recognised in the monitoring protocols for different species, for example, ungulates (Hibert et al., 2011), San Joaquin kit foxes (*Vulpes macrotis mutica*) (Smith et al., 2003), and koala (*Phascolarctos cinereus*) (Sullivan et al., 2002). Knowing the age of scat allows estimates of when an animal was last present. Additionally, reliably aging scat and incorporating it into fecal standing crop methods (scat counts that also include rates of degradation) can enable estimates of population abundance (Sullivan et al., 2002).

The age of scat detected in the field is commonly estimated by visual inspection of scat characteristics (Brown et al., 2015; Harestad & Bunnell, 1987; Paton et al., 2021; White, 2008; Wiles, 1980). However, scat can appear recently deposited when in fact it has been present for some time. In a degradation trial for feral cat (*Felis catus*) scat in Western Australia, 60% of scat present at day 29 appeared ‘unchanged’ from day 1 (Baker et al., 2021). Elk (*Cervus canadensis*) pellets appeared “relatively newly deposited” two years after deposition in boreal forest in northwestern Canada (Jung & Kukka, 2016). Furthermore, the degree of degradation of pellets can be variable (Harestad & Bunnell, 1987), and characteristics can fluctuate by season (Sullivan et al., 2002).

While field experience interpreting the age of scat can be valuable, there have been no formal evaluations of how possum scat characteristics change with age, and if any of the characteristics could be reliably used to estimate the age of scat

3.1. Objectives

The purpose of this trial was to assess the seasonal persistence of possum scat in different environments to better estimate its longevity, and to determine if any characteristics could assist in estimating the age of scat found in the field.

3.2. Methods

There were two scat degradation trials at each location, in winter (August 2023) and summer (February 2024), except for Rakiura where there was no summer trial. Within seasons, the trials began within 16 days of each other and continued until there was no scat left or until the conclusion of the trial at 50 days.

3.2.1. Location

Six locations were used for scat degradation trials to represent different environments and climates. These were Auckland, Mahia Peninsula, Taranaki, Banks Peninsula, Dunedin and Rakiura. The locations were also areas in which large-scale possum control projects were being undertaken. In three locations (Mahia, Dunedin, Taranaki), possum scat dogs were being used to monitor possums, and in one location (Banks Peninsula), scat dogs were intended for use in the future. Unfortunately, the data from Mahia was unreliable and had to be removed from further analysis.

At each location two sites were selected to represent different habitats, one 'open' and one 'vegetated'. Sites were within 250-m of each other. Open sites were exposed to the elements, at least 10 m from any structure or significant vegetation, while vegetated sites were within vegetation common to the local area (e.g., manuka scrub, native bush, macrocarpas). All sites were chosen as representative of the wider area in which possums could be found locally.

3.2.2. Scat collection

Possum scat was obtained locally for each location by leg-hold trappers, except for Auckland where scat was obtained from trapping in Te Puke in the Waikato. During the normal course of trapping operations, trappers collected possum scat that was left on the traps in plastic pottles and this was frozen within 12 hrs of collection. Scat on the traps was presumed to be less than 24 hrs old as traps were checked daily. Scat was collected from male and female adults and juveniles.

To investigate whether variation in possum diet influenced scat degradation rate, we attempted to compare scat collected from wild possums with scat from captive animals (fed on a controlled diet). However, the condition of the scat from captive possums was inconsistent, potentially due to the air conditioning system in the captive facility, or urine contamination due to the possums being caged overnight. Therefore, captive scat was not included in the analysis.

Scat was removed from the freezer 6-12 hours before the trials began.

3.2.3. Experimental set-up

Each site had three groups of scat less than 1 m apart, with 21 pellets per group. The groups were captive scat, and locally sourced single and clumped scat. Wild possum scat can be found in single pellets or in clumps, and this could affect the rate of degradation. Clumps were generally made of three or more pellets, but occasionally two.

3.2.4. Scat monitoring

Scat checks were carried out on trial days 0, 1, 2, 7, 14, 28, and 49. A limit of 50 days was necessary due to the amount of scat available, and the removal of scat over the course of the trial to evaluate characteristics (see section 3.2.8 below). If there was more scat available and it persisted past day 49 it was checked at intervals but could not be compared across all locations as some did not have extra scat available to measure past the 50 days. This is referred to in the results as 'additional persistence' scat. These scats were monitored until disappearance, which, for the purpose of this study, was when the scat was no longer identifiable as a possum scat (relevant to the handlers of scat detection dogs), or until there was no scat remaining.

Scat checks were carried out at approximately the same time within each location, for example, morning, afternoon, or evening. Most checks were carried out in the afternoon, except for the Auckland summer trial which had morning checks.

During each check, scat from each of the groups (captive, single, clumped) and sites (open, vegetated) was recorded as present or not present.

At each site a rain gauge (150 or 250 ml) was placed at the center of the scat groups, and an i-button (Thermochron DS1921G-F5 -40 C to 85 C) was used to monitor air temperature every 180-mins at each site. The amount of precipitation (mm) was recorded at each scat check interval (see section 3.2.7 below), and this was averaged into a daily amount of rain (mm) since the last scat check. Daily average temperatures per location and site were calculated from the i-buttons.

3.2.5. Scat characteristics

At each scat check, three pellets from each scat group were removed from each site, and the characteristics of the scat were recorded (Table 1). The external characteristics were recorded first, and then each scat was broken in half to record the internal characteristics. More than one selection per characteristic group could be made. After individual pellets were evaluated, they were discarded from the trial. An additional characteristic that was recorded was external colour, and external texture originally included 'rough' and 'smooth'. However, these were deemed extraneous and discarded from further investigation.

Table 1. The characteristics and associated variables recorded from possum scat pellets as they aged.

Characteristic Group	Variables					
External smell	Yes	No				
External smell strength	0 - no smell	1 - faint	2 - medium	3 - strong		
External moisture	Mucous/shine	Dull	Moist	Semi-dry	Dry	Wet
External texture	Soft	Firm	Hard	Crumbly		
Internal colour	Lighter	Consistent	Darker			
Internal moisture	Moist	Semi-dry	Dry	Wet		
Internal texture	Soft	Firm	Hard	Crumbly		

3.2.6. Data analysis

3.2.6.1. Scat persistence

The number of days that scat persisted was estimated by taking the mid-point between the last day it had been present and the first date it was recorded as absent (Barnes & Barnes, 1992). For example, if scat was present on day 14 and absent on day 28, it was assumed to have lasted 22 days.

The data was not normally distributed, so non-parametric Kruskal-Wallis tests were used to examine differences in possum scat persistence. Tests were conducted separately for each location (Auckland, Banks Peninsula, Dunedin, Rakiura, and Taranaki) to compare differences in persistence between sites (open and vegetated) and scat groups (single and clumped). For comparisons involving multiple categorical factors, such as site, season, and scat group, the Kruskal-Wallis test was stratified by these variables.

To evaluate seasonal differences in scat persistence within and between locations, a new variable that represented 'any scat remaining' was used as a composite measure, that reflected the presence of any single or clumped scat. A Kruskal-Wallis test was applied to compare scat persistence across seasons. Where significant differences were identified, Dunn post-hoc tests with Bonferroni correction were conducted to assess pairwise comparisons.

Statistical significance was defined as a p-value less than 0.05.

3.2.6.2. Scat characteristics

Scat characteristics were transformed into binary values, with 1 indicating presence and 0 indicating absence at each scat check, except for smell strength, which was recorded on a scale of 0–3 (0 = no smell, 3 = strong smell). Normalised proportions were calculated for each group of characteristics based on the proportion of samples where a characteristic was present relative to the total number of samples per trial day.

Due to the non-normal distribution of the data, non-parametric tests were used. A series of Mann-Whitney U tests (also known as Wilcoxon rank-sum tests) were conducted to compare the distribution of proportions of each characteristic across trial days. The characteristic groups compared included external smell, moisture, and texture, and internal colour, moisture and texture. These tests were performed separately for each characteristic within combinations of location, season (summer and winter), site (open and vegetated), and scat group (single and clumped).

3.3. Results

3.3.1. Scat persistence

There were no significant differences in the persistence of possum scat between open and vegetated sites within locations. By considering if *any* scat persisted, which included both single and clumped scats, scat persisted for the maximum duration of the study period (50 days) in all locations and seasons, except for Auckland in winter, where scat persisted for a maximum of 22 days (Table 2).

Table 2. The number of days single and clumped scat was present at open and vegetated sites at each location during summer and winter trials.

			Days persisted	
Location	**Season**	**Scat group**	**Open site**	**Veg. site**
Auckland	Summer	Clumped	50	39
		Single	39	39
	Winter	Clumped	22	21
		Single	22	21
Banks Peninsula	Summer	Clumped	50	39
		Single	50	50
	Winter	Clumped	50	50
		Single	50	39
Dunedin	Summer	Clumped	50	50
		Single	50	50
	Winter	Clumped	50	50
		Single	39	50
Taranaki	Summer	Clumped	50	50
		Single	50	50
	Winter	Clumped	50	50
		Single	39	39
Rakiura	Winter	Clumped	50	50
		Single	50	39

There were no significant differences in the persistence of scat between open and vegetated sites within locations and season. There were also no significant differences between scat groups within sites for any of the locations. However, in many instances scat was still present at the last scat check, and so the scale of this trial may not have captured differences that could be present under longer trial conditions (Table 3).

Table 3. The maximum number of days (persistence) of any scat at each location from the open and vegetated sites per season for the trial (> 50 days), and additional persistence of scat in some locations/sites, and seasons. 0 >max indicates scat did not persist past the trial maximum, and grey cells indicate where data was not available.

		Max. trial persistence (days)		Additional persistence (days)	
Location	**Season**	**Open site**	**Veg. site**	**Open site**	**Veg. site**
Auckland	Summer	50	39	55	0 >max
	Winter	22	21	0 >max	0 >max
Banks Peninsula	Summer	50	50	64.5	80
	Winter	50	50	111	
Dunedin	Summer	50	50	60	217*
	Winter	50	50		
Taranaki	Summer	50	50	85.5	60
	Winter	50	50	0 >max	
Rakiura	Winter	50	50		

*Still present at the conclusion of this report

The table above shows that scat was recorded as lasting far past the 50 days of the trial in some places, and it is likely that had more scat been available for the trials, it would have persisted past 50 days in additional cases.

There was a significant difference in scat persistence between summer and winter within Auckland (Kruskal Wallace test: χ^2 (1) = 23.05, $p < 0.001$), with winter showing significantly shorter scat persistence compared to summer. There was no seasonal significance in persistence at any of the other locations.

Scat persistence (using any scat remaining) did not differ significantly between locations in the summer (Kruskal Wallace test: χ^2 (3) = 1.91, $p = 0.591$). However, for winter, there was a significant difference in scat persistence between locations (χ^2 (4) = 36.5, $p < 0.001$).

Post-hoc pairwise comparisons using Dunn's test with Bonferroni correction revealed that Auckland had significantly shorter scat persistence compared to Banks Peninsula ($Z = -5.40$, $p < 0.001$), Dunedin ($Z = -5.40$, $p < 0.001$), and Taranaki ($Z = -5.40$, $p < 0.001$) in winter. No significant differences were found between Banks Peninsula, Dunedin, and Taranaki (all $p > 0.05$). These results suggest that Auckland experienced significantly lower scat persistence compared to the other locations during Winter.

3.3.2. Weather

Average daily temperatures were higher for Auckland in both summer and winter for the duration of the trial (Figure 3).

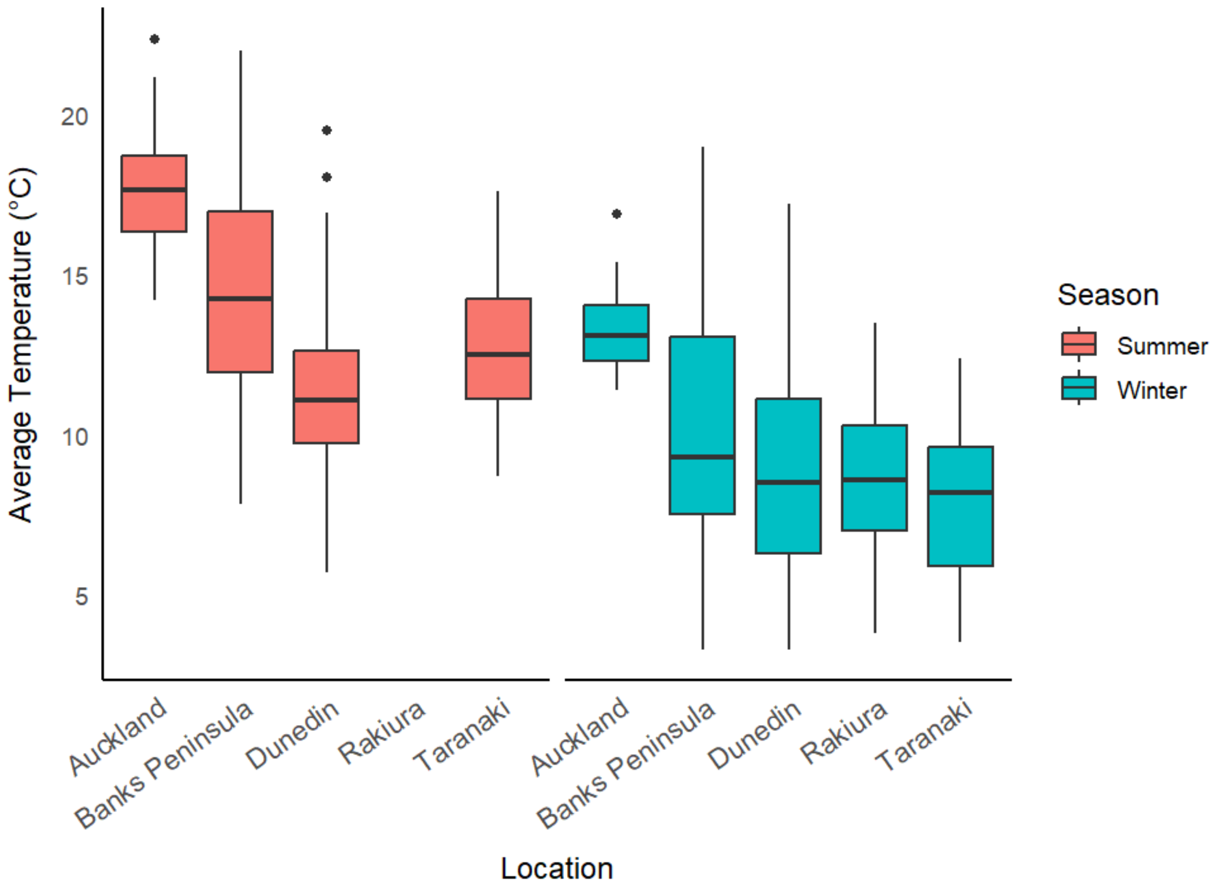


Figure 3. Box plot showing the average daily temperatures for summer and winter in each location over the duration of the trial. Each box represents the interquartile range (IQR) with the central line indicating the median temperature. The whiskers extend to 1.5x the IQR, and dots represent outliers.

The average daily rainfall was highest in Taranaki in both summer and winter (Figure 4) and was slightly higher on average in Auckland during summer than in winter.

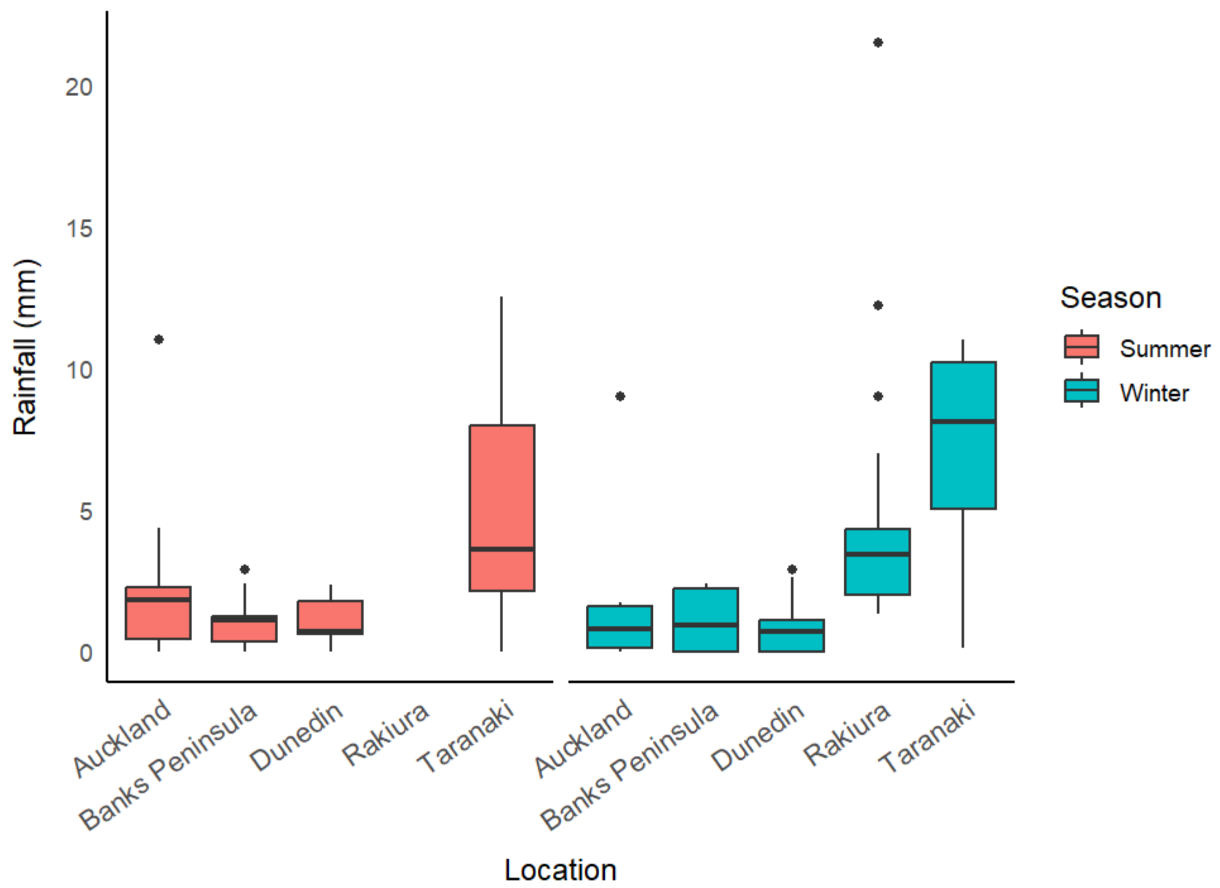


Figure 4. Boxplot showing the average daily rainfall in each location in summer and winter over the duration of the trial. Each box represents the interquartile range (IQR) with the central line indicating the median temperature. The whiskers extend to 1.5x the IQR, and dots represent outliers.

