



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Bennett, EM;Hauser, CE;Moore, JL

Title:

Evaluating conservation dogs in the search for rare species

Date:

2020-04-01

Citation:

Bennett, E. M., Hauser, C. E. & Moore, J. L. (2020). Evaluating conservation dogs in the search for rare species. *Conservation Biology*, 34 (2), pp.314-325. <https://doi.org/10.1111/cobi.13431>.

Persistent Link:

<https://hdl.handle.net/11343/275503>

Evaluating conservation dogs in the search for rare species

1. Emma M Bennett (corresponding author)

Monash University
23 Suburban St, Clunes, VIC 3370
emma.bennett@monash.edu

2. Cindy E Hauser

University of Melbourne
Parkville, VIC 3052

Arthur Rylah Institute for Environmental Research
Dept of Environment, Land, Water & Planning
123 Brown St, Heidelberg VIC 3084

Joslin L Moore
Monash University
Moore Plant Ecology and Conservation Group
Scenic Blvd, Clayton, VIC 3800

Running head: Conservation dogs

Keywords: threatened species, invasive species, eradication, scent, detection dog, cost-benefit analysis, cost-effectiveness

Article impact statement

Key performance measures of the effectiveness of conservation scent dogs include precision, sensitivity, and effort.

Abstract

Detecting rare species is important for both threatened species management and invasive species eradication programs. Conservation scent dogs provide an olfactory survey tool which can have advantages over traditional visual and auditory survey techniques for some cryptic species. From the literature, we identified five measures important for evaluating the use of scent dogs to address conservation objectives: precision, sensitivity, effort, cost and comparison with other techniques. We explore the scale at which performance is

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/cobi.13431](https://doi.org/10.1111/cobi.13431).

This article is protected by copyright. All rights reserved.

evaluated and present examples where field testing under real working conditions is achievable. We also identify significant cost variations between studies, which we attribute to the differences in how the dog and handler are sourced for the study. We provide a framework to address inconsistencies in reporting and recommend guidelines to assist future studies to develop consistent reporting standards. We examine 61 recent studies reporting conservation dog performance, highlighting inconsistencies in the reporting of scent dog performance and find that only four studies reported all five measures. We present consistent reporting guidelines for the key factors that influence evaluation; specifically performance measures and cost in addition to survey objectives, dog training, type of detection task and human influences.

Introduction

Trained scent dogs are increasingly deployed to support conservation efforts, particularly in the search for cryptic species. Dogs and humans have been working together for over 20,000 years (Ruusila & Pesonen 2004), however it is only in the past few decades that the scientific literature has begun to quantify their assistance to the conservation sector. Unlike other methods used to survey for plants and animals, dogs employ their olfactory senses to locate and detect hard-to-find species and provide an alternative to the more common visual and auditory approaches. Whilst their acute sense of smell is useful for detection work, it is likely the ability of the dog to discriminate between odours which provides the detection advantage (Syrotuck 2000).

Dogs have been used to support conservation efforts since the late 19th century when they were first used to find the kiwi in New Zealand (Hutching & Walrond 2007). Later research reported on hunting dogs finding birds for conservation efforts in grasslands (Gutzwiller 1990; Novoa et al. 1996) and scent dogs detecting cryptic entomological pests (Welch 1990; Lewis et al. 1997). From early this century, advances in DNA analysis have seen an increase in the use of dogs for scat detection (Smith et al. 2003; Wasser et al. 2004; Smith et al. 2005). Most recently detection dogs have been used to survey low density species (Kretser et al. 2016; Becker et al. 2017), search in difficult to navigate terrain (Orkin et al. 2016), where alternative methods are invasive

(Duggan et al. 2011), within eradication programs (Hurt et al. 2014) and where other survey techniques are not available (Rolland et al. 2006).

Standardised performance testing of conservation dogs across the sector is difficult due to the range of conservation objectives and different environmental conditions within which conservation dogs work.

However, understanding which targets are likely to be detected and which remain undetected is crucial for reliably interpreting survey data (Hauser & Moore 2016), allocating resources for search efforts (Hauser & McCarthy 2009), establishing presence/absence (Garrard et al. 2008), maximising detection opportunities (Garden et al. 2007) and developing spatial distribution data (Chen et al. 2013). Determining the probability of detection for each target odour at each different site is expensive and time consuming (MacKenzie et al. 2002), with approximately 250 detection opportunities needed to estimate detection probability reliably (Glen & Veltman 2018). This is not practical or even possible for many rare conservation targets. There is also little quantitative peer-reviewed information about dog performance under a range of real survey or environmental conditions (Clare et al. 2015). In a systematic review of all scent dog literature, Johnen et al. (2017) found that precise information on test criteria such as reliability, validity and objectivity of scent detection performance tests were not reported in most publications even though this information was essential to compare findings of similar studies. Whilst they made recommendations for standardising testing and reporting, the authors recognised the difficulty of adequate testing in free searching scenarios typical in conservation work.

Understanding the financial cost as well as the performance of a scent dog is essential for cost-effective planning. Within conservation programs, competing priorities and limited resources mean that actions must be prioritised to maximise the return on investment, by reducing future costs and optimising conservation outcomes (Possingham 2001). For this reason, the most cost-effective solution within this context may be different to the highest performing solution (Cook et al. 2017). It follows that understanding the financial requirements of detection dogs in comparison to existing and alternative survey methods is important when considering their contribution to conservation efforts.

In summary, evaluating the effectiveness of conservation detection dogs to search for rare species must take into consideration the performance of dogs to detect the target species, the financial cost of their deployment and how dogs compare to alternative survey methods. We present a framework for the evaluation of conservation dogs and explore the use of three performance measures and how they can be used to standardise the reporting of conservation scent dog performance. We examine the reported financial costs of engaging dogs and consider how dogs have performed in comparison with other search techniques. We summarise the reporting of performance, costs and comparative studies within the current literature and offer recommendations for standardising future reporting of conservation dog studies.

A framework for evaluation

To evaluate the effectiveness of detection dogs there are a range of factors which need consideration (Fig. 1). In this framework, we limit our focus to the evaluation of dogs already engaged in conservation work or considered trained by their handlers. We identified five aspects of detection dogs necessary for comprehensive evaluation: precision, sensitivity, effort, cost and comparison with other methods. We discuss how standardising performance (precision, sensitivity and effort), provides useful quantitative measures when comparing studies. We examine the efficiency and financial cost of dogs and the benefits of using dogs given particular conservation objectives. Finally, we discuss the comparison of dogs with alternative survey techniques and how this can be considered in evaluating success.

Performance

There are no consistent or agreed-upon measures for the reporting of scent dog performance. Studies of medical scent dogs refer to specificity and sensitivity (Johnen et al. 2013) and explosive scent dog literature includes detection percentage and duration (Gazit & Terkel 2003). Within the conservation scent dog literature, Cablk and Heaton (2006) identified accuracy and reliability as useful measures for reporting conservation dog performance, however, only a few subsequent papers have adopted this terminology. We propose using evaluation measures that can be calculated from the terms from a confusion matrix (Allouche et

al. 2006; Saito & Rehmsmeier 2015) and recommend that precision, sensitivity and effort span the important quantitative dimensions of conservation dog performance.