3.3.3. Scat characteristics

For all characteristics (external smell, moisture, and texture, internal colour, moisture, and texture), no significant differences were found between single and clumped scat across any combination of location, season, and site. Additionally, no significant differences were detected between sites within the same location and season for any characteristic.

Plots for characteristics by locations and seasons are available on request. Given the wide range of environments and locations where possum scat is found, the following plots are from all the data combined, and so can be considered as some widely applied guidelines regardless of season, type of site, and location of scat.

There were two characteristics identified that could be useful for indicating possum scat is 'fresh': smell (presence and strength), and external mucous or shine. Both characteristics were predominantly present on days 0 to 7. Mucous or shine was predominantly recorded as present on day 0 of the trial (Figure 5), which is when scat was 24 hrs old or less, and never past day 7. However, it should be noted that fresh scat does not necessarily have mucous or shine.

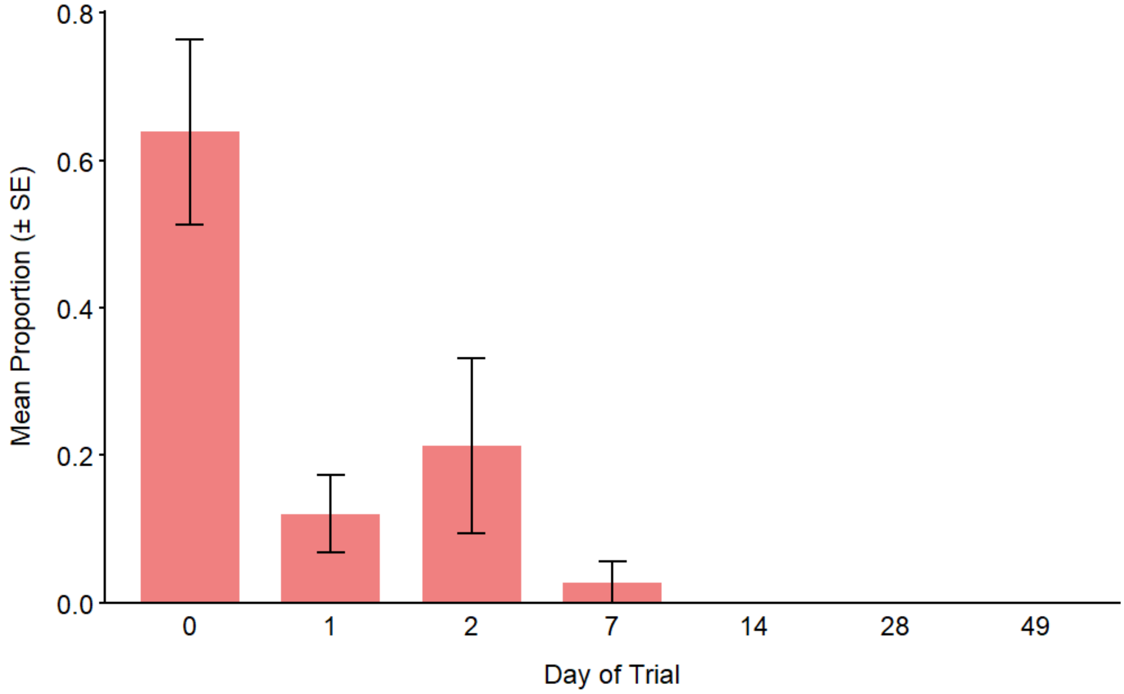


Figure 5. Bar plot showing the mean proportion of scat with external mucous or shine over time (day of the trial) for all locations, sites, and scat groups combined ± standard error.

The presence of external smell could also be useful an indicator of fresh possum scat. By day 14 only ~ 25% of scat had any smell, whereas 100% had smell on days 0,1, and 2 (Figure 6).

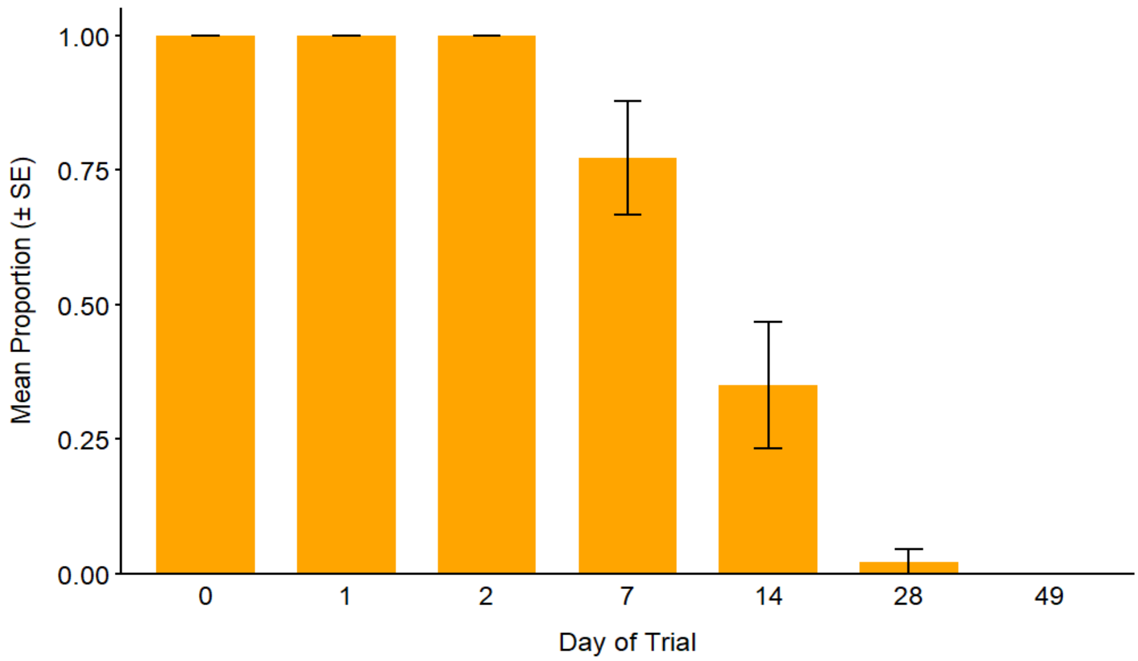


Figure 6. Bar plot showing the mean proportion of scat with external smell per day of the trial for all locations, sites, and scat groups combined ± standard error.

It is worth noting that the presence of smell was judged externally, not when scat was broken in half. On occasion, it was noted that older scat could have increased scent after rain (in Auckland and Taranaki) and could have scent once broken that wasn't present on external examination. The strength of the external smell declined with age of the scat (Figure 7).

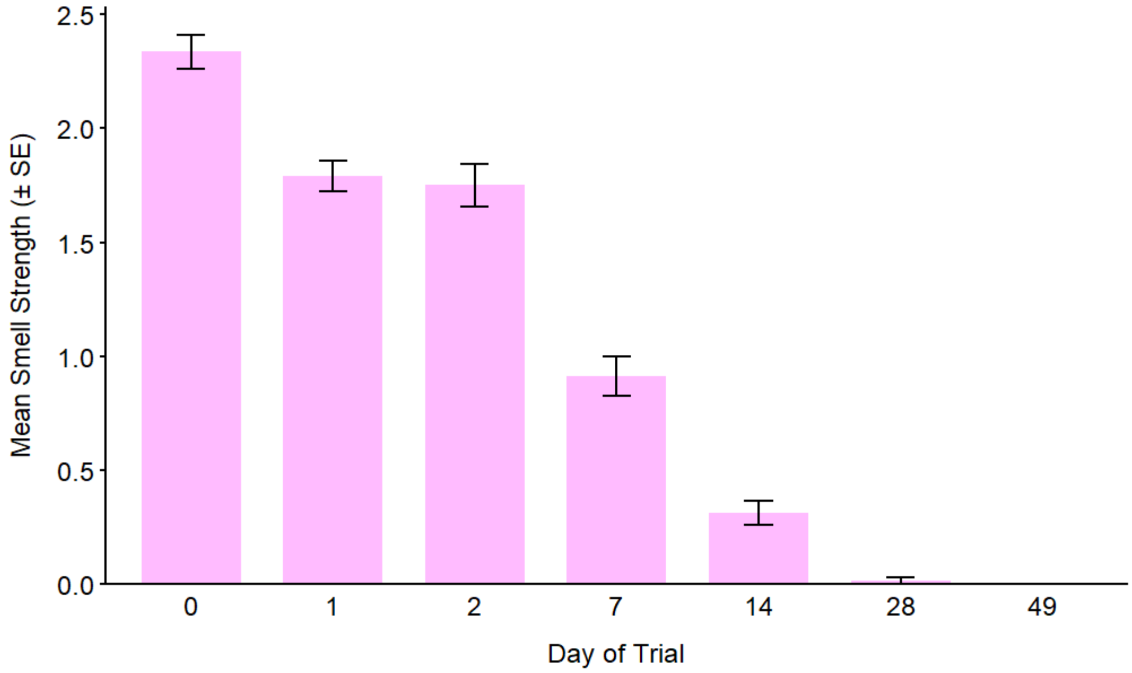


Figure 7. Bar plot showing the mean strength of smell on a scale of 0 (no smell) to 3 (strong smell) from possum scat per day of the trial for all locations, sites, and scat groups combined ± standard error.

By day 7, scat that did have scent had a faint scent, whereas scat on day 0 had the strongest smell. Given the above trends in characteristics, scat that has a strong smell or the presence of mucous or shine can be estimated as most likely within 0-2 days old. Scat with a faint external scent can be estimated as most likely to be within 14 days old.

There were three characteristics that could be useful to help estimate the age of scat as over one week old. These were how crumbly and hard the scat was, and the internal colour when broken in half. Scat was more likely to be older than one week if it was crumbly internally or externally (Figure 8).

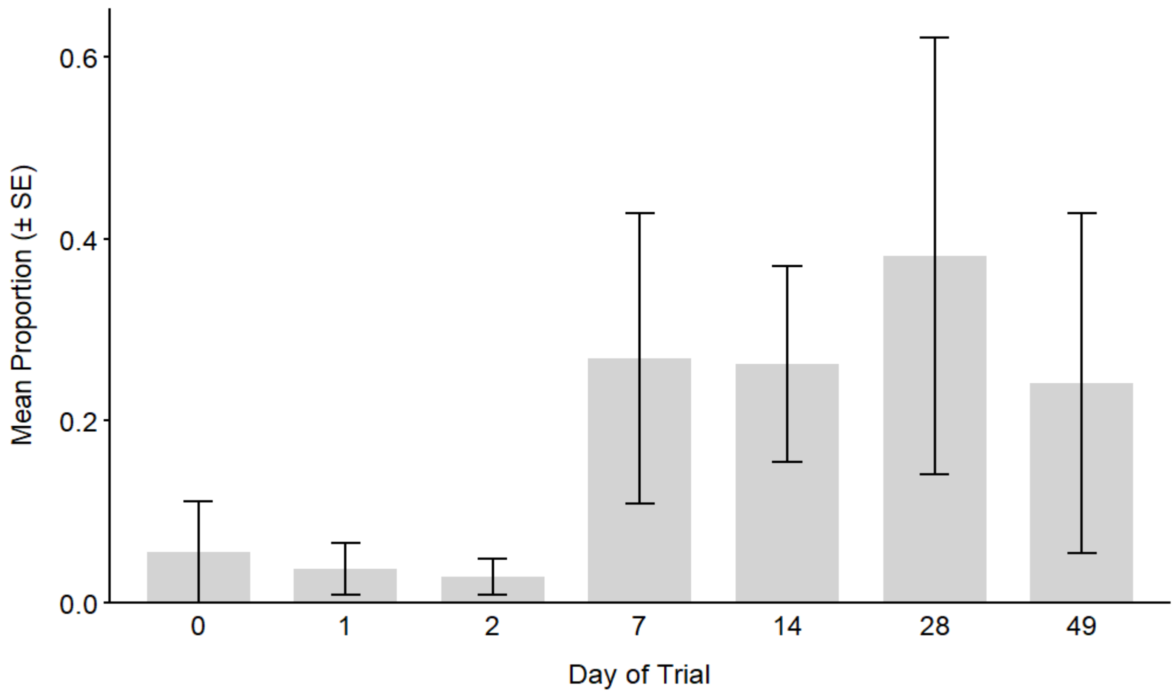


Figure 8. Bar plot showing the mean proportion of possum scat that was crumbly per day of the trial for all locations, sites, and scat groups combined ± standard error.

Scat was also more likely to be older than one week if it was hard (Figure 9), however, scat could also harden very quickly in some conditions (e.g., in Dunedin and Banks Peninsula), and therefore local conditions and recent rain should be considered when using hardness to estimate scat age.

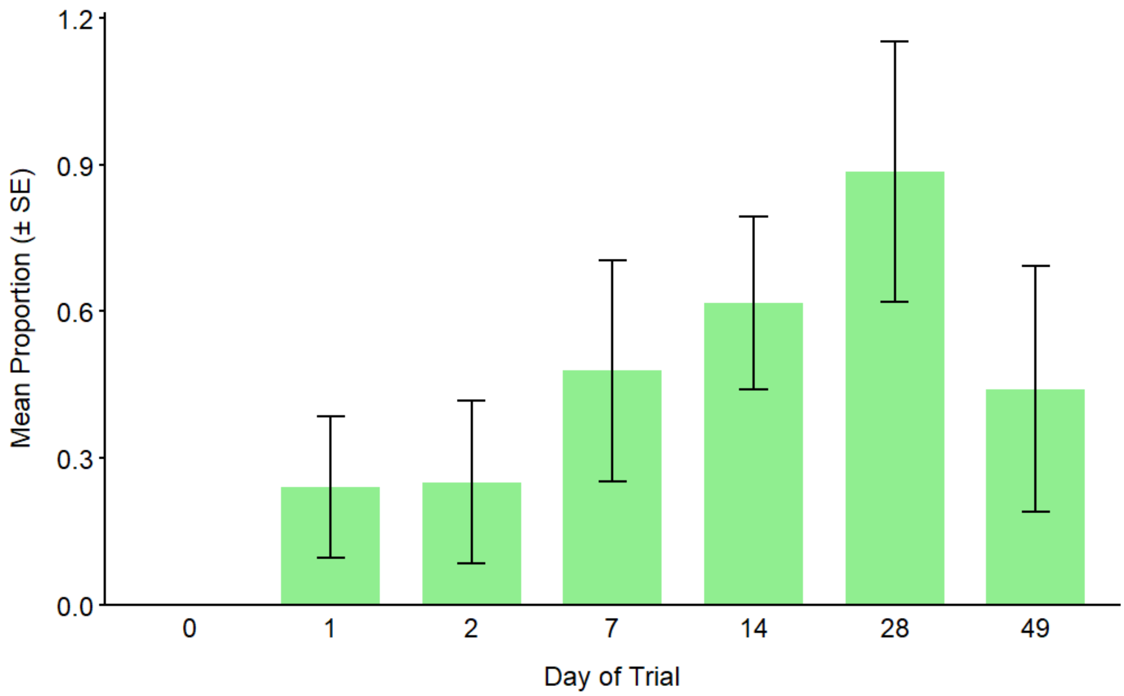


Figure 9. Bar plot showing the mean proportion of possum scat that was hard per day of the trial for all locations, sites, and scat groups combined ± standard error.

One characteristic that was less common but was predominantly found in scat from day 7 onwards, was a darker internal colour when scat was broken in half (Figure 10). This was probably due to moisture changes, with the external layer getting drier as it was exposed to the elements.