Precision

We define precision as the proportion of all alerts that are directed towards a true target (Table 1), which is an equivalent measure to positive predictive value (Saito & Rehmsmeier 2015) and is often termed reliability in conservation dog literature. An alert is a trained response or behaviour change given by the dog to indicate a find and varies between dogs and trainers. Precision provides confidence that the dog is searching for the correct scent and an experienced scent dog is expected to have very high precision, often above 90% (Marchal et al. 2016). Precision is most important for scent matching tasks, where dogs are required to indicate matching scent between visually indistinguishable samples (Kerley & Salkina 2007). It is critical to understand the precision of the dog when relying on scent matching however, precision may be less important where targets can be readily identified by the handler once detected.

Precision demonstrates how well the dog distinguishes the target scent from other odours. This can be measured during simple odour discrimination testing, simple multiple-choice tests where dogs are exposed to identical containers housing target and non-target scents and instructed to indicate the target odour (Porrirt et al. 2015; Marchal et al. 2016). Initial odour discrimination tests are routinely performed during training exercises and attempts have been made to offer standardised procedures (Porrirt et al. 2015). The relevance of simple or indoor odour tests to field based surveys is not well understood, although field based odour discrimination tests have been reported in the conservation literature (Mathews et al. 2013; Leigh & Dominick 2015).

Sensitivity

Sensitivity is the proportion of targets found relative to the total number of targets available (Table 1), which is also termed the true positive rate or recall (Saito & Rehmsmeier 2015). Sensitivity is often termed accuracy within the detection dog literature, which differs from the definition present in the broader scientific literature. Whilst we recognise that the term accuracy is more widely used within conservation dog studies, we

recommend the term sensitivity as it has a clear definition as a basic evaluation measure that is less likely to be misinterpreted.

No detection method currently available for rare species searches is known to locate 100% of all possible targets even with repeat sampling (MacKenzie et al. 2002). Employing a combination of different survey methods often results in higher detection rates than any single method alone (Garden et al. 2007). In some contexts, such as pest or invasive species eradication, the goal is to find all of the target species and thus aim for 100% detection over the life of the program (Hurt et al. 2014). However, many other conservation programs rely on tools with very low detection rates and an sensitivity of 25, 50 or 75% may be an improvement on alternative methods (Mathews et al. 2013).

Cryptic species are by nature difficult to find and there is a risk they will go undetected during surveys. It follows that designs and analysis of survey data should account for detection probability (Hauser et al. 2016). However, measuring sensitivity can be difficult in field-based searches due to the variability of the environments in which conservation projects occur. For example, dogs have demonstrated their capacity to detect whale faeces in the ocean (Rolland et al. 2006) , however, whilst attempts have been made to quantify sensitivity, the challenges of the marine environment and the cost of deploying the dog team and boat make testing unfeasible. Similar challenges can be faced where dogs are searching for live threatened species where limited access and ethical standards prevent the use of live targets for testing purposes. In studies with easily obtainable targets such as scats or plants, quantifying sensitivity has been relatively straightforward through detection experiments (Moore et al. 2011; Hauser et al. 2016).

Effort

Effort is the time spent searching a unit area or transect. It is important for managers to know how long it takes to survey a given area to assist in planning, budgets and logistics (McCarthy et al. 2010). Whilst both active search time and total project time is important for planning, active search time is most important for detection efficiency. Increasing search effort will lead to a higher probability of detection and thus reporting sensitivity relies on understanding the effort it took to achieve (Garrard et al. 2008). A number of factors are

likely to influence the time needed to survey an area, including vegetation type, terrain, weather, the breed of the dog, and the experience and physical fitness of the dog and handler team.

Quantifying effort and understanding how this may impact sensitivity is important for understanding performance and efficiency. In a study looking at the effect of vegetation on detection rates, Leigh & Dominick (2015) found that dogs needed to search longer in denser vegetation to achieve consistent sensitivity. This suggests that changes in vegetation and perhaps varied environmental conditions need appropriate changes to search effort to maintain consistent sensitivity. Understanding the impact of effort may also lead to changes in study designs to maximise the effectiveness of different survey methods. Studies have shown that for the same time taken, searching a greater area with easier accessibility can provide more opportunities for detection than searching a smaller area with reduced accessibility or denser vegetation more thoroughly (Hauser & McCarthy 2009; Huso & Dalthorp 2014).

Efficient detection performance

Efficiency is a measure of the benefit per unit effort and is crucial for resource allocation and effective survey design (Garrard et al. 2008; Hauser & McCarthy 2009; McCarthy et al. 2010). This can be a useful measure when comparing dogs within a study to understand which teams are able to achieve similar detections with less effort and therefore able to survey a larger area given the same time. This can be expressed as a function of dog performance using the relationship:

$$\text{Efficiency} = \frac{\text{Sensitivity}}{\text{Effort}}$$

where sensitivity is the proportion of all targets that were found (Table 1), and effort is the time spent searching divided by the area searched (e.g. hours per ha).

The relative importance of efficiency in evaluating performance depends on the value of sensitivity to the conservation objectives, the size of the area to be surveyed and the time constraints of the project.

Scale of performance evaluation

The measures of precision, sensitivity and effort provide valuable standards that can be reported for both training and in-field assessments (Cablak & Heaton 2006). The quantification of these measures can occur at different scales and it is likely that these measures are not independent of scale. Whilst the scale of field tests often varies between studies, how this impacts detection performance is not well understood. Further studies that address the relationship between simple performance tests and in-field performance would be useful to understand the predictive power of small tests for actual field performance.

The relationship between odour discrimination performance and field based search performance is not well understood. In a single blind study looking at the effects of vegetation on scat detection by dogs, initial odour discrimination tests recorded 100% for both sensitivity and precision, whilst dogs undertaking field based searches detected 83-87% ($n=120$) of targets (Leigh & Dominick 2015). Very few studies have reported both odour discrimination results and free ranging field survey results to allow for comparison. Typically, dogs only continue to field based surveys once they have demonstrated high precision during training and initial testing (often 100%), yet reported sensitivity in the field ranges from as low as 20% (Wasser et al. 2012) to as high as 100% (Cristescu et al. 2015). Undertaking field based detection experiments, which mimic real search conditions, is essential to understand how working duration impacts performance, how formal testing differs from *in situ* evaluation and what targets are likely to be found or missed during surveys.

A good example is how field based testing for conservation dogs has become common practice within the wind farm industry (Arnett 2006; Bennett 2015). Monitoring and reporting of bird and bat fatalities at windfarms is often mandatory and dogs are considered the best practise monitoring tool for carcass detection at windfarms (Paula et al. 2011). As part of management obligations, the sensitivity of carcass detection by dogs in the field is routinely tested and reported several times per year *in situ*. This allows for sensitivity and effort to be quantified under real working conditions providing managers with realistic detection rates and field time expectations. Whilst easily achievable within the wind industry, many conservation programs have complexities which make in-field testing more challenging.