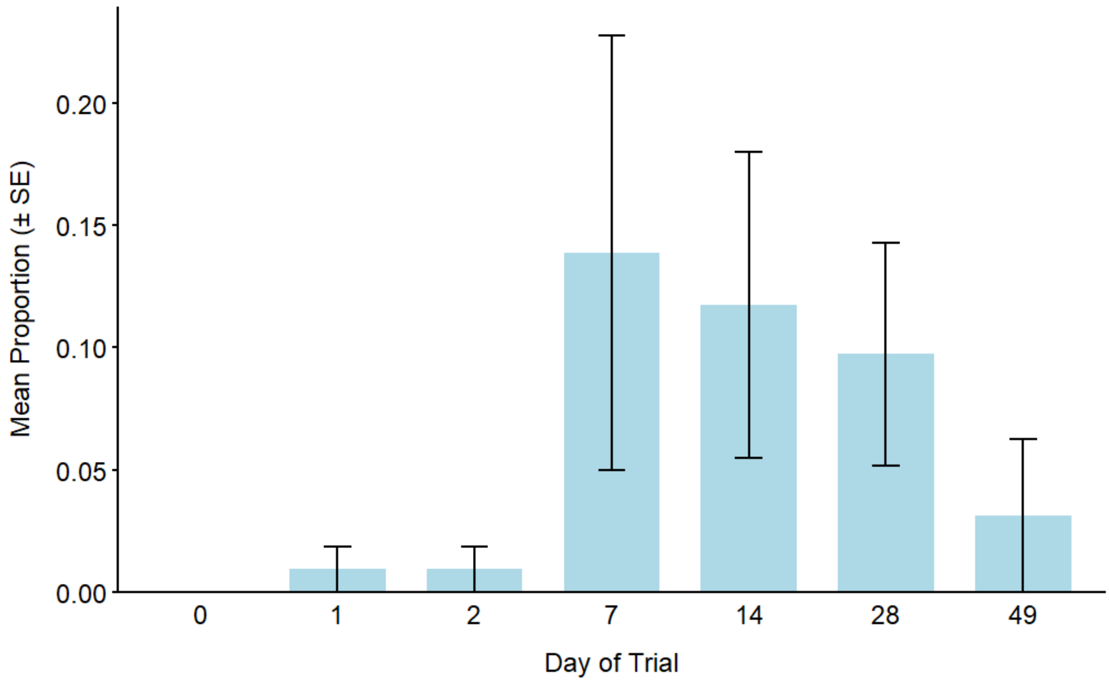


Figure 10. Bar plot showing the mean proportion of possum scat that had a darker internal colour per day of the trial for all locations, sites, and scat groups combined ± standard error.

3.4. Discussion

3.4.1. Scat persistence

The results suggest that within each location, the persistence of scat was not influenced by whether the scat was clumped or in single pellets or if the site was open or vegetated. The lack of significant results implies that scat type and site were not primary drivers of differences in persistence, at least during the conditions and duration of this trial. In many cases, clumped scats fell apart after a short period, resulting in single scats. Although initially being in a clump seemed to keep scat semi-moist for longer, this is a probably reason why there were no significant differences between scat groups.

It is possible that a longer trial period and more replicates of scat group and site (e.g., Jung and Kukka (2016)) would have revealed differences for possum pellets that were not captured in this trial duration. The additional longevity recorded from the 'additional persistence' scats indicates more variation in persistence past 50 days.

Differences in scat persistence between sites (habitats) have been recorded in some cases. Jung and Kukka (2016) found that elk pellets decayed faster in open habitats than forested, 'likely due to exposure to the elements', and macropod pellets disappeared quicker in short, open grass than long, shady grass in Western Australia (Johnson & Jarman, 1987). However, scat does not always deteriorate faster in open environments, black-tailed deer (*Odocoileus hemionus columbianus*) pellets disappeared more rapidly in moist, vegetated sites than dry, bare sites on Vancouver Island (Harestad & Bunnell, 1987). There are not always differences in scat persistence between habitats, Hibert et al. (2011) found no statistical difference in the proportions of decayed ungulate pellets between groups in the sun or shade in Africa.

Seasonal differences were marginally impactful to scat persistence for Auckland. Scat in Auckland persisted for a shorter period in winter than in summer, and for less time in winter than any of the other locations. Different seasons likely capture a variety of nuanced environmental conditions that could be responsible for this difference. Although precipitation is considered an important variable in the persistence of scat (Barnes et al., 1997; Reed et al., 2011; Wallmo et al., 1962; White, 1995), rainfall in Auckland over winter was not as high as in Taranaki, yet the scat in Taranaki persisted longer.

Temperature likely has an impact how long scat lasts, although this if often more of an assumed relationship than a measured one (Brown et al., 2015; Johnson & Jarman, 1987). When maximum and minimum temperatures were measured in relation to the duration of elephant dung piles in the Lope reserve rain forest in Gabon, they did not appear to have as much influence as other variables such as rain, humidity, and diet (fruit content) (White, 1995). Auckland had higher average daily temperatures in winter than the other locations, however, this is unlikely to be the sole reason that scat degraded faster.

An interaction between factors is likely to impact scat persistence in any one context, and different factors can have more or influence. In some cases substrate moisture may play a larger role (Harestad & Bunnell, 1987), while in others, invertebrates have more of an impact (Harestad & Bunnell, 1987; Jachmann & Bell, 2008; Johnson & Jarman, 1987; Wiles, 1980). Invertebrates were primarily responsible for the degradation and disappearance of the scat of 4/5 ungulate species in Africa, with rain secondary (the trial was during the dry season) (Hibert et al., 2011). Furthermore, the timing of different factors can change their level of influence, for example, Johnson and Jarmen (1987) found that rain could eliminate fresh macropod pellets, but had no effect on dry pellets.

In general, the persistence of scat in all locations was longer than expected in both seasons, in most cases lasting > 50 days, and where there was additional scat available, lasting up to 111 days in winter in Banks Peninsula (where only damage to the site caused the loss of remaining scat), and 219 days in Dunedin at the writing of this report. Both locations are known for dry conditions, and the Banks Peninsula open site anecdotally had little invertebrate activity. Johnson and Jarman (1987) found macropod pellets survived well in dry (and cold) conditions, and black-tailed deer pellets remained longer on bare ground, where there was probably less invertebrate activity (Harestad & Bunnell, 1987). The longevity of possum pellets was noted during pre-and post-monitoring of possums throughout a control operation using pellet counts as an index method (Hickling & Thomas, 1990). A small proportion (4.5%) of the pellets present pre-control (in May) survived until the post-poison survey (in August), at Slopedown in Southland, a period of up to 105 days.

Scat can persist for long periods in some conditions. Coyote (*Canis latrans*) scat in California was estimated to persist for up to 31 weeks (217 days) (Kohn et al., 1999), elephant dung can last between 140 to 280 days (Barnes & Barnes, 1992; White, 1995). Black-tailed deer pellets on Vancouver Island lasted for over a year (Harestad & Bunnell, 1987), and elk pellets in semi-arid boreal forest in the Yukon, Canada, were still present two years after deposition (Jung & Kukka, 2016). However, scat can also disappear rapidly in some instances, 16% of moose (*Alces alces*) pellets disappeared within 24 days in the Adrionaks, New York (Kretser et al., 2016), while fresh macropod pellets disappeared with heavy rain in Australia (Johnson & Jarman, 1987), as did fox (*Vulpes vulpes*) scat in Tasmania (Brown et al., 2015).

The duration of time that possum scat can persist in these environments is important to understand when projects are implementing scat dog surveys as a monitoring tool. Generally, after a control operation targeting possums, scat dogs are deployed after a period where it is assumed that scat from the population being targeted would have disappeared, so that scat remaining represents the current population. This study indicates that even at 50 days after fresh scat is deposited, it can remain in the environment, and in some cases for much longer. As such, it would be prudent to only consider fresh scat during surveys as being from the current population. Evaluating the freshness of scat is done subjectively, which can also be problematic. Using scat characteristics to judge the age of scat is addressed in the next section.

3.4.2. Scat characteristics

Although there were no significant differences between sites and scat groups, the distribution of scat characteristics showed some slight differences in patterns between locations and seasons. However, overall, there were some similar trends that could be useful when estimating the age of scat in the field. The external characteristics of smell and mucous/shine indicated fresher scat, while scat that was hard, crumbly, or had a darker internal colour when broken open was more likely to be over one week old.

The presence and strength of odour and presence of patina (shine/luminescence) were also identified as indicators of 'new' (0-28 days old) scat for koala (Sullivan et al., 2002). New pellets were readily identifiable by a strong eucalypt odour, and old pellets (14-29 days) had quickly lost their strong odour and were no longer shiny. Sullivan et al. (2002) found that new pellets had either a strong smell, or a weak external smell with a stronger smell when broken in half.

Using external texture to estimate older scat (i.e., hard and crumbly) was less definitive. Although the proportion of scat that was 'hard' generally increased with age, scat in some locations (Dunedin and

Banks Peninsula) could harden within 24 hours. Hibert et al. (2011) found 'hardness' an unreliable indicator of age, as scats could reach maximum hardness within 1 day. Different factors can also cause scat to appear crumbly when it is fresh, such as diet (roughage can be misinterpreted as crumbliness), and high invertebrate activity. Local conditions should always be considered when estimating the age of scat, and estimates should be interpreted with caution (Jung & Kukka, 2016).

Characteristics indicating age vary in different studies, changes in colour and moisture were the most relevant for aging the scat of San Joaquin kit foxes (Smith et al., 2003), and shade of colour most important for five species of ungulate (Hibert et al., 2011). For foxes in Tasmania, the physical appearance of scat as it aged was so variable that it was not possible to develop an index of scat age (Brown et al., 2015).

The utulisation of scat as an index for population monitoring involves a number of assumptions. Being able to visually age scat allows better estimates of when an animal was last present, and also offers the potential for scat data to be incorporated into population estimation methods that use rates of scat accumulation (and degradation) to estimate abundance (Sullivan et al., 2002).

3.4.3. Conclusion

It is likely that several interacting factors affect the persistence of possum scat, and different influences could have a greater impact depending on habitat. It is possible that a larger sample size and longer trial would have revealed significant differences between possum scat persistence across locations and sites, however, this trial showed that possum scat commonly lasts up to 50 days in some locations, even with high average daily precipitation. This is important to consider for repeat scat surveys using dog teams, where scat could be present from a prior population. The ability to accurately estimate the age of scat in the field could assist scat dog teams to identify only fresh (recent scat) (Smith et al., 2003), though estimates should never be considered conclusive, and local conditions can impact scat characteristics.

4. AGED SCAT TRIAL

Scat surveys are increasingly being used for monitoring populations and making population estimates (Baker et al., 2021; Brown et al., 2015; Kretser et al., 2016; Smith et al., 2005; Wasser et al., 2012), however, one assumption of this method is that scat detections directly relate to the current population of animals. There are limited examples where the age of scat detected by dogs is acknowledged as a potential bias in scat surveys, and little understanding of the limits of a dog's ability to detect scat across different ages.

In studies that trial the success of dogs at detecting scat samples of different ages, there do not appear to be examples where age exceeds the detection limits of dogs. In a controlled trial using fox (*Vulpes vulpes*) scat as the target in Tasmania, dogs successfully indicated on scat that was 126 days old, which was the maximum age used in the trial (Brown et al., 2015). In a controlled field trial where aged scat samples from deer (*Mazama* spp.) were placed along transects in São Paulo State, Brazil, detection was possible up to 'a maximum of 70 days' (Da Silva et al., 2020), though it doesn't appear that samples older than 70 days were present. A trained dog successfully detected spotted-tailed quoll (*Dasyurus maculatus*) scat in a field trial that was 334 days old and 'dried crumbs' (Leigh & Dominick, 2015).

Understanding both the longevity (see Section 3) of possum scat and dogs' ability to detect it at different ages is particularly important when the purpose of scat surveys is to estimate the presence and distribution of the current population (not the past population that may have been removed following pest-control operations). This has been the primary reason for scat surveys for possums (Cowan, 1992; Gormley et al., 2015; Hickling & Thomas, 1990).

Possum scat can persist in the environment for 219 days (see Section 3.3.1). This means that very old possum scat could potentially be detected by dog teams during surveys for possums, which could bias population estimates based on scat presence. However, the presence of scat for long periods does not directly translate into detection by dogs, as many variables influence the probability of detection, including the detection capacity of the dog (Da Silva et al., 2020). An important step in understanding the detection of possum scat by dogs is whether the age of scat impacts the success of detection, and what the potential limit of their ability is when it comes to different aged scat.

4.1. Objectives

The purpose of this trial was to conduct a test in a controlled environment to determine the ability of dogs to detect possum scat of different ages. This trial was considered the first step in understanding the dogs' ability, acknowledging that their performance in a field environment would be affected by a range of variables that are not present in a controlled scenario. Therefore, this trial provides a baseline estimate of the ability of dogs to detect various ages of possum scat.

4.2. Methods

4.2.1. Scat and decoy samples

Possum scat samples were collected from male and female adults during trapping using leg hold traps and were less than 24 hours old at the time of collection. All samples were uniformly aged outside under shelter in Waimauku, Auckland to minimise exposure to the elements and invertebrates. Three scat samples were collected at specified intervals: day 1, 7, 14, 28, 49, 70, 112, 154, 187, and 239, and frozen in individual plastic pottles until the trial.

Non-target rabbit, sheep, and goat scats were collected for use as decoys. The specific age of the decoy scats was unknown but was assumed to be variable from appearance, from visibly fresh (soft, moist with a strong smell) to notably aged (dry, firm, minimal/no scent). Rabbit scats were at least 4 months old as there were no longer rabbits on the property. The decoy animals were present in the regions the dogs commonly worked in, and two out of three dogs had previously worked on goat as a target animal. Six samples each of sheep, goat, and rabbit were used interchangeably to avoid the dogs learning to avoid any one decoy sample. All samples were collected without direct handling, using the pottle as a scoop or gloves. Care was taken to prevent odour contamination by always using clean equipment and plastic pottles were kept separate from each other during storage. Small pieces of the latex gloves were also used as decoys throughout the trial to ensure the dogs were not cueing onto the scent of the gloves.

Samples were presented to the dogs on a base of 1/3 cup of soil. 'Blank' samples consisting only of soil were also used as an experimental control, the idea being that old possum scat may be indistinguishable from soil, but if old scat was present in a container alone, the dog may indicate on it as the 'odd container out', rather than because it recognised it as possum scat. Soil was discarded after each dog's session, so that every session used new soil. The soil was taken from the bottom of a purpose dug hole in an area where possum presence was considered unlikely.

All scat samples were cut in half to increase sample size and immediately refrozen to minimise degradation. This enabled multiple replicates of each age of possum scat so that each trial had 'fresh' samples, therefore avoiding degradation, aging, or contamination issues between trials (Brown et al., 2015).

The possum scats used for pre-trial training were collected from a farm in Norfolk, Taranaki, using the youngest, most inexperienced dog participating in the trial, with the assumption being that scats detected by this dog would be easily detected by the other two dogs. The scats collected appeared fresh and no more than 3 days old.

4.2.2. Trial Setup

The trial setup comprised six concrete cinder blocks, 19 cm high, arranged in a circular pattern around a central drum, labelled A to F (Figure 11).

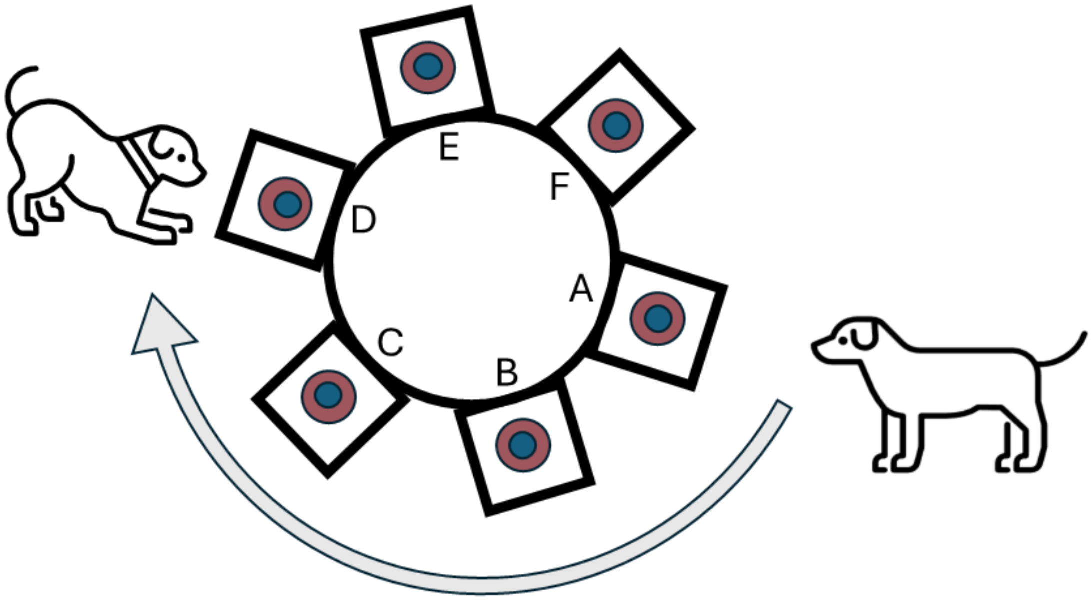


Figure 11. A diagram showing the trial setup, with six blocks around a central drum labelled A to F containing the scat samples. The dog would search the circle of samples and indicate on one of the positions.

Within each cinder block were 8 cm round plastic containers with metal mesh lids, each containing soil, also labelled A to F. Within these were ~ 2.5 cm, round plastic lids which contained either the possum or decoy samples, or a small piece of soil as a visual proxy of a possum scat to ensure the dogs could not visually cue on possum scat. Each possum scat of a particular age and decoy scats were kept in their own plastic pottle and the lids of these pottles were used for the trial. Only the lids with the scats or decoys were moved between containers, which were moved between cinder blocks.