Cost considerations of using a dog

Prioritising management actions must consider the likely benefits of an action and its associated financial and technical constraints. This is important as the most cost-effective solution may differ from the most effective solution, leading to different management outcomes (Cook et al. 2017). Conservation planning generally involves making decisions about where to prioritise investments for the best conservation outcomes and in order to achieve this, the return on investment needs to be understood (Moore et al. 2004; Joseph et al. 2009). Understanding the costs of engaging conservation dogs will help determine if any improvements in detection are worth the financial and temporal investment.

The different costs of purchasing or leasing a dog and its influence on the total project cost was demonstrated by Long et al. (2007a), who determined that purchasing a dog at the beginning of the study could have saved almost 20% of total project costs. Reporting costs provides an understanding of the financial and logistical investment required to use detection dogs and should be considered over the lifetime of the dog when purchasing rather than leasing. Where expenses are itemised in categories, comparisons with alternative techniques and other studies are possible (Clare et al. 2015; Glen et al. 2016). Summarising project costs in terms of project objectives (e.g. cost per scat) is interesting (Arandjelovic et al. 2015; Orkin et al. 2016), but for comparative purposes it is perhaps more useful to report in terms of achievement (total project cost, total area surveyed, probability of detection). Understanding the cost of engaging a detection dog is more useful when the costs of alternative techniques are also considered. Itemising the various costs in terms of dog costs, labour, transport and equipment are particularly useful when publishing studies as prices vary globally.

Comparison with alternative methods

Comparing conservation dogs with other survey tools is essential to find the most effective and/or cost-effective method to survey for rare species (DeMatteo et al. 2019). Comparing with alternative methods must take into consideration the conservation objectives and the project budget. Long et al. (2007a) compared the cost and performance of detection dogs with camera traps and hair snares. They determined that whilst dogs required greater total financial investment, they had a much higher probability of detection for forest carnivores than either camera traps or hair snares and if the same effort necessary to achieve a similar high

probability of detection is accounted for, dogs may be more cost effective. Conversely, Arandjelovic et al. (2015) demonstrated that dogs were faster and found more gorilla scat than people, but the cost per scat found was significantly higher and for the same expense, searches by humans could collect 7 times more samples, albeit over a much longer time period. Understanding the benefits and limitations, particularly the performance and costs of different methods, allows for better comparisons between survey tools. This provides a deeper understanding of how and when conservation dogs may be advantageous over alternative tools.

Relevance to conservation objectives

Whilst quantifying precision, sensitivity, effort, cost and dog performance relative to other methods are all important, their relative importance will vary depending upon the conservation objectives and project constraints (Fig 2). For example, precision becomes less important if the target is easily identified by the handler once found by the dog (Brook et al. 2012) and becomes very important where dogs seek to scent match individuals from visually indistinguishable samples (Kerley & Salkina 2007). Similarly, high sensitivity is vital to eradication programs (Hurt et al. 2014) but sensitivity may not be important where a single detection confirms the target species presence and triggers conservation action (Scheffers et al. 2011). Effort is important for estimating efficiency (Garrard et al. 2008) and provides context for sensitivity. Efficient performance and being able to search large areas within a limited time frame is important within industrial work sites (Paula et al. 2011) but time may not be a critical factor when working with volunteers (Browne et al. 2015). Cost considerations are more important where budgets are limited (Orkin et al. 2016) and less important within industrial and eradication programs (Cherry et al. 2016). The comparison with alternative survey tools, however, is an important consideration for most programs, except where no other detection methods are available.

Consistency of Reporting within the Literature

Standardised reporting enables more robust comparisons among studies, facilitates national and global reviews, enables greater ease of study replication, and leads to improved wildlife research and management

outcomes (Meek et al. 2014). We reviewed recently published studies which reported quantitative information on the performance or cost of detection dogs working on conservation projects. Our aim was to summarise the consistency in reporting and to provide recommendations to conservation managers and researchers for standardised data collection and reporting.

Literature searched

We conducted a literature search using various combinations of the following words: conservation dog, detection dog, working dog, scent dog, scat detection, rare species detection, effective detection, survey techniques, wildlife dog and wildlife detection. Sources searched include JSTOR, Google Scholar and Monash University Library peer reviewed journal search database. Reference lists from each of these papers were evaluated and additional papers included as deemed relevant. A total of 82 relevant studies were sourced. Twenty-one studies were removed as they lacked quantitative information on detection performance and did not include or discuss the financial cost of the dog teams. For the remaining 61 studies information on performance, costs, handler model, bias, comparison with other methods, number of dogs in the study, location, target type and target species was tabulated (Supplementary Information).

Reported performance

Whilst more than half of all the studies reviewed quantified either precision, sensitivity or efficiency, very few reported these measures simultaneously (Fig 3a). Reported precision for dogs in the conservation literature ranged from 13% to 100%, and 20% to 100% for sensitivity. Although there was a large variation in the reported ranges, the majority of studies demonstrated high performance with the median value for precision across the studies being 94% and the median value for sensitivity 84% (Fig 3b).

Whilst effort or efficiency was reported in 43 of the studies, there was little consistency between studies regarding the format and presentation of effort as a measure. Of those studies which reported effort in some form, nine presented time to detection, but only two of the nine presented time to detection per unit area over which the search occurred (i.e. mins/ha). Another three studies reported effort relative to alternative techniques (e.g. 5 times faster than humans), whilst number of targets detected per unit of effort was

reported in eight studies (i.e. captures per km transect, per unit of time, per unit area). Notably, Clare et al. (2015) reported the average daily transect distance achieved by the dogs with a calculated probability of detection for each day and Long et al. (2007b) calculated the search time required to achieve an 80% probability of detection. The remaining studies reported working time, such as four hours per day, but did not indicate the area or transect distance achieved during that time.

More than half of the studies did not report if the survey design was randomised or blind (Fig 3c) with just 11 studies reporting that they were both randomised and blind. Only two studies reported double blind methodology in performance testing and both reported high precision of 75% and 97%, as well as high sensitivity above 85% and 90% (Hoyer-Tomiczek et al. 2016; Matthew 2016).

Whilst previous experiences of the dogs are considered and discussed in the literature, the experience of the handler is rarely reported. There is some evidence that the handler plays a significant role in detection success (Long et al. 2007b; Lit et al. 2011; Jamieson et al. 2018; DeMatteo et al. 2019) however the influence of the handler on the dog's performance is poorly understood and research into the effects of handler experience or handler characteristics on conservation detection dog outcomes would be useful (Beebe *et al.* 2016).

Reported cost

Of the studies reviewed, only 17 reported costs in any form, with the details ranging from full itemisation down to a single sentence. A further twelve mentioned the importance of costs, but provided no detail or elaboration, whilst the remaining studies did not mention cost at all. Interestingly, 10 of the 17 studies which reported costs were published since 2014 and may reflect a growing trend in cost reporting. Perhaps the most useful of the studies which reported costs are those which compare detection dog costs to alternative techniques and provided a full itemisation of the expense categories. Providing itemisation and the rates used to calculate estimates is particularly useful when considering the global variation in prices for wages, equipment and dog handlers. Only four studies reported all of the performance measures, costs and comparison with alternative field methods (Reindl-Thompson et al. 2006; Duggan et al. 2011; Chambers et al. 2015; Clare et al. 2015).