4.2.3. Trial Procedure

All scats were removed from the freezer the night before the trial and thawed to room temperature within their individual pottles. The same scat samples were used for all three dogs within a trial day, but different scats were used for each trial (half the scat was used in Trial 1, and the second half in Trial 2).

The trial followed four stages, 1) Pre-training, 2) Trial 1: aged scat and blank (soil) decoys, 3) Decoy-training, and 4) Trial 2: aged scat and decoy (species and glove) samples. Trial 1 and 2 used possum scat aged 1 - 239 days, the primary difference being that Trial 2 had sheep, goat, and rabbit scat, and gloves as decoys, while Trial 1 had only soil (blanks) in the alternate sample containers.

Both trials and the training sessions took place outdoors in a grass yard. Three dogs completed Trial 1, while two completed Trial 2.

4.2.3.1. *Pre training*

The three dogs were all familiar with possum scat as a target prior to this study, with varying levels of working experience (Table 4). The success of dogs at detecting odour can be highly variable from dog to dog, due to experience, personality, and temperament (Rust et al., 2018), and having three dogs provided a wider range of estimates of ability. The trial set up was novel and so pre-training was undertaken to familiarise them with this, using fresh possum scat samples that were not used during the trial.

Table 4. Gender, age, breed, and work history of the three dogs in the trial: Beau, Mose, and Peggy.

Dog	Gender	Age	Breed	Years in work	Work experience	Possum scat experience
Beau	M	2.3 yrs	Vizsla	< 1	Goat, possum	< 6 months
Mose	M	3 yrs	GSP	2.5	Possum	8 months
Peggy	F	5.10 yrs	GSP X vizsla	5	Goat, possum	4 years

Each dog was exposed to the trial set up of six sample locations in a circular formation and was taught to expect one target odour amongst the six with the others blank. When dogs encountered a target odour, their response was to indicate (already a learned behaviour pre-trial) by pointing and laying down, which if correct was rewarded with a toy thrown next to the target for the dog to retrieve. Each event in which the dog checked the randomly placed odour samples in the circle was termed a 'run'. A run started when the handler asked the dog to 'seek' and ended with the dog performing an indication (correctly or incorrectly), or not indicating to any sample, in which case it was called back to the handler. Pre-training was considered complete when the dogs confidently approached and searched the circle and had been rewarded for a correct indication.

4.2.3.2. *Decoy training*

Prior to Trial 2, dogs had decoy training with the purpose of providing exposure to sheep, goat, rabbit, and glove decoy scents alongside possum scat. Each dog had two consecutive training sessions of 6 runs each, (12 runs in total) on the decoy scents. During each session, fresh possum scat was the target, and the number and type of decoy was increased from a single blank sample to having all decoys present simultaneously. This phase was completed when each dog confidently ignored non-target scents. Two of the dogs had previously worked with goat as a target, and when first exposed to goat scat they both gave an indication. They were not rewarded and did not indicate on goat again.

4.2.3.3. Trials 1 & 2

Trials 1 and 2 were on separate days, with the decoy training on the day in-between. Each trial had two sessions of 15 runs, with a break of approximately 120 mins in-between. The first session had three replicates of scats aged 1, 7, 14, 28 and 49 days old, and the second session had three replicates of scats aged 70, 112, 154, 187, and 239 days old. The same handler was used for all the trials and training.

Scats were presented incrementally by age (Schoon, 2005) and three individual replicates of each age were presented in a row. This was designed to determine each dog's detection threshold, with the expectation that three consecutive unsuccessful runs would signify the point of failure in detection. For each run, six samples were available for the dog to check, one was the target possum scat, and the other five samples were soil decoys (Trial 1), or 1 goat, 1 sheep, 1 rabbit, 1 glove, and 1 soil decoy (Trial 2). The positions of possum, decoy, and blank samples were randomised for each run.

For Trial 2, the plastic containers labelled A-F remained in the cinderblocks, and the scat and soil decoys were moved amongst them. The container lids were wiped with alcohol wipes between each run to provide a consistent scent across lids. For Trial 2, the plastic containers were collected between runs and brought to a sorting table where the target and decoy samples were moved amongst them, and all the lids were wiped with alcohol wipes before the containers were placed randomly in the cinder blocks. The positions were randomised for each run. This method adjustment was made to prevent the dogs using human scent that could have been present from the assistant changing sample containers between runs at the cinder blocks, as the assistant paused slightly longer at the target position and the position of the container the target was being swapped with. All containers were handled the same way, using latex gloves, alcohol wipes and tongs.

Both trials were double blind and between each run the dog and handler were positioned out of sight. When the run was ready, the handler directed the dog to begin the search with a ‘seek’ command while the assistant who had placed the scats stood out of sight to avoid any cues to the dog. The dogs freely approached the setup and were able to search the blocks in any order or direction they chose. When the dog indicated on a brick, the handler verbalised the brick's label. The assistant responded with 'correct' for a successful identification or 'no' for a false positive. Successful indications were rewarded with a toy thrown next to the brick for the dog to retrieve. If the indication was false, the handler directed the dog to ‘seek on’ until it correctly indicated the target. After a successful indication the dog and handler withdrew out of sight until the next run.

A run was only considered successful if the dog correctly identified the target on its first indication; subsequent correct indications after an initial error were not recorded as successful but were used as training opportunities by the handler.

Indications from the dog were recorded as true positives (indicating on the target possum scat), or false positives (indicating on any other brick that was not the target). The number of passes was used as a metric of difficulty for the dog in identifying the target and was counted as one when the dog indicated correctly on the first pass of the target (first time it checked the cinder block containing the target), or two or more if the dog checked the target but did not indicate until a later pass. If the dog checked the target but did not indicate until a second pass and did not indicate on a decoy, this was considered a successful run. However, the data was also analysed a second time where only a correct indication on the first pass of the target was considered successful (see the Appendix).

The two dogs that completed Trial 1 & 2 were both exposed to six scat samples of each age, while the dog who completed Trial 1 was exposed to three scat samples of each age. For each trial, there was the potential for 30 correct indications of target odour.

4.2.4. Data analysis

Due to the non-normal distribution of data (as confirmed by the Shapiro-Wilk test), non-parametric tests were used throughout the analysis. These included Fisher's exact test for comparing proportions of success rates and the Wilcoxon rank-sum test (Mann-Whitney U) for comparing success rates between trials and dogs. Additionally, the Kruskal-Wallis rank-sum test was used to evaluate the distribution of pass numbers across different scat ages.

To assess whether success rates varied across different scat ages for each dog and to compare success rates between dogs, Fisher's exact tests were used for each trial individually. Fisher's exact test was particularly suitable due to the small sample sizes and the binary nature of the success rates (1 = success, 0 = failure).

Wilcoxon rank-sum test (Mann-Whitney U) was applied to assess differences in the success rates of detecting possum scat across all ages between Trial 1 (soil decoys) and Trial 2 (species decoys) for Peggy and Mose individually (Beau participated only in Trial 1), to see if the presence of decoys in Trial 2 influenced the success rates for each dog. The presence of ties in the data, due to multiple instances of identical success rates, necessitated the use of an approximation with continuity correction, though this is unlikely to have substantially affected the interpretation of the results.

A Kruskal-Wallis rank-sum test was performed to evaluate whether the distribution of the number of passes before indication varied across different scat ages. The analysis was conducted separately for Trial 1 and Trial 2 due to potential differences in the experimental setups (soil vs. species decoys).

4.3. Results

The mean success rate of the dogs at identifying possum scat of all ages was high for Trial 1 and Trial 2 (Trial 1: M = 88%, SD = 33% (Range 78-100%), Trial 2: M = 88%, SD = 32% (Range 67-100%).

There was no significant difference in overall success rates between scat ages (for all dogs) in Trial 1 (Fisher's exact test: $p = 0.935$) or Trial 2 (Fisher's exact test: $p = 0.131$) (Figure 12).

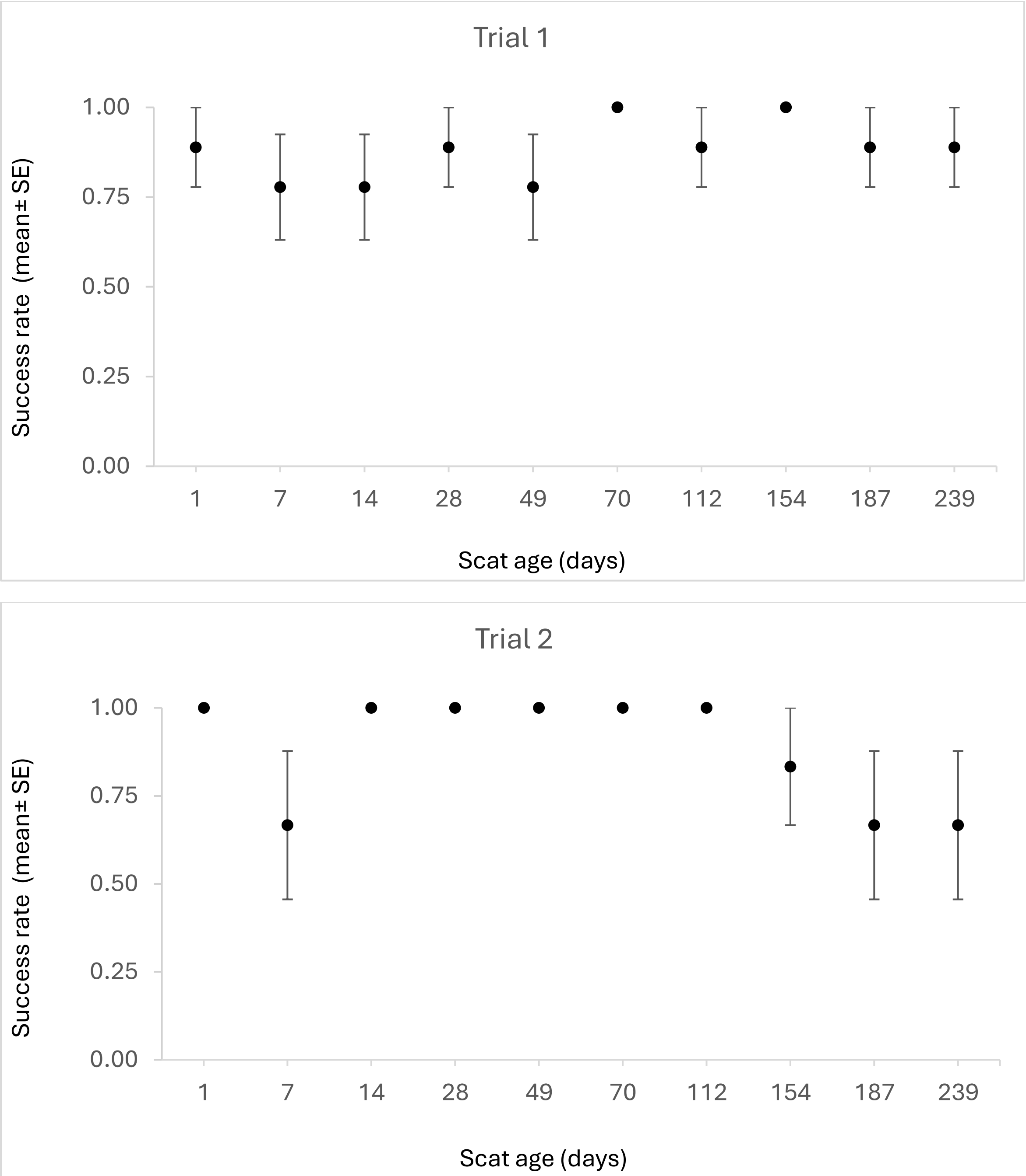


Figure 12. Plot of mean success rate from all dogs combined per scat age for Trial 1 (above), with Peggy, Mose, and Beau, and Trial 2 (below), with Peggy and Mose ± standard error.

Across all scat ages, there were no statistically significant differences in success rates between the trials, suggesting that the dogs were as successful at indicating all ages of scat regardless of trial (presence of species decoys or blanks). However, it should be noted that the small sample sizes, high number of ties, and low variability in success rates likely reduced the statistical power to detect any meaningful differences.

There were no significant differences in the performance between dogs (Peggy, Mose, or Beau) in Trial 1 (Fisher's exact test: $p = 0.366$) or Peggy and Mose in Trial 2 (Fisher's exact test: $p = 1.0$) (Figure 13).

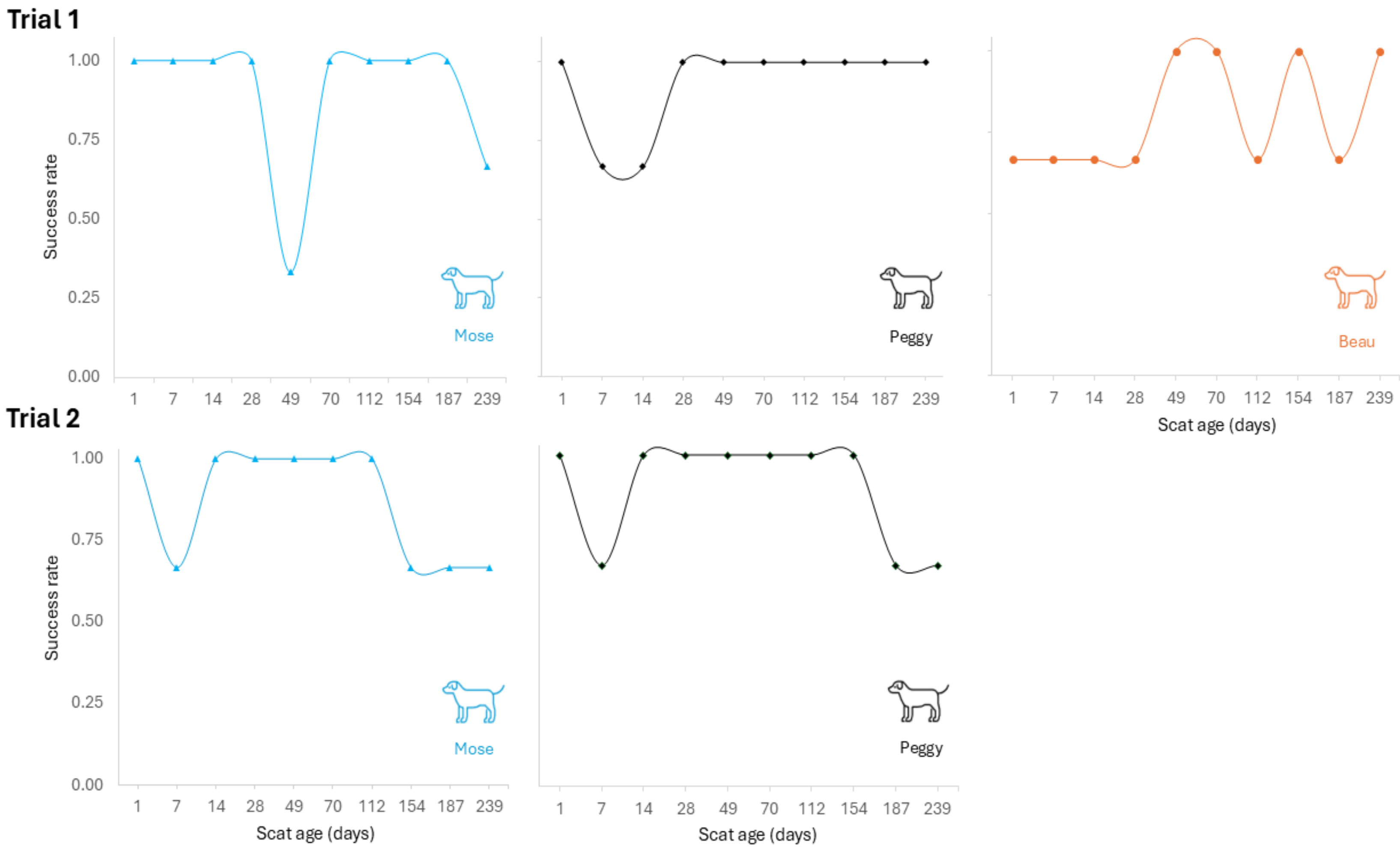


Figure 13. The success rate of each dog as a proportion of correct indications for each scat age for Trial 1 (soil decoys) and Trial 2 (species decoys). Success rates were calculated as a proportion of correct indications from three replicates per scat age per trial.

There was no significant difference in the mean success rates between trials for Peggy (Trial 1: M = 92%, SD = 13%; Trial 2: M = 89%, SD = 14%; Wilcoxon rank-sum test: W = 55, p = 0.651) or Mose (Trial 1: M= 94%, SD = 10%; Trial 2: M = 90%, SD = 13%; Wilcoxon rank-sum test: W = 58, p = 0.480), or in success by age of scat for either Peggy (Fisher's exact test: Trial 1: p = 1.0, Trial 2: p = 1.0) or Mose (Fisher's exact test: Trial 1: p = 0.20, Trial 2: p = 1.0). These results suggest that the age of the scat did not significantly affect the dogs' ability to correctly identify it as possum, regardless of whether the decoys were soil (Trial 1) or scat from other species (Trial 2).