In three separate studies, the cost comparison between camera traps and detection dogs differed substantially, with Clare et al. (2015) reporting that dogs cost three times more than cameras in the search for bobcats in the USA, Long et al. (2007a) reporting dogs costing 1.5 times more than cameras for forest carnivores and Glen et al. (2016) finding cameras and detection dogs to be similar in cost when searching for feral cats in New Zealand. Differences in wages for personnel between countries may account for some of the variation reported, however the cost of the dog and handler were also substantially different between the studies. This difference in dog handler cost is most profoundly highlighted where professional dog handlers from the USA were engaged to work in Africa for 3 months at a total cost of \$98,000 USD (Arandjelovic et al. 2015), compared to a local police dog handler employed in China in a similar study for a total cost of \$2,851 USD (Orkin et al. 2016).

The cost differences associated with engaging dogs and handlers from different sources was raised by Nussear et al. (2008) and Long et al. (2007a), though it is not discussed further in more recent literature. A range of different models for dog and handler were mentioned in the studies reviewed although the impact to cost was not evaluated. These include:

- Leasing or purchasing a dog from a professional conservation detection dog supplier
- Contracting a professional conservation detection dog handler and their dog
- Partnering with a local police dog service
- Purchasing one's own dog and contracting professional training and support
- Purchasing one's own dog and self-training
- Working with dog clubs and volunteers to train for scent detection.

Ethical considerations

A number of studies purchased dogs either from shelters or from professional sources for the purpose of the study, with dogs sometimes working with multiple handlers. However most studies failed to mention the ongoing care and living arrangements of the dogs once the study concluded. Such factors can influence the long-term costs of a project and can raise additional animal welfare concerns. In New Zealand, all dogs and handlers used in conservation are certified and registered with the Government Department of Conservation

(DOC) and issues of welfare plus continuity of care post-retirement are the responsibility of the handler under legislation. In Australia, a number of state governments have purchased dogs for the purpose of eradication programs although an internet search failed to find any reference to care after retirement or indeed any guidelines concerning their welfare. Orkin et al. (2016) highlights the advantages of engaging a local police officer and his police dog for their study in China in terms of care and housing for the dog, in addition to logistical advantages such as facilitating travel and interactions with permit officials and field assistants. One study in New Zealand which utilised 20 dogs, worked with members of two dog clubs to undertake odour discrimination on two species of reptiles as a test of feasibility for future field surveys (Browne et al. 2015). By using volunteers, they deferred responsibility of the dogs to private handlers whilst still demonstrating high sensitivity in odour discrimination (above 90%).

Comparative studies

Twenty-seven studies reviewed considered the effectiveness of conservation dogs as a tool compared with alternative survey methods. Notably, early research by Reindl-Thompson et al. (2006), Long et al. (2007a), Nussear et al. (2008) and Cablk and Heaton (2006) focussed on comparing detection dogs with existing search methods and set a precedent for reporting in conservation dogs studies which has not been maintained in more recent years.

Eleven studies compared dogs with human surveyors, although three of these simply made limited comments without providing supporting evidence, e.g. that dogs can reduce time and cost compared to human based surveys (Nielsen et al. 2016), dogs are five times faster than humans (Bennett 2015) and dogs are 19 times faster than humans (Cristescu et al. 2015). Hurt et al. (2014) utilised dogs to search for an invasive plant in areas already searched by human surveyors and found that dogs detected 42% of plants overlooked by humans and were able to detect them prior to flowering. In one of the few blind and randomised trials, Mathews et al. (2013) found sensitivity for bat carcasses in varying grass lengths was 75% for the dog team (n=61) and just 20% for human searches (n=60), reporting similar results to an earlier study by Arnett (2006).

Two studies investigated the use of dogs as an alternative to live trapping. Duggan et al. (2011) found that in their search for cryptic rodents less than 1 hour of survey by two trained dogs was equivalent to twice daily

live trapping surveys but that the use of dogs was more expensive. More recently, McGregor et al. (2016) found that the use of dogs to locate, track and bail (i.e. contain in a stationary position) feral cats was four times more efficient than leg hold traps and allowed for higher capture rates. Furthermore, they found that the use of dogs did not result in any injuries to the cats, whereas trapping has the additional risk of mortality or debilitating injury.

Recommendations for future reporting

In order to evaluate the effectiveness of any conservation action, there must be a clear research hypothesis or management goal against which to benchmark success (Lindenmayer & Likens 2010). This remains true for evaluating the effectiveness of conservation detection dogs. It is also important to consider how previous experience of both the dog and handler may impact performance and is useful to include when reporting performance measures (DeMatteo et al. 2019). To facilitate comparisons between studies and to help determine if methods from one study can be applied successfully to other environmental contexts, we make the following recommendations for future reporting (Table 2).

We recommend measuring performance based on the three quantitative measures of precision, sensitivity and effort. Whilst there are a range of scales available to undertake performance testing, lessons from the wind farm industry demonstrate that in-field testing occurring under real survey conditions with known locations of targets is achievable on a routine basis for some targets (Paula et al. 2011). Studies that investigate if small scale field tests can provide a valuable measure of in-field performance would be useful for many conservation programs as they can have logistical advantages over large scale and field trials and may be the only feasible option for some targets.

Future research would benefit if both precision and sensitivity are presented as the proportion of the total sample size detected. Presenting information within a confusion matrix so that confidence intervals can be calculated from sample size would provide consistent quantification of detection success, which is both useful for analysis and easily understood by practitioners. To ensure scale is considered when interpreting results, it

is our recommendation that effort is reported as area (or transect) per unit of time. Additionally, reporting environmental variables, such as weather, vegetation type and topography, alongside effort is important for understanding the influence of search conditions on search time. Reporting effort in this way provides a practical guide for planners about the time required to achieve search area objectives and will help in forecasting project budget costs.

As with many researchers in the conservation sector (Armsworth 2014; Evans et al. 2015; Cook et al. 2017), we encourage the reporting of project costs, particularly where detection dogs are compared to alternative survey techniques. When reporting costs, it would also be useful to specify the type of handler and dog (professional/amateur/purchase/lease) and to itemise expense categories to enable comparisons with other projects and survey techniques to be considered. We also recommend reporting the environmental conditions under which the search is conducted, information around the target scent and any proxies used as well as design parameters such as blinding and randomisation.

Discussion

Our framework identifies key measures to evaluate performance of conservation scent dogs and, if adopted in future reporting, would allow for comparison between studies. Our review of the recent literature has shown that reporting of performance for conservation detection dog studies has been varied and inconsistent which can lead to poor comparisons between studies, incorrect conclusions around performance results and makes replication or improvements of the study design difficult. However much of the information is readily available and it is awareness rather than degree of difficulty which is restricting the reporting of these measures.

With the increase in use of scent dogs in conservation research there are deeper questions that will need specific research efforts to address and are promising avenues for further research. Studies which quantify the effect of spatial scale on detection performance could reveal the assessment designs and intensity required to estimate precision and sensitivity in free ranging field surveys. There are also interesting questions about how the experience of the handler and dog influences performance and if the inclusion of costs in the evaluation

makes the most cost-effective solution different from the most effective. As dog and handler selection may have a large impact on project cost and effectiveness, we suggest further research that examines the advantages and disadvantages of various handler models alongside evaluation of field performance and cost. Further studies, which seek to clarify consistent field methods for performance evaluation, would also be welcome.

Scent dogs are a promising tool for conservation monitoring with the potential to increase detection opportunities for rare and cryptic species however evaluation is crucial if they are to be effectively integrated in conservation monitoring programs.

Acknowledgments

We thank A. Camaclang for her supportive and insightful comments on a previous manuscript. We thank the three anonymous reviewers and regional editor for their valuable feedback and suggestions. C.E.H. was supported by an ARC Discovery Project DP160100745. E.M.B. was supported by a Monash University RTP Stipend.

Supporting Information

A summary table of the 82 publications sourced and 61 publications reviewed (Appendix S1) is available online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature Cited

- Allouche O, Tsoar A, Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43**:1223-1232.
- Arandjelovic M, Bergl RA, Ikfuingei R, Jameson C, Parker M, Vigilant L. 2015. Detection dog efficacy for collecting faecal samples from the critically endangered Cross River gorilla (*Gorilla gorilla diehli*) for genetic censusing. *Royal Society open science* **2**:140423.
- Armsworth PR. 2014. Inclusion of costs in conservation planning depends on limited datasets and hopeful assumptions. *Annals of the New York Academy of Sciences* **1322**:61-76.
- Arnett EB. 2006. A preliminary evaluation on the use of dogs to recover bat fatalities at wind energy facilities. *Wildlife Society Bulletin* **34**:1440-1445.
- Becker MS, et al. 2017. Using dogs to find cats: detection dogs as a survey method for wide-ranging cheetah. *Journal of Zoology* **302**:184-192.
- Bennett E. 2015. Observations from the use of dogs to undertake carcass searches at wind facilities in Australia. Pages 113-123. *Wind and Wildlife*. Springer.

- Brook SM, et al. 2012. Integrated and novel survey methods for rhinoceros populations confirm the extinction of *Rhinoceros sondaicus annamiticus* from Vietnam. *Biological Conservation* **155**:59-67.
- Browne CM, Stafford KJ, Fordham RA. 2015. The detection and identification of tuatara and gecko scents by dogs. *Journal of Veterinary Behavior: Clinical Applications and Research* **10**:496-503.
- Cablk ME, Heaton JS. 2006. Accuracy and reliability of dogs in surveying for desert tortoise (*Gopherus agassizii*). *Ecological Applications* **16**:1926-1935.
- Chambers CL, Vojta CD, Mering ED, Davenport B. 2015. Efficacy of scent-detection dogs for locating bat roosts in trees and snags. *Wildlife Society Bulletin* **39**:780-787.
- Chen G, Kéry M, Plattner M, Ma K, Gardner B. 2013. Imperfect detection is the rule rather than the exception in plant distribution studies. *Journal of Ecology* **101**:183-191.
- Cherry H, Constantine A, Primrose K, Hauser C, Tasker K. 2016. It takes a village: detection dogs, partnerships and volunteers aid hawkweed eradication in mainland Australia. Pages 164-170. *Proceedings of the 20th Australasian Weeds Conference*, eds R. Randall, S. Lloyd and C. Borger.
- Clare JD, Anderson EM, Macfarland DM, Sloss BL. 2015. Comparing the costs and detectability of bobcat using scat - detecting dog and remote camera surveys in central Wisconsin. *Wildlife Society Bulletin* **39**:210-217.
- Cook CN, Pullin AS, Sutherland WJ, Stewart GB, Carrasco LR. 2017. Considering cost alongside the effectiveness of management in evidence-based conservation: A systematic reporting protocol. *Biological Conservation* **209**:508-516.
- Cristescu RH, Foley E, Markula A, Jackson G, Jones D, Frère C. 2015. Accuracy and efficiency of detection dogs: a powerful new tool for koala conservation and management. *Scientific Reports* **5**:8349.
- DeMatteo KE, Davenport B, Wilson LE. 2019. Back to the basics with conservation detection dogs: fundamentals for success. *Wildlife Biology* **1**.
- Duggan JM, Heske EJ, Schooley RL, Hurt A, Whitelaw A. 2011. Comparing detection dog and livetrapping surveys for a cryptic rodent. *The Journal of Wildlife Management* **75**:1209-1217.
- Evans MC, Tulloch AIT, Law EA, Raiter KG, Possingham HP, Wilson KA. 2015. Clear consideration of costs, condition and conservation benefits yields better planning outcomes. *Biological Conservation* **191**:716-727.
- Garden JG, McAlpine CA, Possingham HP, Jones DN. 2007. Using multiple survey methods to detect terrestrial reptiles and mammals: what are the most successful and cost-efficient combinations? *Wildlife Research* **34**:218-227.
- Garrard GE, Bekessy SA, McCarthy MA, Wintle BA. 2008. When have we looked hard enough? A novel method for setting minimum survey effort protocols for flora surveys. *Austral Ecology* **33**:986-998.
- Gazit I, Terkel J. 2003. Explosives detection by sniffer dogs following strenuous physical activity. *Applied Animal Behaviour Science* **81**:149-161.
- Glen AS, Anderson D, Veltman CJ, Garvey PM, Nichols M. 2016. Wildlife detector dogs and camera traps: a comparison of techniques for detecting feral cats. *New Zealand Journal of Zoology* **43**:127-137.
- Glen AS, Veltman CJ. 2018. Search strategies for conservation detection dogs. *Wildlife Biology* **2018**:wlb.00393.
- Gutzwiller KJ. 1990. Minimizing Dog-Induced Biases in Game Bird Research. *Wildlife Society Bulletin (1973-2006)* **18**:351-356.