The Kruskal-Wallis test revealed no significant difference in the distribution of pass numbers across scat ages for Trial 1 ($X^2 = 5.94$, df = 9, p = 0.746) or Trial 2 ($X^2 = 5.81$, df = 9, p = 0.759), indicating that the number of passes made by the dogs before indicating on the target was consistent across different scat ages in both trials.

There was no obvious difference in pass number between trials, the mean number of passes in Trial 1 was 1.33 (SD = 0.58) and the median was 1. In Trial 2, the mean number of passes was 1.16 (SD = 0.42) and the median was 1.

4.4. Discussion

All three dogs were capable of detecting possum scat up to 239 days old in a controlled trial. After the success of the dogs at Trial 1, methods were adjusted slightly to make sure the dogs were not able to take cues from anything other than the target scat, and they were just as successful in Trial 2. Success in general was high, and the non-significant difference between success rates for different scat ages or in the number of passes per age indicates that detecting the older scat was not any more difficult for the dogs. There was a dip in performance in Trial 2 for the last few scat ages for both Mose and Peggy, though this difference in performance to Trial 1 was not significant. As a result of having a small number of opportunities for each dog to detect each age of scat per trial, even a 2/3 success rate of 65% looks like a significant dip when comparing the performance of individual dogs. However, it is also possible that with larger sample sizes this pattern could be significant.

Given other examples of the detection of scat samples with age by dogs of 126 days in a controlled trial, and up to 334 days in a field trial (Brown et al., 2015; Da Silva et al., 2020; Leigh & Dominick, 2015), the success of the dogs in this trial on detecting scat of 239 days old with no apparent difficulty is unlikely to be an anomaly.

In Trial 1 blank samples contained soil, as scat samples in the field eventually degrade into soil and become undetectable. We assumed that as they aged, the scent of possum scat could become indistinguishable from the surrounding soil. However, in this trial we either did not reach that point, or we unintentionally cued the dogs onto the target in some way. In between each run the target position was changed. To do this, the assistant paused longer at the cinder block that had the previous target, and the cinder block where the new target would be placed. It is possible that scent pooling from the assistant assisted the dogs in detecting the target. However, the previous target cinderblock was always visited last, and the final movement of the assistant was to walk in a loop around all the cinderblocks and wipe them with an alcohol wipe. During decoy training and Trial 2, no dogs indicated on the decoy glove samples, so these could not have acted as a cue.

The presence of decoy samples from local species in Trial 2 did not impact success rates between trials. This is despite the decoy samples for sheep and goat being in most cases fresher than the target possum scat, and therefore likely to have a stronger smell. This may have assisted the dogs in discriminating the possum scat, and further trials with older decoy scats could increase the difficulty for the dogs in discriminating the possum target and provide different results.

In Trial 1, there was no statistical difference between the performance of the dogs, despite their different levels of working experience. During their previous work-related training, the age of possum scat was not purposefully controlled, but it was considered likely that Peggy (the most experienced dog) had been exposed to and rewarded for a wider range of scat ages than Beau (the least experienced). Beau detected possum scats of all ages in Trial 1. Dogs can quickly generalise between different sources or variations of target odour beyond what they have been trained on (Oldenburg Jr et al., 2016). For example, DeChant and Hall (2021) found that dogs spontaneously generalised within a 10-fold concentration range lower than the training stimulus.

Despite this, the chemical profile of a target odour can differ significantly over time and depending on the substrate. Exposure to a wide range of samples across the full spectrum of targets will allow successful generalisation depending on operational aims (Rust et al., 2016). In the context of possum scat surveys, where the detection of the full spectrum of aged scats could be problematic, it could be more worthwhile to focus training on scat within a particular age period so that dogs focus their search on specific aged scat. For example, DeChant and Hall (2021) rapidly trained dogs to discriminate on specific concentrations of odour within a spectrum of previously detected odour. Training dogs to detect fresh koala (*Phascolarctos cinereus*) scat only has been trialled by Detection Dogs for Conservation at the University of the Sunshine Coast (*R. Cristescu, pers.comm)*.

The ability of dogs to detect possum scat of various ages in a controlled setting does not represent their propensity to detect or indicate on samples in the field. Detection probabilities of samples in the field are affected by a range of variables aside from age, including terrain, weather, time of day, dog fatigue, previous recognition/reward, and distance. The detection probability of scat has been found to decline with distance and age (Baker et al., 2021; Da Silva et al., 2020). Cat (*Felis catus*) scat over 20 days old proved more difficult for a dog to detect during transect trials, and after 30 days, the detection probability of scats on the transect line itself was 50% (Baker et al., 2021). Different substrates can also impact detection success (Chilcote et al., 2018), and recognition of scats of different ages can vary with seasonal conditions (Brown et al., 2015). During possum scat degradation trials (Section 3), Peggy was directed to search over the area where a single clump of possum scat 49 days old was located and she showed no obvious interest. A newly trained possum scat dog in Dunedin (Scout) showed no interest in scat that was 219 days old, however, there was fresh possum scat nearby which was suitable distracting.

Further research on the ability and propensity of dogs to indicate aged scat in the field is an important next step. During this trial, we expected to possibly reach the limit of the dogs' ability, which is why samples were introduced incrementally in age. We did not see this occur, and further research that aims to detect a 'limit' would be interesting. In most cases, dog handlers and trainers set odour thresholds intentionally or unintentionally during training and work experience, as opposed to reaching the potential of the dogs ability (Leigh & Dominick, 2015).

5. SCENT MATCHING TRIAL

The ability to identify individuals is valuable in various monitoring contexts, such as determining if a pest captured is the one detected, estimating population abundance (e.g., using mark-recapture methods (Kerley, 2010), and addressing key questions about home range size, dispersal, pre- and post-control assessments, distinguishing remnant versus reinvading populations, and estimating the number of individuals remaining at low densities for removal.

Some animals have a unique appearance and can be distinguished visually, i.e., from camera trap footage. However, many animals are difficult to identify to an individual level and in these cases DNA genotyping can be used. This is not necessarily an effective method in every case, for example, when genetic diversity in the population is low or samples are degraded (Wasser et al., 2009). This results in a high cost for a potentially low rate of success.

Working dogs have been used to identify individuals, starting with the identification of people. For example, dogs match crime scene objects to a perpetrator in criminal lineups (Schoon, 1996) that are accepted as evidence in a court of law in some countries (Tomaszewski & Girdwoyn, 2006). Search and rescue dogs also identify individuals, tracking a specific person or persons scent from an item with their associated odour. Dogs have also been used to identify individual animals; Amur tigers (*Panthera tigris altaica*) and maned wolves (*Chrysocyon brachyurus*) from scat samples (Kerley & Salkina, 2007; Wasser et al., 2009), and Eurasian beavers (*Castor fiber)* from anal scent glands (Rosell et al., 2020).

The process of identifying individuals using dogs is called scent-matching. Samples are collected from the field and presented to a dog or dogs in a controlled setting. Typically, the dog is exposed to a sample from a specific individual, then tasked to match this odour with one from a range of samples from other individuals of that species (McKay, 2014). Essentially, the dog shows the handler the match by sitting next to it, and if there is no match, the sample is labelled as a "new" individual within that collection (Kerley, 2010). Scent-matching differs from the more traditional species detection (Section 1) as the target scent is changed frequently (within a species), so a long-term olfactory memory is not required.

Scent-matching has been used successfully to estimate population abundance using a capture-recapture design with scent-matching of samples by dogs representing the 'capture' (Kerley, 2010). Dogs identified 12 unique tigers from an area where scats were collected, which was consistent with tracking studies from the same area. The dogs could match multiple scats deposited over a 4-year period from one tiger with an accuracy of 100% (Kerley & Salkina, 2007).

Scent-matching offers a cost-effective alternative to expensive genotyping (Rosell et al., 2020), with dogs in some cases outperforming genetic testing (Harrison, 2006; Wasser et al., 2009). Possums at low densities often aggregate (Sweetapple & Nugent, 2009), leading to low genetic diversity and reduced success with DNA genotyping. Scent-matching could provide a reliable method for individual identification of possums in New Zealand.

5.1. Objectives

The objective was to do some initial exploration into the potential of scent-matching for possums. The identifying factor that dogs use to match individuals is unknown, and whether all species scat can be matched remains to be seen.

5.2. Methods

5.2.1. Scat samples

Scat was collected from 11 possums (7 females, 2 males, 2 unknown) that were shot on private property for pest control. Six of the possums (possums A-F) were to be used as targets, and the remaining five (possums G-K) were for use as additional decoys (Wasser et al., 2009).

At least 9 pellets were collected per individual using gloves and tweezers and were stored in separate plastic pottles and frozen (or aged and then frozen, details below). Pellets were defrosted for two hours before being dehydrated at 45°C for 15 hours. Samples from one individual were placed in the dehydrator at a time in small paper bags to avoid cross contamination. All the samples used for training and the trial were dehydrated so they could be stored and used without thawing-freezing repeatedly, which can impact the dogs ability to identify individuals (Wasser et al., 2009).

From each possum, three pellets were frozen immediately until they could be dehydrated ‘fresh’ (1-day old), while the remaining six pellets were aged before being frozen until they could be dehydrated. Three were aged to 3-days old, and three to 5-days old, so that for each individual possum there were three samples of 1,3, and 5-days old. The samples were aged by leaving them outside under cover, protected from precipitation and separated by individual to prevent cross contamination. Aged samples were used as a proxy for samples collected over time from the same individual (Kerley & Salkina, 2007; Wasser et al., 2009). The purpose of having different aged samples was to test the dogs’ ability to discriminate the same individual from different samples (Kerley & Salkina, 2007).

Horse dung was used initially as a surrogate for possum scat, and this was collected from individual horse stalls and dehydrated following the same methods as for the possum scat except it was dehydrated on the same day of collection (not frozen or aged).

5.2.2. Scent-matching potential

To verify that scent-matching for possums would be applicable in a practical context it was important to consider if possums interact with the scat of other possums in the wild in any way that would cause scent contamination. To investigate this, possum pellets were placed in front of camera traps in a low-density possum population on Otago Peninsula in Dunedin, and a high-density population on Rakiura, and the footage monitored to check for interactions with the scat from wild possums.

5.2.3. Dogs

Two dachshunds were used for this trial. Heidi, a female 5 years old and Frankie, a male 8 years old. Both dogs had previously been trained to detect small pieces of kong (a type of rubber) and had some possum tracking training (scent trails). Heidi was introduced to the concept of scent-matching much earlier than Frankie and had approximately 70 training sessions that included different types of kong, horse dung, different types of tea bag, and possum scat. Frankie had approximately 10 training sessions with kong and possum scat. Sessions would usually include 6 runs/chances to detect the target.

5.2.4. Training

Training for scent-matching required that both dogs be familiar with a 'smeller', a cup containing the target which was presented to them by the handler, and a lineup, a selection of samples to search for the correct target (a 'match' to the smeller). These concepts were new for both dogs.

During scent-matching training the dogs most often received a reward for sniffing the smeller (alternated at times with no reward which was found to be unsuccessful), and again when they indicated on the correct target in the lineup. The dogs were asked to search the lineup independently, and on a correct indication the handler used a clicker to communicate to the dog they were correct, and then the dog returned to the handler to be rewarded. If the dog made an incorrect indication, it was asked to 'seek on' and try again. If the dog continued to make incorrect indications it was asked to return to the handler and did not receive a reward and the target position was reset. The dogs were moved on to the next step in training when they were at least 80% successful in the task, and ideally at 100% for one session (of 6 runs).

The general training process below was followed in full for Heidi (H), Frankie's (F) training was streamlined to the first and last few steps. Further details on training can be provided on request.

- (H&F) The smeller concept was introduced, using food to get the dog to sniff inside the cup, which was replaced with the target and the dog was rewarded for sniffing the target inside the cup.
- (H&F) The concept of a line up was introduced (initially a circle and then later a straight line) using kong as the target and increasing blank decoys one at a time to a maximum of five.
- (H) The kong was switched to horse A and the dog was asked to find horse A in the lineup with blank decoys added one at a time to a maximum of five.
- Two samples from horse B were added as decoy samples and the dog was asked to continue to select horse A.
- (H) Horse C was added to the decoy samples along with horse B and the dog was asked to continue to select horse A.
- (H) The dog was given alternating targets in the smeller for horse B and C, and the lineup contained the target, the other horse as a decoy, and a blank. There was no horse A. The main difference in this step is that there were past targets as decoys, and the target was changed more frequently.
- (H&F) Possum scat was introduced as the target. Possum 1 was always used as the target and possums 2-5 were used as decoys. A sit indication was introduced (building on from a stand and point).
- (H&F) Possum A (from the trial samples) was introduced as the target, with 5 decoys of different possums from the trial samples.
- (H&F) Double blind runs for each dog 3x matching possum A1 (1-day old sample) to possum A1, 3x matching possum A1 to possum A3 (3-day old sample), and 3x matching possum A1 to possum A5 (5-day old sample). The decoy samples remained consistent.

5.2.5. Trial

5.2.5.1. Structure

The trial was run over three days with a total of 84 runs, 42 per dog. Each dog had two sessions per day, on day one the dogs had nine runs in each session, on days two and three the dogs had six runs per session.

The number of sessions was reduced to decrease fatigue and stress for the dogs. Each dog rested for at least two hours between sessions.

Each dog was presented with a total of six individual possums to match (A-F), with three different aged scats (1,3, and 5-days old). In total, 36 of the runs (87%) used 1-day old scat as the target, and there were 24 runs with a 3- and 5-day old target respectively. The sniffer was always 1-day old scat, except for on day three (see details below). The dogs had the opportunity to match an individual possums scat a minimum of six times, nine times for possums A and B on trial day one.

On each trial day, the dogs were presented with two different individuals as the target, one individual per session, which they were asked to identify from among five decoy individuals. The decoy individuals were all 1-day old samples, and were randomised between individuals B-K, so that dogs were first exposed to a target as a decoy (except for possum A), and then a decoy became a target. After the dog was given an individual to match, the individual was removed from further trials, so that dogs were not asked to ignore previous targets. The number of runs matching an individual and the age of the scat differed by trial day (see below). The session targets were identical for each dog, with different randomised target and decoy positions per run. Fresh sniffer samples were used for each dog, and multiple samples from each decoy possum were used interchangeably.

<u>Day one</u>

Dogs each had two sessions of nine runs. Each session had a different target possum and three different ages of scat progressing from 1-3-5-days old, with three runs per age. The smeller was always day-1 (e.g., possum A1), and the target increased in age (e.g., A1x3, A3x3, A5x3).

<u>Day two</u>

Dogs each had two sessions of six runs. Each session had a different target possum and three different ages of scat progressing from 1-3-5-days old, with *two* runs per age. The smeller was always day-1 (e.g., possum C1), and the target increased in age (e.g., C1x2, C3x2, C5x2). The difference to trial day one was a reduction in the number of runs, which resulted in a reduction in the number of replicates per age target.

<u>Day three</u>

Dogs each had two sessions of six runs. Each session had a different target possum. In the first session, all the target scats were 1-day old (e.g., E1x6), and no aged scats were used. In the second session, the age of the sniffer scat was changed for the first time, and matched the target scat age (e.g., sniffer and target: F1x2, F3x2, F5x2).

5.2.5.2. Set-up

The line-up was a straight line perpendicular to the handler with six plastic PVC pipes with 12 cm high stainless-steel cups inside. The PVC pipe kept the cups stable, and the depth of the cups prevented the dogs from getting close to the samples. The scat samples were in 5 cm high glass jars that were moved between the cups for each run using tweezers. New steel cups were used for each session, and all surfaces sterilised with alcohol wipes. The handler was positioned ~ 1 m from the first sample position (position 6), and the assistant was positioned ~1.5 m from the line up during runs. The handler was blind to the position of the target and the assistant was blind during each run to avoid giving the dog cues as to the correct target position. During a run the handler would say the position the dog indicated on out loud, to which the assistant would confirm if it was correct. The same procedure for reward was followed as described in the training section, except that on day one, Heidi was given multiple opportunities to indicate correctly while

Frankie was given a maximum of two attempts. On days two and three both dogs were given a maximum of two attempts before being recalled to the handler.

If a dog indicated on the target sample correctly on its first attempt it was considered a successful run. If the dog false indicated on a decoy and then correctly on the target, it was considered a failed run.

5.2.6. Data analysis

A logistic regression model was fitted using maximum likelihood estimation to investigate the effects of scat age, target position, trial day, dog, and the interaction between dog and trial day on the likelihood of a successful identification by the dogs. The model demonstrated a good fit, with a residual deviance of 83.51 (on 78 degrees of freedom) and an AIC of 95.51, indicating that the model adequately explained the variability in the data.

A Chi-Square Goodness of Fit Test was used to assess whether the distribution of target samples, false indications, and an over-selected decoy possum across lineup positions were evenly distributed. A Pearson's Chi-Square Test of Independence was employed to compare the distribution of successful first indications across sample positions to evaluate if their distribution was dependent on sample position.

5.3. Results

5.3.1. Scent-matching trial

The success rate, measured as the proportion of correct first indications during the trial, was very similar between the dogs, with Frankie correct 43% of the time and Heidi 52%. This difference was not statistically significant (Pearson's Chi-squared test with Yates' continuity correction, $X^2 = 0.42955$, df = 1, p-value = 0.5122)

If the runs were considered successful when the dogs indicated correctly on their second attempt (having false indicated on a decoy on their first try), the scores increased to 83% for Frankie and 69% for Heidi. This alternative metric is provided as this pattern of behaviour became recurring for Frankie from session two on day two, where he would give a false indication on his first run (consistently on decoy possum G), and then correctly indicate the target when asked to try again. This happened on 8/12 runs on day three.

Overall, the dogs could successfully match the scat sample presented in the smeller with the target sample 64% of the time when the target scat was 1-day old, but performance decreased significantly with age (log regression, estimate = -0.6663, SE = 0.1904, $p < 0.001$), suggesting that as scat samples aged, the dogs were less likely to correctly identify the target (Figure 14).

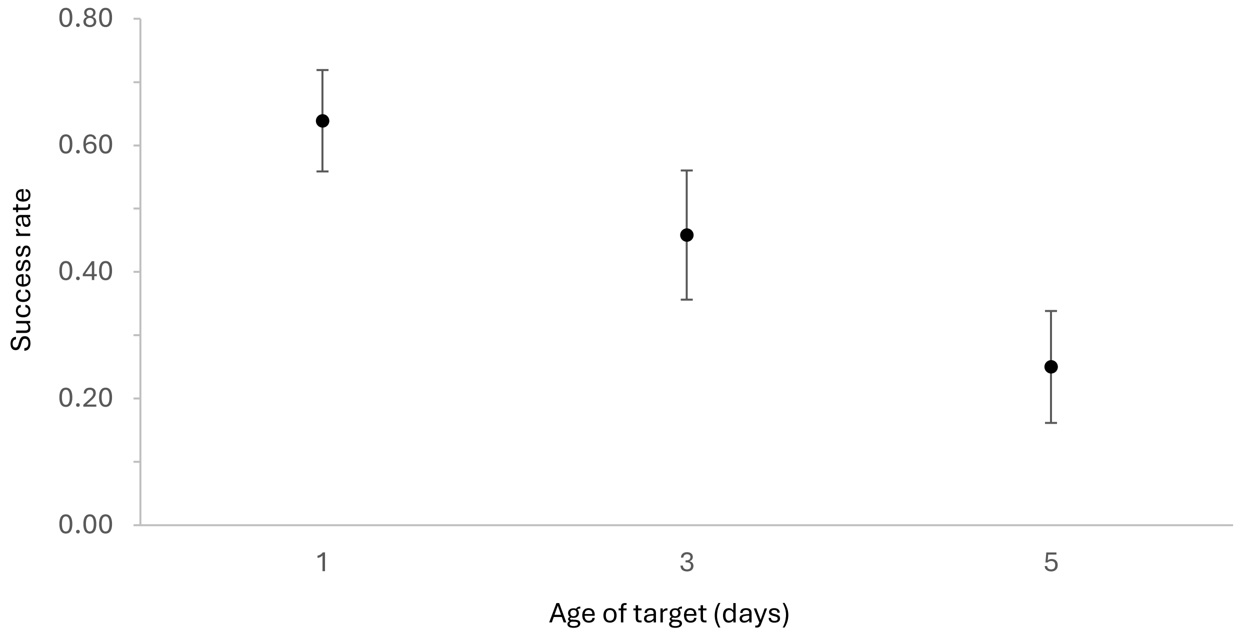


Figure 14. The overall success rate of the dogs as a proportion of the total number of successful first indications per age of the target scat (1,3, or 5-days old) ±SE.

This reduction in success with age of the target scat was present for both dogs, and more pronounced for Heidi (Figure 15).

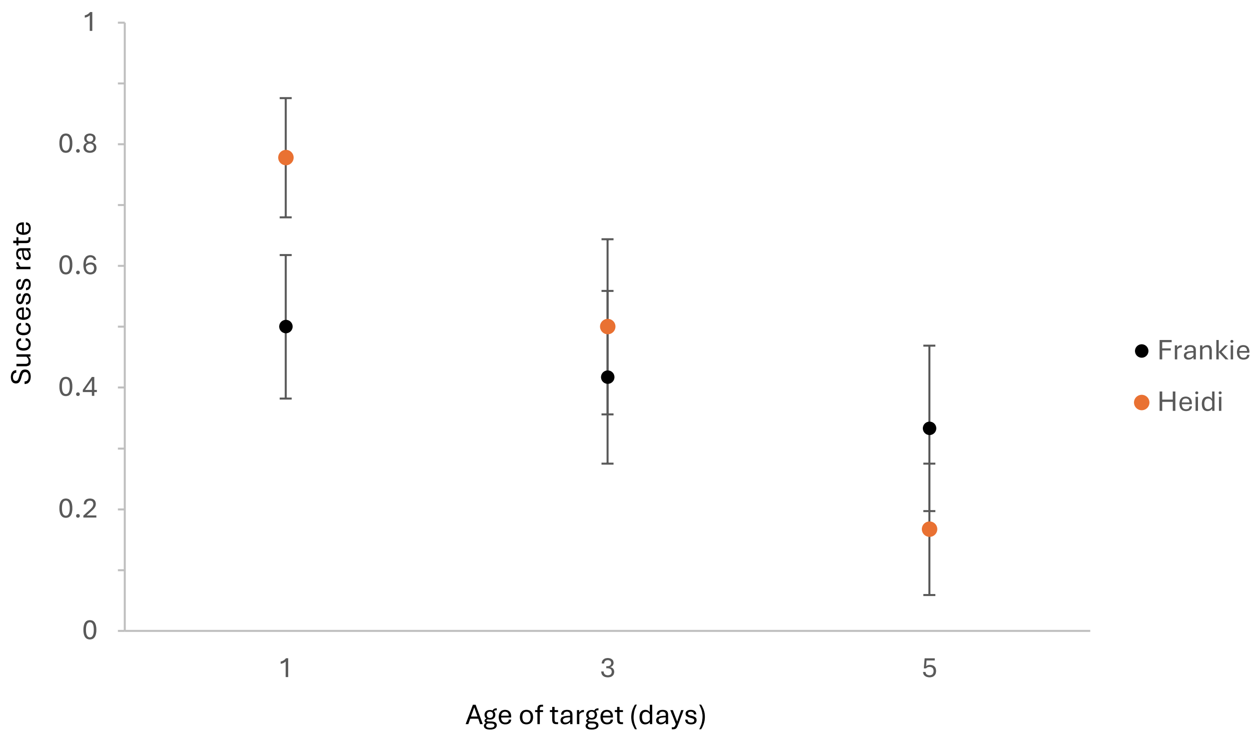


Figure 15. The success rate of Frankie (black) and Heidi (orange) as a proportion of the total number of successful first indications per age of the target scat (1,3, or 5-days old) ±SE.

Performance significantly decreased over time, with trial day having a significant negative effect on the likelihood of success (Estimate = -1.7784, SE = 0.5502, p = 0.0012) for both dogs (Figure 16).

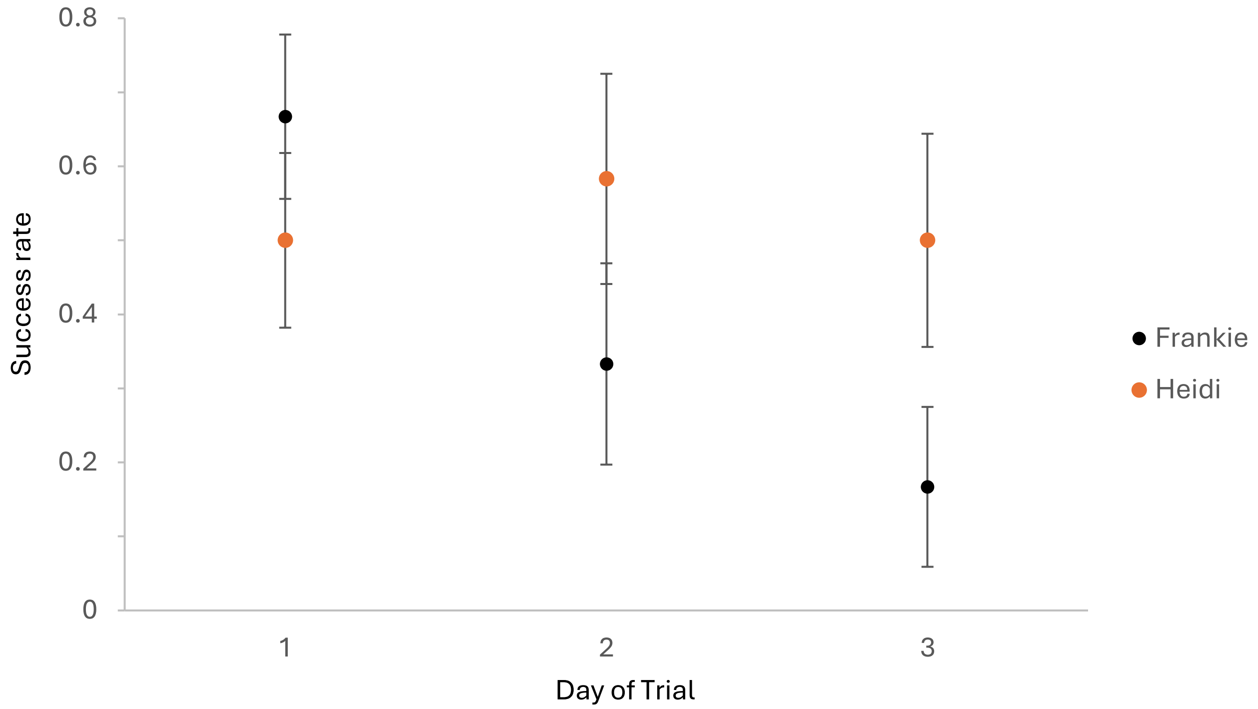


Figure 16. The success rate of Frankie (black) and Heidi (orange) as a proportion of the total number of successful first indications per trial day ±SE.

Heidi's baseline performance on day one was lower than Frankie's, but the difference did not reach statistical significance (Estimate = -2.3196, SE = 1.3422, p = 0.0840). However, the interaction between dog and day was significant (Estimate = 1.5485, SE = 0.6876, p = 0.0243), indicating that Heidi's performance improved relative to Frankie's across the trial days.

Trial set up

The random position of the target in the lineup was distributed evenly across positions 1-6 (Chi-squared test for given probabilities: $X^2 = 0.57143$, df = 5, p-value = 0.9893). However, the distribution of successful first indications was not even across sample positions (Pearson's Chi-squared test: $X^2 = 13.573$, df = 5, p-value = 0.01856).

The log regression model indicated that the position of the target in the lineup had a significant positive effect on the likelihood of success (Estimate = 0.5062, SE = 0.1789, p = 0.0047), indicating that some positions were more likely to lead to successful indications (Figure 17).

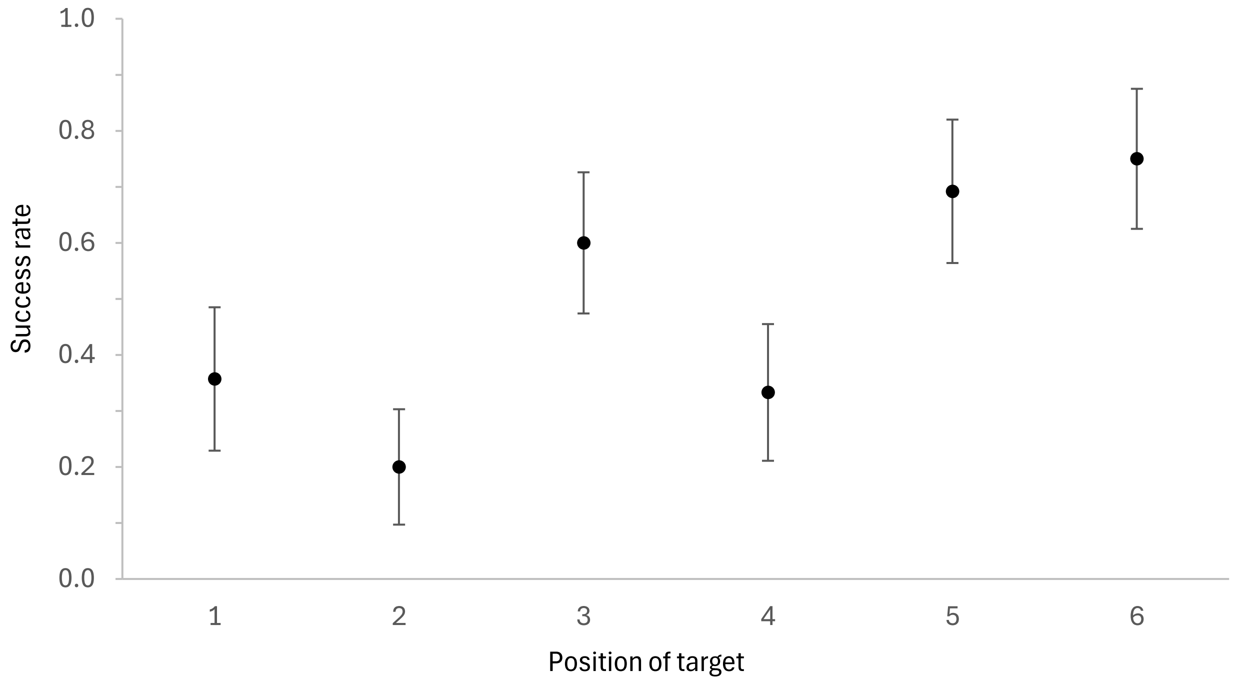


Figure 17. The overall success rate as a proportion of the total number of correct first indications for both dogs combined based on the position of the target sample ±SE.

Combining data for both dogs, the target was most likely to be successfully identified when it was in position six, closely followed by position five, the two closest positions to where the dog began each run, and where the handler was positioned. Therefore, the dogs were more likely to be successful if the target was encountered sooner in the lineup.

Looking at the impact of target position on the success of the dogs independently, both dogs were more likely to succeed when the target was in position six (Figure 18), while position five was the next most likely for Heidi, and position three for Frankie.

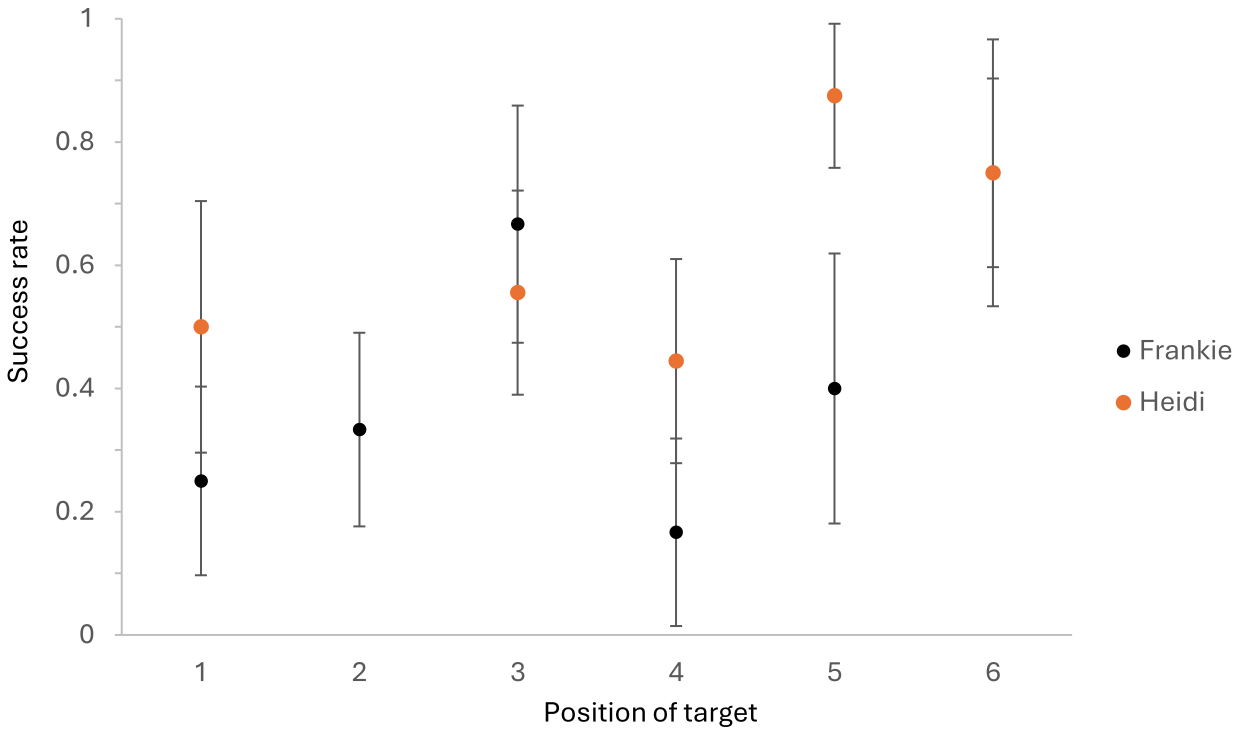


Figure 18. The success rate for Frankie (black) and Heidi (orange) as a proportion of the total number of correct first indications based on the position of the target ±SE.

When dogs gave a false indication instead of correctly identifying the target, the distribution of false indications across positions was not even (X^2 = 43.143, df = 5, p-value = 3.457e-08). When either dog gave a false indication, they were more likely to do so on position 6 than any other position (Figure 19).