- Hauser C, Moore JL. 2016. Estimating detectability using search experiments. Pages 71-75 in Wilson J, Pantta FD, and Lindgren C, editors. *Detecting and Responding to Alien Plant Incursions*. Cambridge University Press, Cambridge, UK.
- Hauser CE, McCarthy MA. 2009. Streamlining 'search and destroy': cost-effective surveillance for invasive species management. *Ecology Letters* **12**:683-692.
- Hauser CE, Weiss J, Guillera-Arroita G, McCarthy MA, Giljohann KM, Moore JL. 2016. Designing detection experiments: three more case studies. Pages 171-178. *Proceedings of the 20th Australasian Weeds Conference*, eds R. Randall, S. Lloyd and C. Borger, Perth, Western Australia.
- Hoyer-Tomiczek U, Sauseng G, Hoch G. 2016. Scent detection dogs for the Asian longhorn beetle, *Anoplophora glabripennis*. *EPPO Bulletin* **46**:148-155.
- Hurt A, Guscio D, Tirmenstein DA, Richards N. 2014. Using Search Dogs for Biological Eradication Programs – A Tale about Dyer's Woad (*Isatis tinctoria* L.) in Schwarzländer M, and Gaskin JF, editors. *Proceedings of the 3rd northern rockies invasive Plants council conference*, Washington.
- Huso MMP, Dalthorp D. 2014. Accounting for unsearched areas in estimating wind turbine-caused fatality. *The Journal of Wildlife Management* **78**:347-358.
- Hutching G, Walrond C. 2007. Threatened species- Land management and conservation, Te Ara - the Encyclopedia of New Zealand, Available from <http://www.TeAra.govt.nz/en/photograph/14099/lassie> (accessed 23 August 2019).
- Jamieson TJ, Baxter GS, Murray PJ. 2018. You Are Not My Handler! Impact of Changing Handlers on Dogs' Behaviours and Detection Performance. *Animals (Basel)* **8**.
- Johnen D, Heuwieser W, Fischer-Tenhagen C. 2013. Canine scent detection—Fact or fiction? *Applied Animal Behaviour Science* **148**:201-208.
- Johnen D, Heuwieser W, Fischer-Tenhagen C. 2017. An approach to identify bias in scent detection dog testing. *Applied Animal Behaviour Science* **189**:1-12.
- Joseph LN, Maloney RF, Possingham HP. 2009. Optimal Allocation of Resources among Threatened Species: A Project Prioritization Protocol. *Conservation Biology* **23**:328-338.
- Kerley LL, Salkina GP. 2007. Using Scent-Matching Dogs to Identify Individual Amur Tigers from Scats. *The Journal of Wildlife Management* **71**:1349-1356.
- Kretser H, Glennon M, Whitelaw A, Hurt A, Pilgrim K, Schwartz M. 2016. Scat-detection dogs survey low density moose in New York. *Alces* **52**.
- Leigh KA, 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**:745-752.
- Lewis VR, Fouche CF, Lemaster RL. 1997. Evaluation of dog-assisted searches and electronic odor devices for detecting the western subterranean termite. *Forest Products Journal* **47**:7.
- Lindenmayer DB, Likens GE. 2010. The science and application of ecological monitoring. *Biological Conservation* **143**:1317-1328.
- Lit L, Schweitzer J, Oberbauer AM. 2011. Handler beliefs affect scent detection dog outcomes. *Conservation Biology* **14**:387-394.
- Long RA, Donovan TM, Mackay P, Zielinski WJ, Buzas JS. 2007a. Comparing scat detection dogs, cameras, and hair snares for surveying carnivores. *Journal of Wildlife Management* **71**:2018-2025.
- Long RA, Donovan TM, Mackay P, Zielinski WJ, Buzas JS. 2007b. Effectiveness of scat detection dogs for detecting forest carnivores. *Journal of Wildlife Management* **71**:2007-2017.
- MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA. 2002. Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One. *Ecology* **83**:2248-2255.

- Marchal S, Bregeras O, Puaux D, Gervais R, Ferry B. 2016. Rigorous Training of Dogs Leads to High Accuracy in Human Scent Matching-To-Sample Performance. *PLoS One* **11**:e0146963.
- Mathews F, Swindells M, Goodhead R, August TA, Hardman P, Linton DM, Hosken DJ. 2013. Effectiveness of search dogs compared with human observers in locating bat carcasses at wind-turbine sites: A blinded randomized trial. *Wildlife Society Bulletin* **37**:34-40.
- Matthew EE. 2016. The use of a sniffer dog for amphibian conservation ecology. Page 135. Environmental Sciences. North-West University, South Africa.
- McCarthy MA, Thompson CJ, Hauser C, Burgman MA, Possingham HP, Moir ML, Tiensin T, Gilbert M. 2010. Resource allocation for efficient environmental management. *Ecology Letters* **13**:1280-1289.
- McGregor HW, Hampton JO, Lisle D, Legge S. 2016. Live-capture of feral cats using tracking dogs and darting, with comparisons to leg-hold trapping. *Wildlife Research* **43**:313-322.
- Meek PD, et al. 2014. Recommended guiding principles for reporting on camera trapping research. *Biodiversity and Conservation* **23**:2321-2343.
- Moore J, Balmford A, Allnutt T, Burgess N. 2004. Integrating costs into conservation planning across Africa. *Biological Conservation* **117**:343-350.
- Moore JL, Hauser CE, Bear JL, Williams NSG, McCarthy MA. 2011. Estimating detection effort curves for plants using search experiments. *Ecological Applications* **21**:601-607.
- Nielsen TP, Jackson G, Bull CM. 2016. A nose for lizards; can a detection dog locate the endangered pygmy bluetongue lizard (*Tiliqua adelaidensis*)? *Transactions of the Royal Society of South Australia* **140**:234-243.
- Novoa C, Catusse M, Ellison L. 1996. Capercaillie (*Tetrao urogallus*) summer population census comparison of counts with pointing dogs and route census *Gibier Faune Sauvage* **13**:11.
- Nussear KE, Esque TC, Heaton JS, Cablk ME, Drake KK, Valentin C, Yee JL, Medica PA. 2008. Are wildlife detector dogs or people better at finding desert tortoises (*Gopherus agassizii*)?
- Orkin JD, Yang Y, Yang C, Yu DW, Jiang X. 2016. Cost-effective scat-detection dogs: unleashing a powerful new tool for international mammalian conservation biology. *Scientific Reports* **6**:34758.
- Paula J, Leal MC, Silva MJ, Mascarenhas R, Costa H, Mascarenhas M. 2011. Dogs as a tool to improve bird-strike mortality estimates at wind farms. *Journal for Nature Conservation* **19**:202-208.
- Porritt F, Mansson R, Berry A, Cook N, Sibbald N, Nicklin S. 2015. Validation of a short odour discrimination test for working dogs. *Applied Animal Behaviour Science* **165**:133-142.
- Possingham HP 2001. The business of biodiversity: applying decision theory principles to nature conservation. Australian Conservation Foundation, Fitzroy, Australia.
- Reindl-Thompson SA, Shivik JA, Whitelaw A, Hurt A, Higgins KF. 2006. Efficacy of scent dogs in detecting black-footed ferrets at a reintroduction site in South Dakota. *Wildlife Society Bulletin* **34**:1435-1439.
- Rolland RM, Hamilton PK, Kraus SD, Davenport B, Gillett RM, Wasser SK. 2006. Faecal sampling using detection dogs to study reproduction and health in North Atlantic right whales (*Euhalaena glacialis*). *Journal of Cetacean Research and Management* **8**:121.
- Ruusila V, Pesonen M. 2004. Interspecific cooperation in human (*Homo sapiens*) hunting: the benefits of a barking dog (*Canis familiaris*). *Annales Zoologici Fennici* **41**:545-549.
- Saito T, Rehmsmeier M. 2015. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE* **10**:e0118432.
- Scheffers BR, Yong DL, Harris JBC, Giam X, Sodhi NS. 2011. The World's Rediscovered Species: Back from the Brink? *PLOS ONE* **6**:e22531.
- Smith DA, Ralls K, Cypher BL, Maldonado JE. 2005. Assessment of scat-detection dog surveys to determine kit fox distribution. *Wildlife Society Bulletin* **33**:897-904.