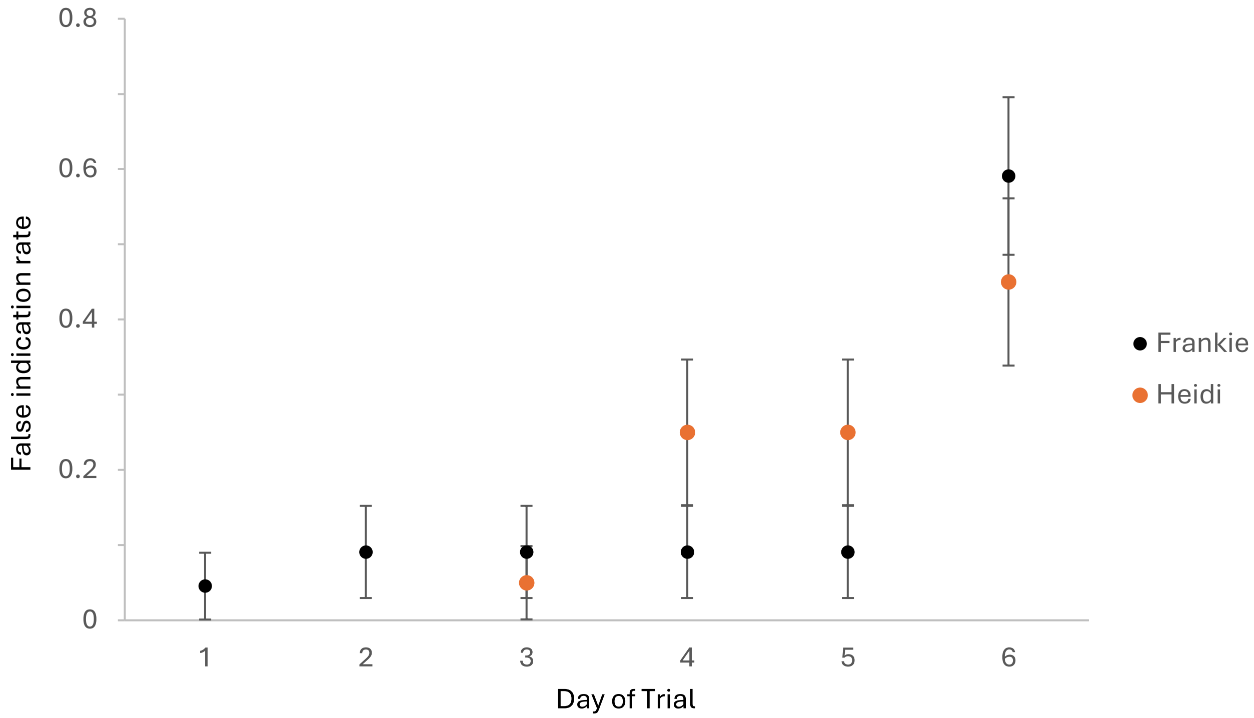


Figure 19. The false indication rate for Frankie (black) and Hedidi (orange) as a proportion of the total number of false indications per sample position ±SE.

Dogs were also more likely to select a particular possum when they made a false indication, possum 'G'. Heidi selected possum G 50% of the time when she false indicated, and Frankie 55%, while other possums (B-K) were selected 4.5-15% of the time by either dog.

A Chi-Square Goodness of Fit Test for the distribution of possum G across positions 1-6 shows it was distributed evenly (X^2 = 4.8033, df = 5, and p-value = 0.4404), and in fact was more frequently in position one and four than five or six (Table 5) and therefore position alone was not likely to be the reason it was selected.

Table 5. The total number of times possum G was present as a decoy sample per lineup position.

Position	No. of times present
1	13
2	7
3	9
4	15
5	9
6	8

The number of times that possum G was available as a decoy was also unlikely to be a factor in it being frequently selected, given the similar availability of some of the other decoy possums (Table 6).

Table 6. The total number of times each decoy possum was present in the lineup during the trial.

Possum	No. of times present
B	8
C	18
D	27
E	34
F	46
G	61
H	58
I	64
J	55
K	49

5.3.2. Scent-matching potential

The camera trap footage was not analysed in any formal way. In the high-density population on Rakiura, possums did not show any obvious interest in the scat from other possums. On the Otago Peninsula in the very low-density population, possums appeared very interested in the foreign possum scat. They did not cause any obvious scent contamination, but this needs to be explored further. The possible use of foreign possum scat as a bait for possums is also worth investigation, as traps in the low-density population that were next to the cameras caught two possums within days after the scat had been placed that had avoided capture for months previously. There were also several observations during the scat degradation trials (Section 3) of fresh possum scat deposits near the scat groups being monitored. During a study in which Tasmanian devil (*Sarcophilus harrisii*) scats were being monitored using camera traps, devils were seen depositing scats in the same manner as the possums above, and they concluded they had artificially created new 'latrines' (Paton et al., 2021).

5.4. Discussion

Both dogs could match scat samples from individual possums, but the overall success rates were reasonably low (Frankie 42%, Heidi 56%), particularly for samples over one day old. Dogs matching individual people in the Netherlands had similar results, the average success rate was 46% when decoys were blank (with no odour), and 59% when decoys were from non-target people (Schoon, 2005). However, trials where dogs were matching individual animals from scat have had much higher success rates. Dogs correctly identified individual Amur tigers 87% of the time with 58 scats from 25 individuals in Russia (Kerley & Salkina, 2007). In Washington state, dogs could consistently match 18 scat samples from five closely related maned wolves that were on the same diet (Wasser et al., 2009). Dogs were also able to scent-match Eurasian beavers from anal secretion glands with an average accuracy of 88.9%, sensitivity of 66.7% and specificity of 93.3% (Rosell et al., 2020).

There were a few factors that likely influenced the dogs' performance and success rate. Both performed better during training, and it's possible that increased stress, an increased number of unknown decoys and the longer times between runs (due to the randomisation of all sample positions that had to be repositioned) affected the dogs. Rosell et al. (2020) used only 1-3 decoys per run, which could have simplified the task. It can be difficult to predict how successful a dog will be in a trial from how it performs during training (Jezierski et al., 2010), and success can also decline due to temperament or losing interest in the task (Kerley & Salkina, 2007).

Previous studies have used multiple samples from individuals over time (Kerley & Salkina, 2007; Wasser et al., 2009), as opposed to using one sample for an individual that has been aged for different time periods. This is an important difference and could impact the variance between samples in an unknown way, and it is possible that matching scats of different ages was too much of an increase in difficulty. Schoon (2005) describes the effect of ageing of the target odour on the performance of dogs recognising the target perpetrator in a line-up. The dogs were highly successful at matching odours collected on the same day, but results dropped and become more variable with older samples. Dogs matching individual tigers performed better when samples were purposefully 'aged' for 48 hours, and the age of some of the samples used for testing the dogs were unknown (Kerley & Salkina, 2007). However, the scent profile of possum scat could alter in different ways over time, and the tiger scat dogs likely received more training on samples of different ages than the dogs in this trial. Rosell et al. (2020) used only fresh samples and one per individual, which could partly explain the dogs' high success rates in that trial. In some trials some of the scats were split in half and the dog had to match the other half (Kerley & Salkina, 2007; Wasser et al., 2009).

It is possible that the dogs didn't fully comprehend the task, and as a result when the difficulty increased with aged samples their performance declined. Smith et al. (2003) found that when there was no target in the lineup the dogs were more likely to give a false indication. If the dogs in this trial didn't recognise the aged scat samples as being a match to the fresh scat in the sniffer, they could be more likely to throw a false indication than to return to the handler with no indication. Training that incorporates runs in which there is no target could potentially mitigate this by giving the dogs a process to follow when they cannot identify the target.

Introducing new target individuals frequently in early training, and having multiple samples for training could increase the dogs grasp of the scent-matching task concept (Kerley & Salkina, 2007; Wasser et al., 2009). The dogs in this trial may have been imprinted too thoroughly on a single individual and therefore not fully comprehended the task as matching individuals so much as matching samples. Having multiple samples per individual that are collected over time would ensure the dogs are identifying individuals (Kerley & Salkina, 2007). Increasing the difficulty of the task in more incremental steps during training

(Rosell et al., 2020) and training for a longer period of time (Kerley & Salkina, 2007) could improve success rates.

A straight lineup perpendicular to the handler was problematic in this trial. Both the likelihood of success and of false indication was related to position in the lineup and were both more likely to occur on the position closest to the handler (position 6). This implies that when the target was in position 6 the dogs could have had a successful indication by chance. More commonly, the chance of false indication increases as a dog moves down a straight line-up, as the probability of finding a correct sample becomes higher per container (Johnen et al., 2017; Wasser et al., 2009). The dogs in this study moved back towards the start of the lineup before giving a false indication, potentially because this was the position closest to the handler and location of reward. A circular arrangement of samples would be more appropriate for this task (Kerley & Salkina, 2007; Schoon, 2005), which prevents the increasing probability of finding a correct target at each sample because a circle does not have a recognisable end. Introducing multiple targets per run (Kerley & Salkina, 2007), runs with no target (Johnen et al., 2017) and rewarding the dog at the source of the scent could also be effective strategies (Rosell et al., 2020; Wasser et al., 2009).

Dogs can show favouritism towards some samples for unknown reasons and it is best if these samples can be identified and removed during training to avoid misidentifications (Kerley & Salkina, 2007). Both dogs in this trial indicated disproportionately on decoy possum G when they false indicated, and Frankie appeared to favour possum G over the target, as he developed a pattern of indicating first on possum G and then the target when given a second chance at the same run. It is highly unlikely the sample was contaminated due to strict handling protocols, and multiple samples from possum G were used interchangeably. Dogs scent-matching tigers favoured three out of 58 scats, from three different tigers (Kerley & Salkina, 2007). This propensity of dogs to get fixated on samples is important to consider for scent-matching in a practical context to avoid misidentifications.

The performance of dogs is most often a reflection of the training process and trial design, rather than an ability to achieve the task (Kerley & Salkina, 2007). Scent-matching to identify individuals has been proven successful for some species including humans (Kerley & Salkina, 2007; Rosell et al., 2020; Schoon, 1996; Wasser et al., 2009). Dogs have even be trained to discriminate breeder from subordinate wolves using scat samples (Bottaro & Marucco, 2024). This trial has shown that dogs can discriminate the scat of an individual possum from among a range of different possum scats. This promising initial result requires further development to ensure the dogs can match individuals and not just samples, to explore the impact of the age of samples on detection, and to understand the various factors affecting the dogs' performance. Once success rates are improved reliably, scent-matching could be a cost-effective and efficient alternative to DNA genotyping for population monitoring where identifying individuals is important (Kerley & Salkina, 2007). Training dogs is time consuming and expensive, but once trained, the results are immediate (Rosell et al., 2020).

6. SUMMARY

This study highlights the potential of scat surveys to enhance possum control by targeting areas with scat detections, thereby reducing the need for blanket control efforts. Results from scat surveys in Taranaki throughout a possum eradication suggest that integrating scat detection dogs with grid-based spatial control systems can enhance the cost-effectiveness of eradication programs.

The persistence of possum scat is important to consider, and it varied little between sites, locations, and seasons. In most cases, scat lasted for the full duration of the 50-day trial, especially in drier locations like Dunedin and Banks Peninsula, where additional persistence beyond 50 days was recorded (up to 217 days in Dunedin). Scat characteristics such as external smell and mucous/shine were found to be reasonable indicators of freshness, with these features predominantly present in scat less than seven days old. After this period, scat became harder, crumblier, and had a darker internal colour, useful indicators in identifying scat older than one week.

The aged scat trials demonstrated that detection dogs were highly effective at identifying possum scat and could still detect scats up to 239 days old in a controlled environment. This reinforces the utility of scat detection in low-density possum populations, but also raises an important factor to consider in the timing of scat surveys and interpretations of survey results. Older scat may indicate historical rather than current possum presence, potentially skewing estimates of possum population presence and distribution if not accounted for. By targeting fresh scat, surveys may be more relevant and accurate.

The scent-matching trial showed that dogs are capable of matching scats from individual possums. This highlights the potential for scent-matching to provide a practical alternative to DNA genotyping in monitoring low-density possum populations, where genetic diversity can be limited and genotyping less effective. The use of scent-matching could improve identification and tracking of individual possums in control operations, offering a cost-effective and efficient tool for pest management.

Overall, this project provides an initial exploration into some of the factors important to possum scat surveys, and the potential utilisation of scat dogs within eradication projects. Dog teams are a valuable tool but require well-structured and thorough trials to capture results in a meaningful way that can be interpreted with confidence.

REFERENCES

Anderson, D. P., Ramsey, D. S., Nugent, G., Bosson, M., Livingstone, P., Martin, P. A., Sergeant, E., Gormley, A. M., & Warburton, B. (2013). A novel approach to assess the probability of disease eradication from a wild-animal reservoir host. *Epidemiol Infect*, *141*(7), 1509-1521. https://doi.org/10.1017/S095026881200310X

Baker, G. B., Candy, S., Robinson, S., Friend, J. A., Holdsworth, M., Jensz, K., Page, M., & Algar, D. (2021). Effectiveness of dogs for detecting feral cat scats in wheatbelt reserves of Western Australia. *Wildlife Research*, *48*(8), 690-700.

Barnes, R., Asamoah‐Boateng, B., Naada Majam, J., & Agyei‐Ohemeng, J. (1997). Rainfall and the population dynamics of elephant dung‐piles in the forests of southern Ghana. *African Journal of Ecology*, *35*(1), 39-52.

Barnes, R., & Barnes, K. (1992). Estimating decay rates of elephant dung piles in forest. *African Journal of Ecology*, *30*(4), 316-321.

Barnes, R., & Dunn, A. (2002). Estimating forest elephant density in Sapo National Park (Liberia) with a rainfall model. *African Journal of Ecology*, *40*(2), 159-163.

Becker, M. S., Durant, S. M., Watson, F. G., Parker, M., Gottelli, D., M'soka, J., Droge, E., Nyirenda, M., Schuette, P., & Dunkley, S. (2017). Using dogs to find cats: detection dogs as a survey method for wide‐ranging cheetah. *Journal of Zoology*, *302*(3), 184-192.

Bengsen, A. (2014). Effects of coordinated poison-baiting programs on survival and abundance in two red fox populations. *Wildlife Research*, *41*(3), 194-202.

Bengsen, A., Robinson, R., Chaffey, C., Gavenlock, J., Hornsby, V., Hurst, R., & Fosdick, M. (2014). Camera trap surveys to evaluate pest animal control operations. *Ecological Management & Restoration*, *15*(1), 97-100.

Bennett, E., Jamieson, L. T., Florent, S. N., Gill, N., Hauser, C., & Cristescu, R. (2022). Detection dogs provide a powerful method for conservation surveys. *Austral Ecology*, *47*(4), 894-901. https://doi.org/10.1111/aec.13162

Bottaro, R., & Marucco, F. (2024). Will dogs sniff out social rank? The use of detection dog in wolf scat identification. *Journal of Veterinary Behavior*, *75*, 46-53.

Breedt, B., & King, C. M. (2022). Distribution and detectability of mammalian pests in the Waikato Region. *New Zealand Journal of Zoology*, *49*(1), 37-52.

Brown, W., Ramsey, D., & Gaffney, R. (2015). Degradation and detection of fox (Vulpes vulpes) scats in Tasmania: evidence from field trials. *Wildlife Research*, *41*(8), 681-690.

Campbell, D., Swanson, G., & Sales, J. (2004). Methodological insights: comparing the precision and cost‐effectiveness of faecal pellet group count methods. *Journal of Applied Ecology*, *41*(6), 1185-1196.

Chilcote, B., Rust, L., Nizio, K. D., & Forbes, S. L. (2018). Profiling the scent of weathered training aids for blood-detection dogs. *Science & Justice*, *58*(2), 98-108.

Comer, S., Speldewinde, P., Tiller, C., Clausen, L., Pinder, J., Cowen, S., & Algar, D. (2018). Evaluating the efficacy of a landscape scale feral cat control program using camera traps and occupancy models. *Scientific Reports*, *8*(1), 5335.

Cowan, P. (1992). The eradication of introduced Australian brushtail possums, Trichosurus vulpecula, from Kapiti Island, a New Zealand nature reserve. *Biological Conservation*, *61*(3), 217-226.

Da Silva, A. R., De Oliveira, M. r. L., & Duarte, J. M. B. (2020). Incorporating the sampling effectiveness of detection dogs in the faecal standing crop method. *European Journal of Wildlife Research*, *66*(4). https://doi.org/10.1007/s10344-020-01388-9

DeChant, M. T., & Hall, N. J. (2021). Training with varying odor concentrations: implications for odor detection thresholds in canines. *Animal Cognition*, *24*(4), 889-896.

Glen, A. S., Anderson, D., Veltman, C. J., Garvey, P. M., & Nichols, M. (2016). Wildlife detector dogs and camera traps: a comparison of techniques for detecting feral cats. *New Zealand Journal of Zoology*, *43*(2), 127-137. https://doi.org/10.1080/03014223.2015.1103761

Glen, A. S., Garvey, P. M., Innes, J., & Buchanan, F. (2023). What can wildlife-detection dogs offer for managing small mammalian predators? *Wildlife Research*.