- Smith DA, Ralls K, Hurt A, Adams B, Parker M, Davenport B, Smith M, Maldonado JE. 2003. Detection and accuracy rates of dogs trained to find scats of San Joaquin kit foxes (*Vulpes macrotis mutica*). *Animal Conservation* **6**:339-346.
- Syrotuck WG 2000. Scent and the Scenting Dog. Barkleigh Productions, Pennsylvania.
- Wasser SK, Davenport B, Ramage ER, Hunt KE, Parker M, Clarke C, Stenhouse G. 2004. Scat detection dogs in wildlife research and management: application to grizzly and black bears in the Yellowhead Ecosystem, Alberta, Canada. *Canadian Journal of Zoology* **82**:475-492.
- Wasser SK, Hayward LS, Hartman J, Booth RK, 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*). *PloS one* **7**:e42892.
- Welch JB. 1990. A detector dog for screwworms (Diptera: Calliphoridae). *Journal of economic entomology* **83**:1932-1934.

Table 1 A confusion matrix or contingency table for evaluating conservation scent dogs, summarising: true positive (TP), the number of targets positively alerted on; false positive (FP), the number of non-targets incorrectly alerted on; false negative (FN), the number of targets missed; and true negative (TN), the number of non-targets ignored. These provide the data needed for calculating sensitivity, precision, accuracy and other statistical measures.

		Targets:	
		Presence	Absence
Dogs Response	Alert	TP	FP
	No Alert	FN	TN

Sensitivity (aka true positive rate, recall)	$\frac{TP}{TP + FN}$
Precision (aka reliability, positive predictive value)	$\frac{TP}{TP + FP}$
Accuracy	$\frac{TP + TN}{TP + FP + TN + FN}$

Table 2 Key factors identified as important to standardise the reporting of conservation dog studies

Key Factors	Reporting recommendations
Identify Survey Objective	Define research hypothesis or management goal
Performance Measures	Record total time spent surveying and area searched
	Present field sensitivity and precision as proportions and record sample sizes
Scale	Present effort as time per unit area or transect
	Report the area searched, transect or patch size
Costs	Specify the type of handler (professional/amateur/purchase/lease)
	Present total cost of dog surveys and total survey effort achieved (or cost per unit effort)
	Specify the currency in which the costs are presented
Type of Scent detection task	Specify length of the study
	Specify if search strategy is free ranging or controlled
	Report environmental conditions such as weather, vegetation, topography
Human influences	Describe target samples
	Describe differences between real and proxy targets used in testing
	Report design parameters such as blinding and randomisation
	Report previous experience of the dog handler
	Report previous experience of the dog

Figure Legend

Figure 1 Framework for evaluation of conservation detection dogs showing the five measures (medium-grey circles) that are important for evaluating conservation detection dogs. Scale should be considered when measuring precision, sensitivity, effort and cost to determine cost effectiveness. Determining cost effectiveness allows for useful comparisons of detection dogs with other tools and search methods, which is crucial for overall evaluation.

Figure 2 Precision, sensitivity, efficiency, cost and comparison of how dogs perform relative to other methods are all important when evaluating conservation dogs, however their relative importance will vary depending upon the conservation objectives being addressed and who the participating organisation or individuals are.

Figure 3 (a) Reported occurrences of the five parameters needed for effective evaluation of conservation detection dogs. (n=61) (b) Distribution of reported precision and sensitivity results from the reviewed studies with the centre line of the box representing the median value, the boxes indicating first and third quartile and the whiskers showing maximum and minimum values reported. (c) Reported design parameters showing the proportion of studies which reported the various parameters and those that did not.

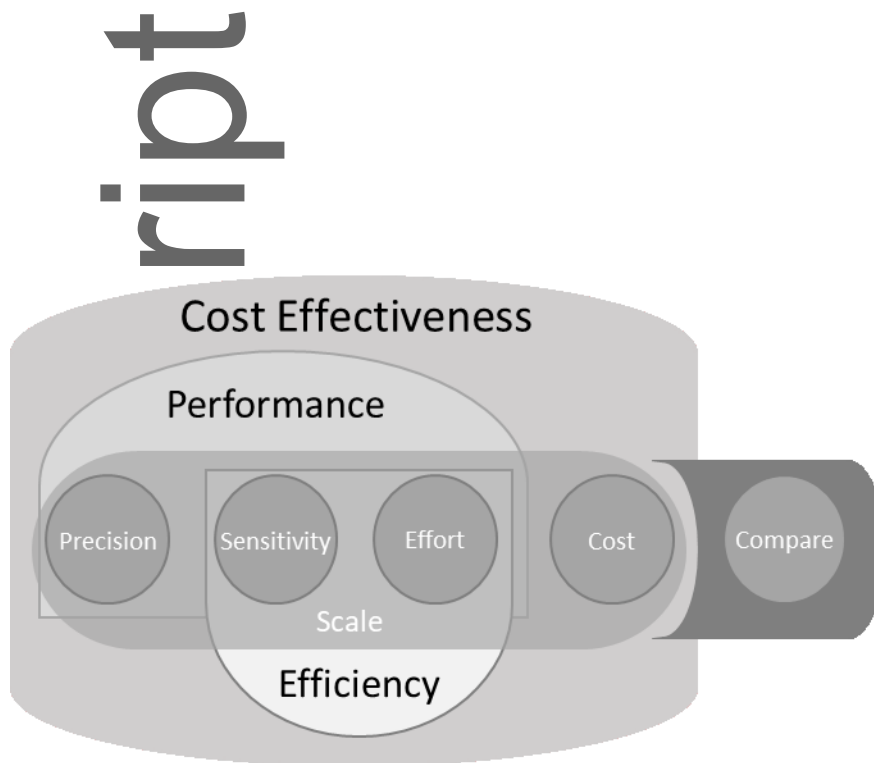


Figure 1 Framework for evaluation of conservation detection dogs showing the five measures (medium-grey circles) that are important for evaluating conservation detection dogs. Scale should be considered when measuring precision, sensitivity, effort and cost to determine cost effectiveness. Determining cost effectiveness allows for useful comparisons of detection dogs with other tools and search methods, which is crucial for overall evaluation.

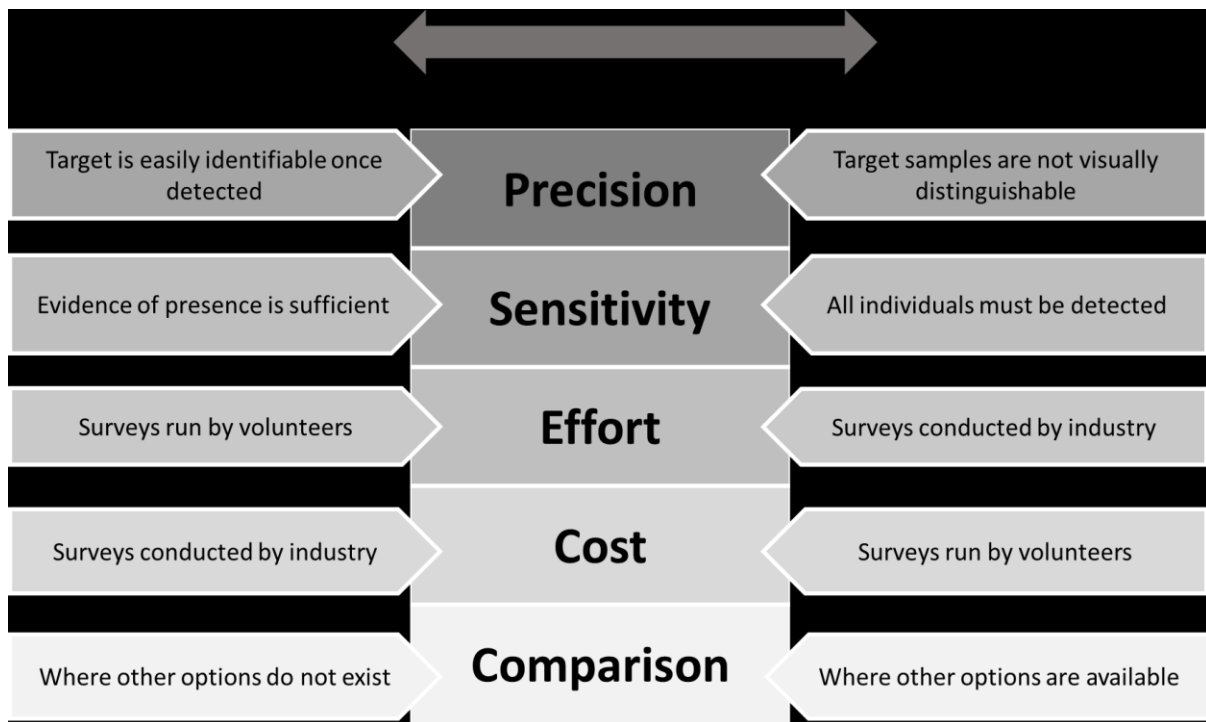


Figure 2 Precision, sensitivity, effort, cost and comparison of how dogs perform relative to other methods are all important when evaluating conservation dogs. However, their relative importance will vary depending upon the conservation objectives being addressed and who the participating organisation or individuals are.

Author
Mar

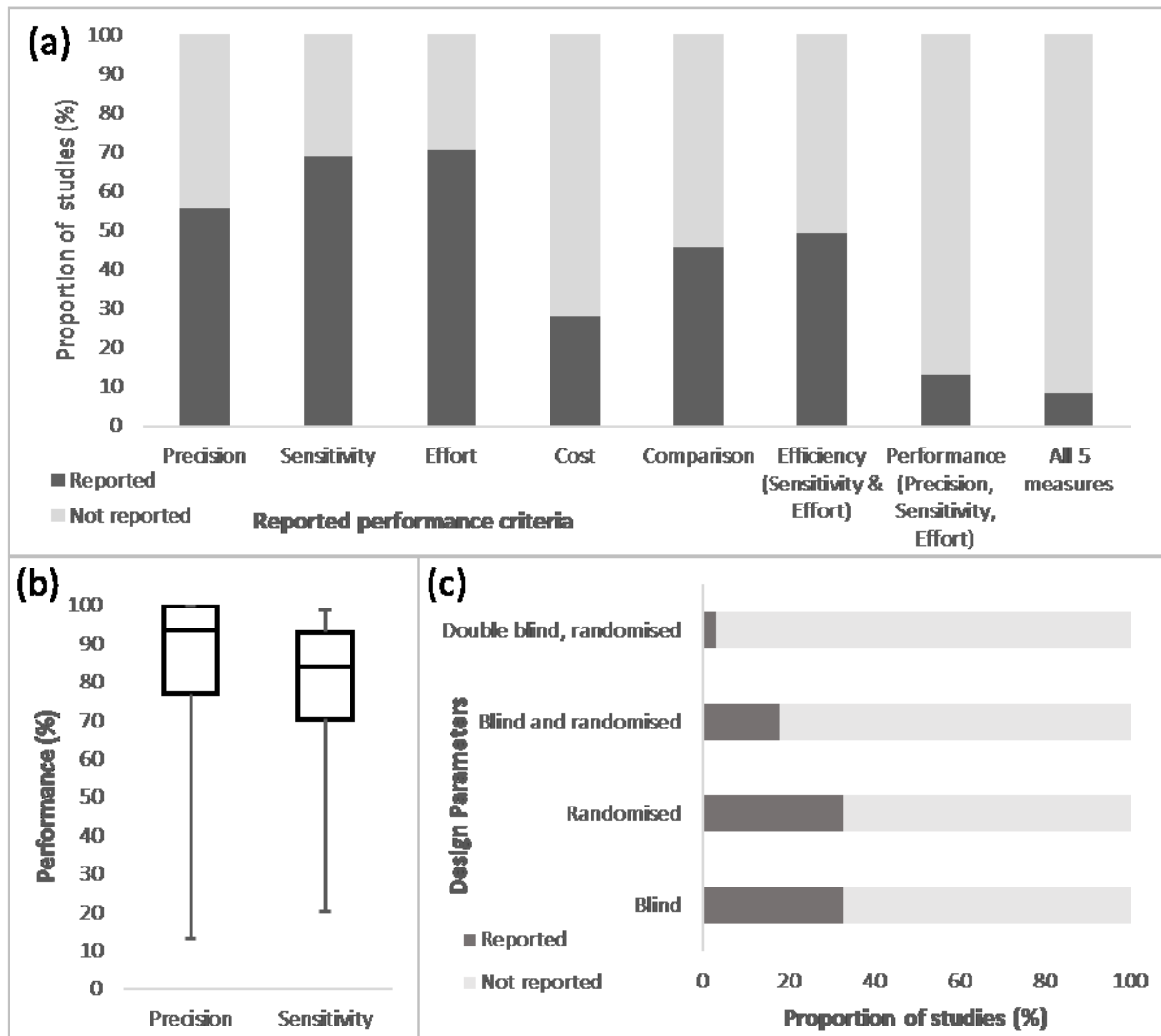


Figure 3 (a) Reported occurrences of the five parameters recommended for effective evaluation of conservation detection dogs within papers reporting at least one measure. (n=61) (b) Distribution of reported precision and sensitivity values from the reviewed studies with the centre line of the box representing the median value, the boxes indicating first and third quartile and the whiskers showing maximum and minimum values reported (n=61). (c) Reported design parameters of the 61 studies showing the proportion of studies which reported the various parameters and those that did not.

Author