Gormley, A. M., Forsyth, D. M., Griffioen, P., Lindeman, M., Ramsey, D. S., Scroggie, M. P., & Woodford, L. (2011). Using presence-only and presence-absence data to estimate the current and potential distributions of established invasive species. *J Appl Ecol*, *48*(1), 25-34. https://doi.org/10.1111/j.1365-2664.2010.01911.x

Gormley, A. M., Forsyth, D. M., Wright, E. F., Lyall, J., Elliott, M., Martini, M., Kappers, B., Perry, M., & McKay, M. (2015). Cost-Effective Large-Scale Occupancy-Abundance Monitoring of Invasive Brushtail Possums (Trichosurus Vulpecula) on New Zealand's Public Conservation Land. *PLoS One*, *10*(6), e0127693. https://doi.org/10.1371/journal.pone.0127693

Gsell, A., Innes, J., de Monchy, P., & Brunton, D. (2010). The success of using trained dogs to locate sparse rodents in pest-free sanctuaries. *Wildlife Research*, *37*(1), 39-46.

Harestad, A. S., & Bunnell, F. L. (1987). Persistence of black-tailed deer fecal pellets in coastal habitats. *The Journal of Wildlife Management*, 33-37.

Harrison, R. L. (2006). A Comparison of Survey Methods for Detecting Bobcats. *Wildlife Society Bulletin*, *34*(2), 548-552. https://doi.org/10.2193/0091-7648(2006)34[548:Acosmf]2.0.Co;2

Hibert, F., Maillard, D., Fritz, H., Garel, M., Abdou, H. N., & Winterton, P. (2011). Ageing of ungulate pellets in semi-arid landscapes: how the shade of colour can refine pellet-group counts. *European Journal of Wildlife Research*, *57*, 495-503.

Hickling, G. J., & Thomas, M. D. (1990). (FWE 90/10).

Hurt, A., Guscio, D., Tirmenstein, D. A., Richards, N., Burch, A., & Marler, M. (2014). Using search dogs for biological eradication programs—A tale about Dyer's Woad (Isatis tinctoria L.). Proceedings of the 3rd Northern Rockies Invasive Plants Council Conference,

Jachmann, H., & Bell, R. H. V. (2008). The use of elephant droppings in assessing numbers, occupance and age structure: a refinement of the method. *African Journal of Ecology*, *22*(2), 127-141. https://doi.org/10.1111/j.1365-2028.1984.tb00686.x

Jamieson, L. T. J., Hancock, A. L., Baxter, G. S., & Murray, P. J. (2021). How quoll-ified are northern and spotted-tailed quoll detection dogs? *Wildlife Research*, *48*(4), 376-384. https://doi.org/10.1071/wr19243

Jezierski, T., Gorecka-Bruzda, A., Walczak, M., Swiergiel, A. H., Chruszczewski, M. H., & Pearson, B. L. (2010). Operant conditioning of dogs (Canis familiaris) for identification of humans using scent lineup. *Animal Science Papers and Reports*, *28*(1), 81-93.

Johnen, D., Heuwieser, W., & Fischer-Tenhagen, C. (2017). An approach to identify bias in scent detection dog testing. *Applied Animal Behaviour Science*, *189*, 1-12.

Johnson, C., & Jarman, P. (1987). Macropod studies at Wallaby Creek. 6. A validation of the use of dung-pellet counts for measuring absolute densities of populations of macropodids. *Wildlife Research*, *14*(2), 139-145.

Jung, T. S., & Kukka, P. M. (2016). Influence of habitat type on the decay and disappearance of elk Cervus canadensis pellets in boreal forest of northwestern Canada. *Wildlife Biology*, *22*(4), 160-166.

Kerley, L. L. (2010). Using dogs for tiger conservation and research. *Integrative zoology*, *5*(4), 390-396.

Kerley, L. L., & Salkina, G. P. (2007). Using scent-matching dogs to identify individual Amur tigers from scats. *The Journal of Wildlife Management*, *71*(4), 1349-1356.

Kohn, M. H., York, E. C., Kamradt, D. A., Haught, G., Sauvajot, R. M., & Wayne, R. K. (1999). Estimating population size by genotyping faeces. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1420), 657-663.

Kretser, H., Glennon, M., Whitelaw, A., Hurt, A., Pilgrim, K., & Schwartz, M. (2016). Scat-detection dogs survey low density moose in New York. *Alces: A Journal Devoted to the Biology and Management of Moose*, *52*, 55-66.

Leigh, K. A., & Dominick, M. (2015). An assessment of the effects of habitat structure on the scat finding performance of a wildlife detection dog. *Methods in Ecology and Evolution*, 6(7), 745-752.

Long, R. A., Donovan, T. M., Mackay, P., Zielinski, W. J., & Buzas, J. S. (2007). Comparing scat detection dogs, cameras, and hair snares for surveying carnivores. *The Journal of Wildlife Management*, *71*(6), 2018-2025.

Long, R. A., Donovan, T. M., MacKay, P., Zielinski, W. J., & Buzas, J. S. (2011). Predicting carnivore occurrence with noninvasive surveys and occupancy modeling. *Landscape Ecology*, *26*, 327-340.

MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G., & Franklin, A. B. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, *84*(8), 2200-2207.

Mackenzie, D. I., & Royle, J. (2005). Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology*, *42*(6), 1105-1114. https://doi.org/10.1111/j.1365-2664.2005.01098.x

McKay, S. (2014). *Scent-matching of olfactory samples via proxy by domestic dogs Canis lupus familiaris* University of the Witwatersrand, Faculty of Science, School of Animal, Plant ...].

Nugent, G., Fraser, K., & Sweetapple, P. (1997). Comparison of red deer and possum diets and impacts in podocarp-hardwood forest, Waihaha Catchment, Pureora Conservation Park. *Science for conservation (Wellington, NZ)*, *50*(581.5223099338), 20.

O'Malley, T. D. (2023). *Rat and Possum Dynamics in an Eradication Landscape: Novel Tools, Movement Patterns and Device Interactions* ResearchSpace@ Auckland].

Oldenburg Jr, C., Schoon, A., & Heitkönig, I. M. (2016). Wildlife detection dog training: a case study on achieving generalization between target odor variations while retaining specificity. *Journal of Veterinary Behavior*, *13*, 34-38.

Paton, A. J., Buettel, J. C., & Brook, B. W. (2021). Evaluating scat surveys as a tool for population and community assessments. *Wildlife Research*, *49*(3), 206-214.

Reed, S. E., Bidlack, A. L., Hurt, A., & Getz, W. M. (2011). Detection distance and environmental factors in conservation detection dog surveys. *The Journal of Wildlife Management*, *75*(1), 243-251. https://doi.org/10.1002/jwmg.8

Robley, A., Gormley, A. M., Forsyth, D. M., & Triggs, B. (2014). Long-term and large-scale control of the introduced red fox increases native mammal occupancy in Australian forests. *Biological Conservation*, *180*, 262-269.

Rosell, F., Kniha, D., & Haviar, M. (2020). Dogs can scent-match individual Eurasian beavers from their anal gland secretion. *Wildlife Biology*, *2020*(2), 1-7.

Rust, L., Nizio, K. D., & Forbes, S. L. (2016). The influence of ageing and surface type on the odour profile of blood-detection dog training aids. *Analytical and bioanalytical chemistry*, *408*, 6349-6360.

Rust, L., Nizio, K. D., Wand, M. P., & Forbes, S. L. (2018). Investigating the detection limits of scent-detection dogs to residual blood odour on clothing. *Forensic Chemistry*, *9*, 62-75.

Schoon, G. (1996). Scent identification lineups by dogs (Canis familiaris): experimental design and forensic application. *Applied Animal Behaviour Science*, *49*(3), 257-267.

Schoon, G. (2005). The effect of the ageing of crime scene objects on the results of scent identification line-ups using trained dogs. *Forensic science international*, *147*(1), 43-47.

Sherley, G. (1992). SCIENCE & RESEARCH SERIES N0. 46.

Smith, D. A., Ralls, K., Cypher, B. L., Clark, H. O., Kelly, P. A., Williams, D. F., & Maldonado, J. E. (2006). Relative abundance of endangered San Joaquin kit foxes (Vulpes macrotis mutica) based on scat–detection dog surveys. *Southwestern Naturalist*, *51*(2), 210-220.

Smith, D. A., Ralls, K., Cypher, B. L., & Maldonado, J. E. (2005). Assessment of scat-detection dog surveys to determine kit fox distribution. *Wildlife Society Bulletin*, *33*(3), 897-904.

Smith, D. A., Ralls, K., Hurt, A., Adams, B., Parker, M., Davenport, B., Smith, M., & Maldonado, J. E. (2003). Detection and accuracy rates of dogs trained to find scats of San Joaquin kit foxes (Vulpes macrotis mutica). Animal Conservation Forum,

Sullivan, B. J., Baxter, G. S., & Lisle, A. T. (2002). Low-density koala populations in the mulgalands of south-west Queensland. I. Faecal pellet sampling protocol. *Wildlife Research*, *29*(5), 455-462. https://doi.org/https://doi.org/10.1071/WR00110

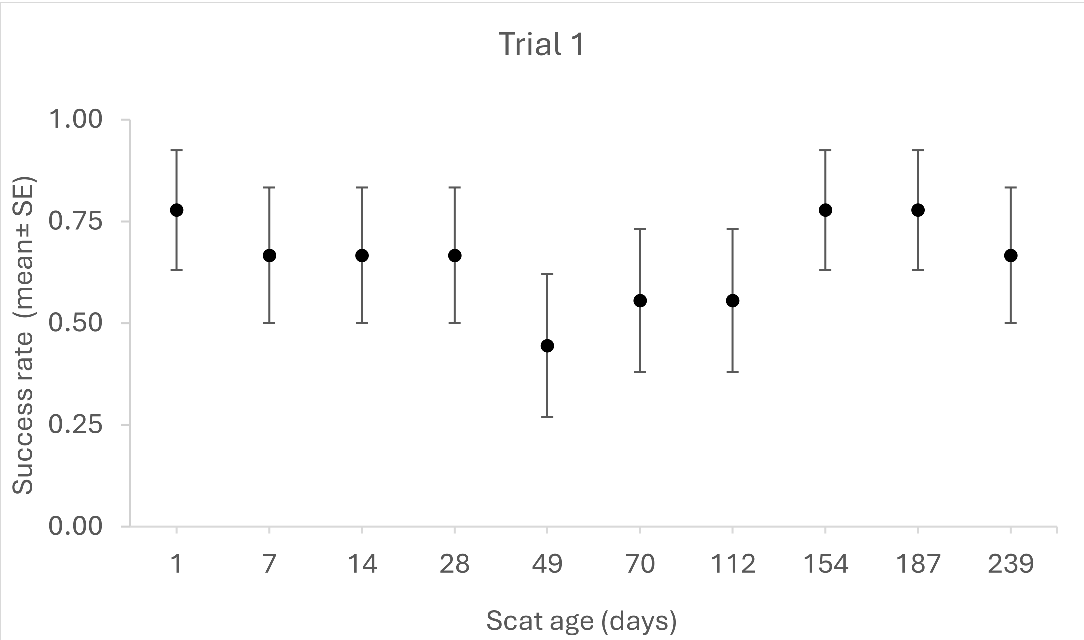
Sweetapple, P., & Nugent, G. (2009). Possum demographics and distribution after reduction to near–zero density. *New Zealand Journal of Zoology*, *36*(4), 461-471.
Sweetapple, P., & Nugent, G. (2020). *Refining detection and control tools and strategies for eradication of mainland low-density possum populations* (Contract Report: LC3819).
Thomas, M. L., Baker, L., Beattie, J. R., & Baker, A. M. (2020). Determining the efficacy of camera traps, live capture traps, and detection dogs for locating cryptic small mammal species. *Ecology and Evolution*, *10*(2), 1054-1068.
Thompson, C. M., Royle, J., & Garner, J. D. (2012). A framework for inference about carnivore density from unstructured spatial sampling of scat using detector dogs. *The Journal of Wildlife Management*, *76*(4), 863-871. https://doi.org/10.1002/jwmg.317
Tomaszewski, T., & Girdwoyn, P. (2006). Scent identification evidence in jurisdiction (drawing on the example of judicial practice in Poland). *Forensic science international*, *162*(1-3), 191-195.
Veale, A. J., & Etherington, T. R. (2023). *Assessing mustelid dispersal and the Predator Free*

Taranaki trapping programme using population genomics (Contract Report: LC4241).
Wallmo, O., Jackson, A., Hailey, T., & Carlisle, R. (1962). Influence of rain on the count of deer pellet groups. *The Journal of Wildlife Management*, *26*(1), 50-55.
Wasser, S. K., Hayward, L. S., Hartman, J., Booth, R. K., Broms, K., Berg, J., Seely, E., Lewis, L., & Smith, H. (2012). Using detection dogs to conduct simultaneous surveys of northern spotted (Strix occidentalis caurina) and barred owls (Strix varia).
Wasser, S. K., Smith, H., Madden, L., Marks, N., & Vynne, C. (2009). Scent‐matching dogs determine number of unique individuals from scat. *The Journal of Wildlife Management*, *73*(7), 1233-1240.
White, L. J. (1995). Factors affecting the duration of elephant dung piles in rain forest in the Lopé Reserve, Gabon. *African Journal of Ecology*, *33*(2), 142-150.
White, L. J. T. (2008). Factors affecting the duration of elephant dung piles in rain forest in the Lopé Reserve, Gabon. *African Journal of Ecology*, *33*(2), 142-150. https://doi.org/10.1111/j.1365-2028.1995.tb00789.x
Wiles, G. J. (1980). Faeces deterioration rates of four wild ungulates in Thailand. *Natural History Bulletin of the Siam Society*, *28*, 121-134.

APPENDIX: AGED SCAT TRIAL

The data was analysed identically but a pass number more than 1 was regarded as a fail even if a correct indication was given on a subsequent pass.

The results were all the same in terms of significance, but the success rates were slightly lower (Figure 20). Overall, the mean success from all dogs for Trial 1: M = 66%, SD = 47% (Range 44-78%), and Trial 2: M = 77%, SD = 43% (Range 50-100%).



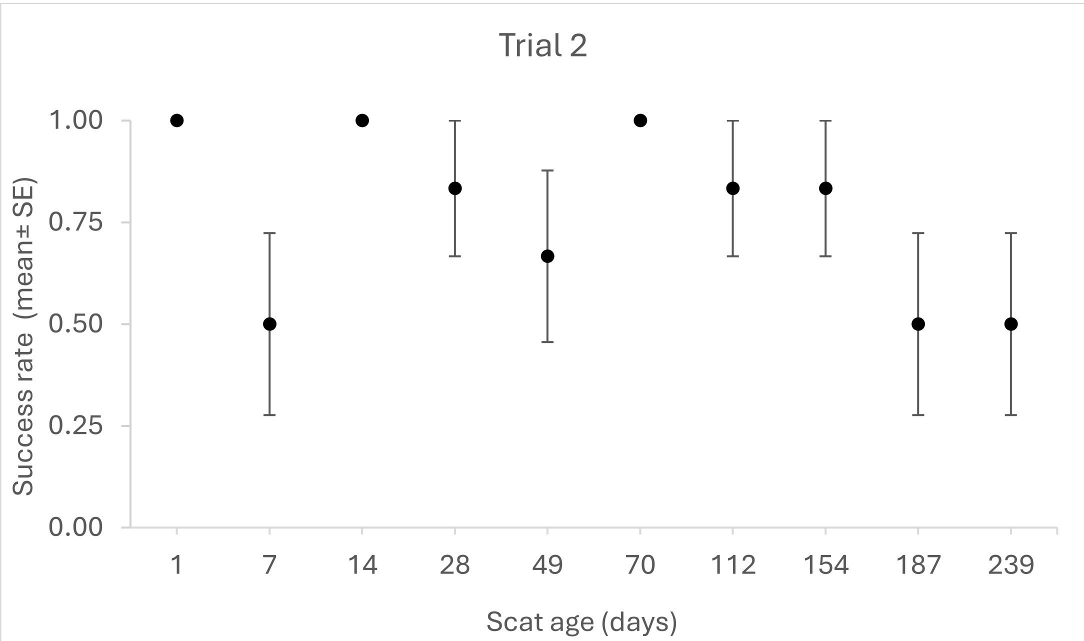


Figure 20. Plot of mean success from all dogs combined per scat age for Trial 1 (above), with Peggy, Mose, and Beau, and Trial 2 (below), with Peggy and Mose, where pass number > 1 was considered a fail ± standard error.

Across all scat ages, no statistically significant differences were found between the trials. However, it should be noted that the small sample sizes, high number of ties, and low variability in success rates likely reduced the statistical power to detect any meaningful differences.

There was no significant difference in overall success rates between scat ages (for all dogs) in Trial 1 (Fisher's exact test: $p = 0.933$) or Trial 2 (Fisher's exact test: $p = 0.129$). There was still no significant in performance between dogs (Peggy, Mose, and Beau) in Trial 2 (Fisher's exact test: $p = 0.433$) or Peggy and Mose in Trial 2 (Fisher's exact test: $p = 0.360$).

There was still no significant difference in overall success rate between trials for Peggy (Trial 1: $M = 73\%$, $SD = 12\%$; Trial 2: $M = 70\%$, $SD = 14\%$; Wilcoxon rank-sum test: $W = 53$, $p = 0.840$), but there was a very close to significant difference for Mose (Trial 1: $M= 56\%$, $SD = 10\%$; Trial 2: $M = 83\%$, $SD = 13\%$; Wilcoxon rank-sum test: $W = 25$, $p = 0.046$), which could indicate a learning effect. There was no significant difference in success by age of scat for either Peggy (Fisher's exact test: Trial 1: $p = 0.6$, Trial 2: $p = 0.44$) or Mose (Fisher's exact test: Trial 1: $p = 0.44$, Trial 2: $p = 1.0$).